ELSEVIER

Surface Science 318 (1994) 379–394

surface science

Experimental and simulated tunneling spectra of the polar ZnO surfaces

Paul M. Thibado *[,a], Gregory S. Rohrer [b], Dawn A. Bonnell [c]

[a] *Physics Department, University of Pennsylvania, Philadelphia, PA 19104, USA*
[b] *Department of Materials Science and Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA*
[c] *Department of Materials Science and Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA*

Received 24 January 1994; accepted for publication 28 June 1994

Abstract

Scanning tunneling microscopy was used to study the geometric and electronic structure of the two polar surfaces of zinc oxide in ultra-high vacuum. Even though the topographies of the surfaces are similar, the tunneling spectra differ on each surface plane (Zn-terminated or O-terminated), differ depending on the proximity to mesas, and are affected by the adsorption of O_2. Calculations of tunneling spectra allow these data to be interpreted within the framework of established models for tunneling to semiconductor surfaces. These calculations demonstrate that changes in the spectra are due to changes in the local electronic properties of the surface. For example, features in tunneling spectra recorded at the edge of mesas can be related to spatially localized electronic states in the band-gap region, and changes induced by the adsorption of O_2 can be explained by a reduction in the surface charge density. Negative differential conductance (NDC) was observed and a model which uses a varying density of states combined with tip-induced band bending is presented.

1. Introduction

Zinc oxide crystallizes in the hexagonal wurtzite structure. This structure may be schematically described as alternating parallel planes composed entirely of Zn or O ions stacked along the *c*-axis. These crystallographically distinct planes are conventionally differentiated by labeling the Zn-plane (0001) and the O-plane ($000\bar{1}$). Single crystals have a well-defined polarity, being terminated on one side by an O plane and on the other macroscopically separated side by a total charge neutralizing Zn plane. It has been known for some time that these different surfaces have very different physical and chemical properties [1]. For example, the O surface etches more rapidly in acidic solutions and is less resistant to abrasive wear than is the Zn surface [1]. In addition, the morphologies of the polar surfaces differ. ZnO grows into needle-shaped single crystals bounded on one end by a flat O surface and on the other by a steep pyramid [1]. It has also been observed that the conductivity [2], work function [3], and electronic structure of these surfaces differ [4,5]. Although it might seem obvious that these differences in properties occur as a result of the outermost layer of ions being either Zn or O, the mechanism by which this final layer of ions influences these proper-

* Corresponding author. Fax: +1 215 573 2128.

SSDI 0039-6028(94)00423-4

ties is certainly not clear. Furthermore, it is unclear how a complex property, such as abrasive wear, is related to the last layer of ions. It is likely that the differences in the atomic termination influence the surface morphology, defect structure, segregation of impurities and defects, as well as local electronic characteristics, and it is these factors that are directly responsible for the differences in the surface properties.

In an effort to learn more about the geometric and electronic structural differences between the ZnO polar surfaces, a scanning tunneling microscope (STM) was used to record topographic images and spatially resolved tunneling spectra of both surfaces. These surfaces were studied after annealing in ultra-high vacuum (UHV) and after dosing with oxygen. While the topographic images share common features, the tunneling spectra are influenced by the surface plane, the proximity to mesas, and the adsorption of O_2. A computational model was used to understand how differences in the tunneling spectra were related to changes in physical parameters and the local electronic properties of the surface. The model was originally described by Bono and Good [6], further developed by Feenstra and Stroscio [7,8], and takes into account the effects of bias-dependent band bending, tunneling through the depletion region of the semiconductor, and a finite surface charge density.

2. Experimental technique

The zinc oxide single crystal used in this study was optically transparent, colorless, and unintentionally doped. The electronic properties of this nonstoichiometric compound are determined by a native population of point defects. These defects have ionization energies in the range of 0.01 to 0.1 eV and, thus, act as shallow donors yielding a small extrinsic conductivity [9,10]. Unfortunately, the concentration of these defects, which are thought to be oxygen vacancies and zinc interstitials, is determined by a combination of chemical and kinetic factors and, therefore, depends on the thermal history of the sample. Typical concentrations for these defects are between 10^{16} to 10^{17} per cm^3 [9,10]. In this experiment, the total concentration of these defect states was undetermined and it is likely that it varied somewhat in response to the in vacu thermal treatments that were used to prepare the surfaces. Nevertheless, based on the clarity of the crystal and reported donor concentrations for other undoped ZnO crystals, it appears appropriate to assume that the donor concentration was in the range of 10^{17} cm^{-3}.

The anisotropic etching behavior of the ZnO(0001) and $(000\bar{1})$ surfaces was used to determine the polarity of the crystal [1]. Each face was mechanically polished to eliminate all optically visible defects by sequentially reducing the size of abrasive alumina grinding media to 0.05 μm. The O-terminated surface of the crystal was polished and analyzed first, then the Zn-terminated surface was polished and returned to the UHV-STM chamber for analysis.

For each surface, the sample was clamped into firm contact with a Si wafer that was heated to 930°C for one to two hours. The temperature was measured with a thermocouple at the surface of the wafer, then confirmed with an optical pyrometer. This temperature more accurately represents the temperature of the heater than that of the sample. The sample did not radiate as much light, remained transparent, and was presumably at a much lower temperature, probably closer to 700°C. This annealing procedure results in oxygen sublimation, which could increase the surface charge density and local carrier concentration. It is also possible that an oxygen vacancy concentration profile exists from the surface to the bulk. After annealing, the crystal exhibited sharp LEED patterns that did not change with further annealing. The O surface exhibited 1×1 symmetry while the Zn surface exhibited the $\sqrt{3} \times \sqrt{3}$ reconstruction, consistent with earlier work by van Hove and Leysen [11]. Auger electron spectroscopy showed that these heat treatments removed carbon from both surfaces, but a small amount of potassium impurity was detected on the Zn surface. After recording STM images and spectra from each of the untreated surfaces, O_2 was leaked into the chamber and then pumped away before re-examining the surfaces. Beginning with a background pressure of 5×10^{-10} Torr, the O surface was exposed to 2×10^{-7} Torr of O_2 for one hour, while the Zn surface was exposed to 1×10^{-3} Torr of O_2 for a few minutes.

Images were recorded in constant current mode with a sample bias of 2.75 (-3.0) V and tunneling

currents of 0.6 and 0.05 (0.6) nA's for the Zn- (O-) terminated surface. A mechanically formed Pt tip was sharpened and cleaned by applying a 90 V potential and passing 1 μA of current between the sample and tip for several minutes. It was necessary to repeat the tip sharpening and cleaning procedure periodically to maintain resolution and junction stability. Tunneling spectra were acquired at preselected locations during image acquisition. When the tip reached one of the selected locations, the feedback loop was momentarily disabled, at which time the sample bias was ramped linearly from -3.0 to 3.0 V (O surface) or 3.0 V to -3.0 (Zn surface) in less than 100 ms while simultaneously recording the current. After a brief stabilization interval, the process was repeated so that each spectrum was the average of ten sequential measurements. Once the image was recorded, each I–V curve could be correlated to a specific point on the image. Current measurements below 1 pA, the detection limit of the sensor, were disregarded.

3. Results

Atomic resolution images were never observed, despite repeated attempts. The primary observation was that while the surfaces are faceted, they are not particularly flat. Individual terraces were on the order of 100×100 Å^2, which suggests that a considerable fraction of the surface atoms reside on terrace edges. Constant current images of the Zn-terminated and O-terminated surfaces taken at a tunneling current of 0.6 nA and sample biases of 3 and -3 V, respectively, are shown in Fig. 1. Under the sample preparation conditions described above, the differences between the two surfaces are more apparent in the systematic and reproducible variation in the tunneling spectra than in the topographic images. This result suggests that while the geometric microstructure of the surfaces are similar, the electronic properties differ. Minor drift in the lateral directions does occur during imaging, but since this paper quantifies only the electronic properties this drift is neither

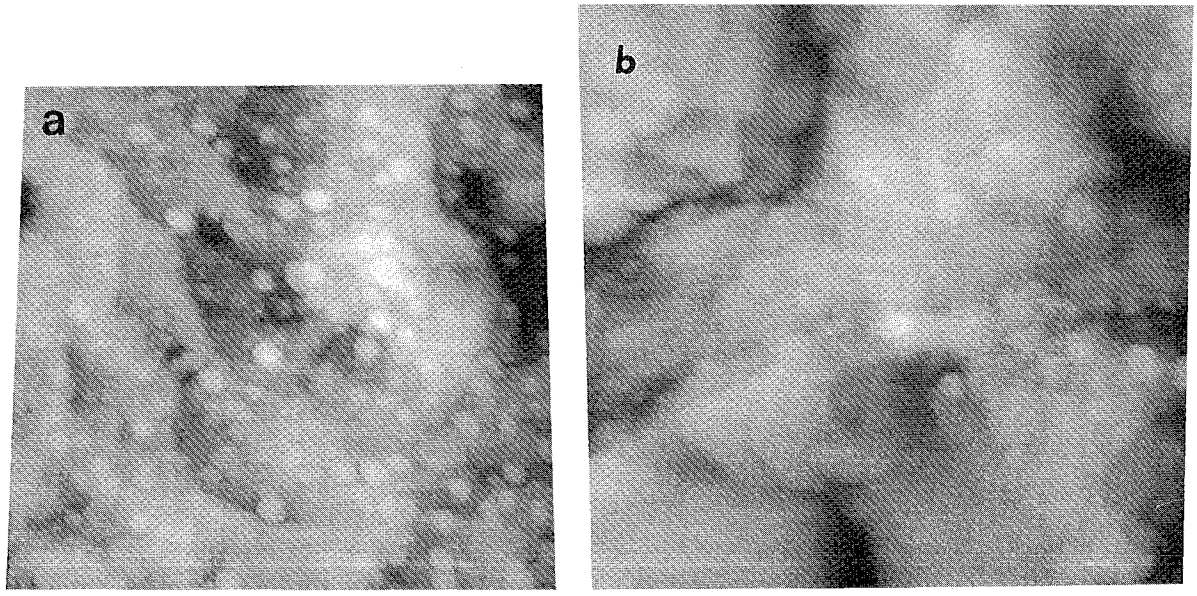

Fig. 1. (a) Constant current STM image of Zn-terminated surface 0.1×0.1 μm^2 with a black to white scale of 30 Å. This image was acquired with a sample bias of 3 V, and tunneling current of 0.6 nA. (b) Constant current STM image of O-terminated surface 0.1×0.1 μm^2 with a black-to-white scale of 30 Å. This image was acquired with a sample bias of -3 V, and a tunneling current of 0.6 nA. These images illustrate the degree of roughness as well as the geometric structure present on the mesa-like features.

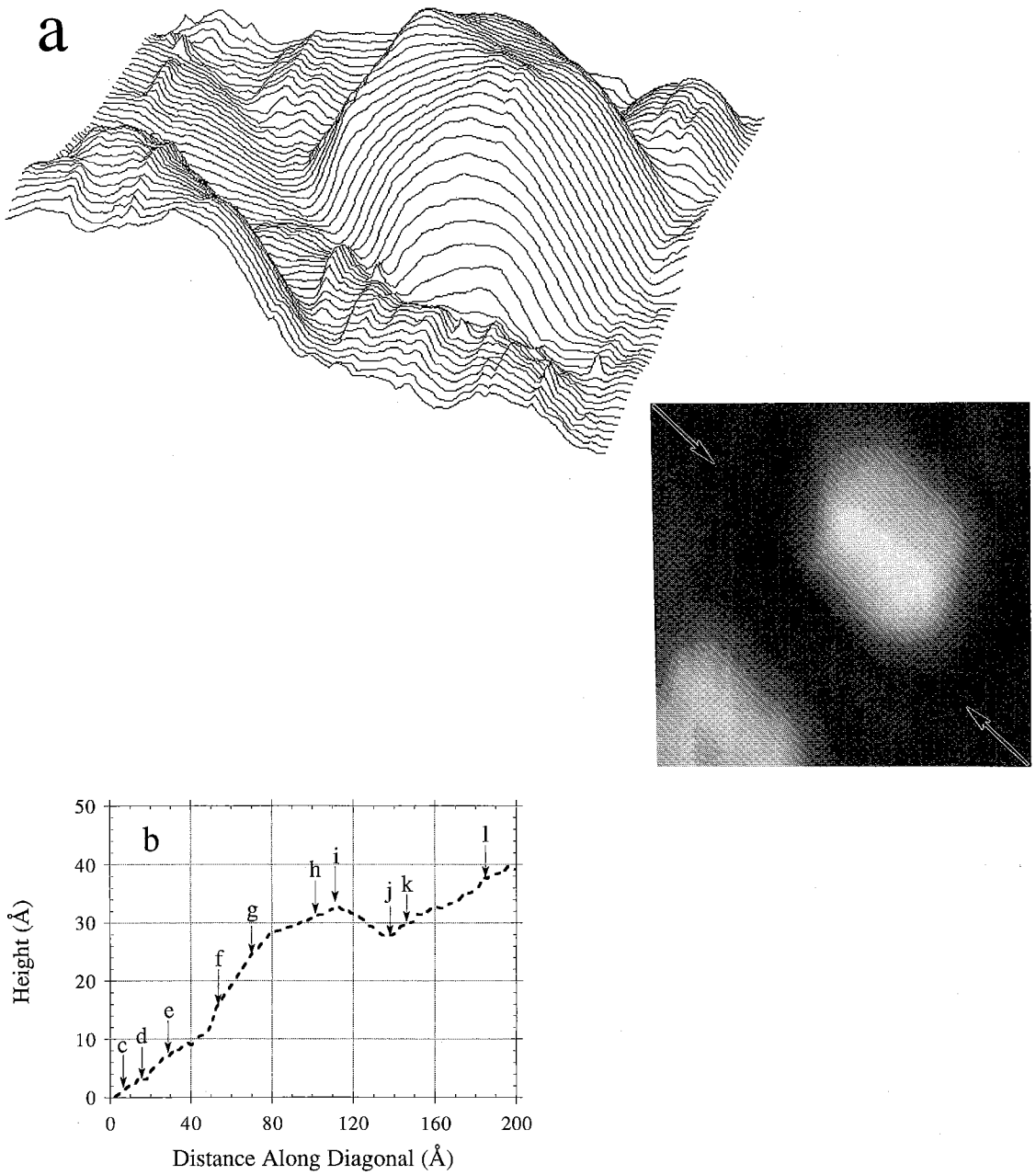

Fig. 2. Constant current STM image of Zn-terminated surface, profile along diagonal, and selected I–V curves along diagonal. (a) The 3D rendering of a 140×140 Å^2 area with a height of 20 Å. The inset is a 2D gray scale representation of the same data. Both images are shown after a plane subtraction was performed. (b) Line profile extracted from the image along the diagonal from the upper left corner to the lower right corner (without plane subtraction) with selected I–V locations labeled with figure letters. (c–ℓ) Selected I–V spectra along the diagonal profile.

quantified nor corrected. Drift in the vertical direction was corrected with a background subtraction. Upon correction, all mesa heights were integral multiples of unit cell dimensions.

3.1. Zn surface

Fig. 2 illustrates the manner in which tunneling spectra vary with their proximity to mesas. The topographic image in Fig. 2a shows two mesa-like structures protruding 10–12 Å from a flat region of the Zn-terminated surface and Fig. 2b shows the topographic profile along a line from the upper left- to the lower right-hand corner of the imaged area in Fig. 2a. This profile is plotted without a background subtraction to illustrate the measured geometry of the mesa. The overall slope of the sample is due primarily to thermal drift which is significantly compounded during the spatially-resolved spectroscopy experiment (due to the time required for the acquisi-

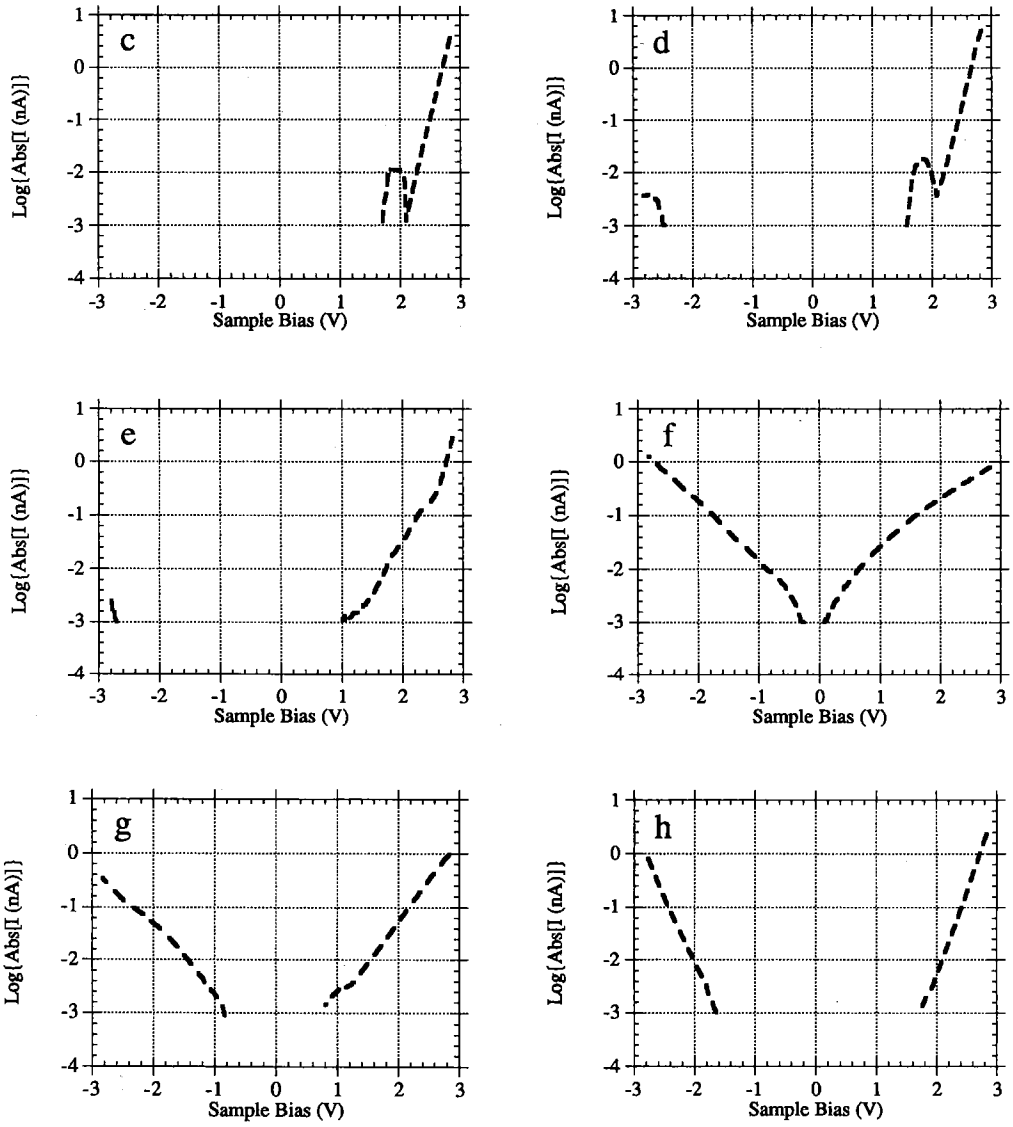

Fig. 2 (continued).

tion of I–V curves). Along this profile, which crosses the mesa, 100 I–V spectra were recorded and a representative subset of these are shown in Fig. 2c-l. The spectra can be categorized into 3 characteristic groups; those recorded far from the edge of a mesa (c, d, k, l), those recorded near an edge of a mesa (e, h, i, j), and those recorded directly at an edge of a mesa (f, g) (see Fig. 2b). Representatives of each of the three groups were selected for a more detailed analysis and are shown in Fig. 3 together with a characteristic I–V spectrum from a flat region of the surface after treatment in O_2 (Fig. 3d).

The interpretation of these data will be based on the characteristic features of each spectrum. The two important characteristics of spectra from the Zn-terminated surface (Fig. 3a) are that no current is detected for applied sample biases (V_a) between -3 and 1.6 V and that at higher biases, the current first increases, then drops about a factor of 3 before increasing again. The decreasing current (with increasing bias) phenomenon is referred to as negative differential conductance (NDC). Note, unlike other observations of NDC the results presented here are not spatially localized, but occur over extended areas on the surface. The tunneling spectrum representative of areas near the edge of the mesa (see Fig. 3b) is characterized by an absence of current between -2.7 and 1.0 V and steep rises in the current at $V_a = -2.7$ V and $V_a = 2.5$ V. The tunneling spectrum representative of those recorded on the edge of a mesa (Fig. 3c) is characterized by continuous increases in the current (on both sides of the Fermi level) and the absence of inflections or sharp increases. Finally, the spectrum that is representative of the flat Zn-terminated surface after it has been exposed to oxygen (Fig. 3d) is similar to that recorded near an edge of the mesa and is characterized by sharp increases in the tunneling current at $V_a = -2.2$ V and $V_a = 2.3$ V.

3.2. O surface

Spatially resolved tunneling spectra from the O-terminated surface were presented in an earlier letter

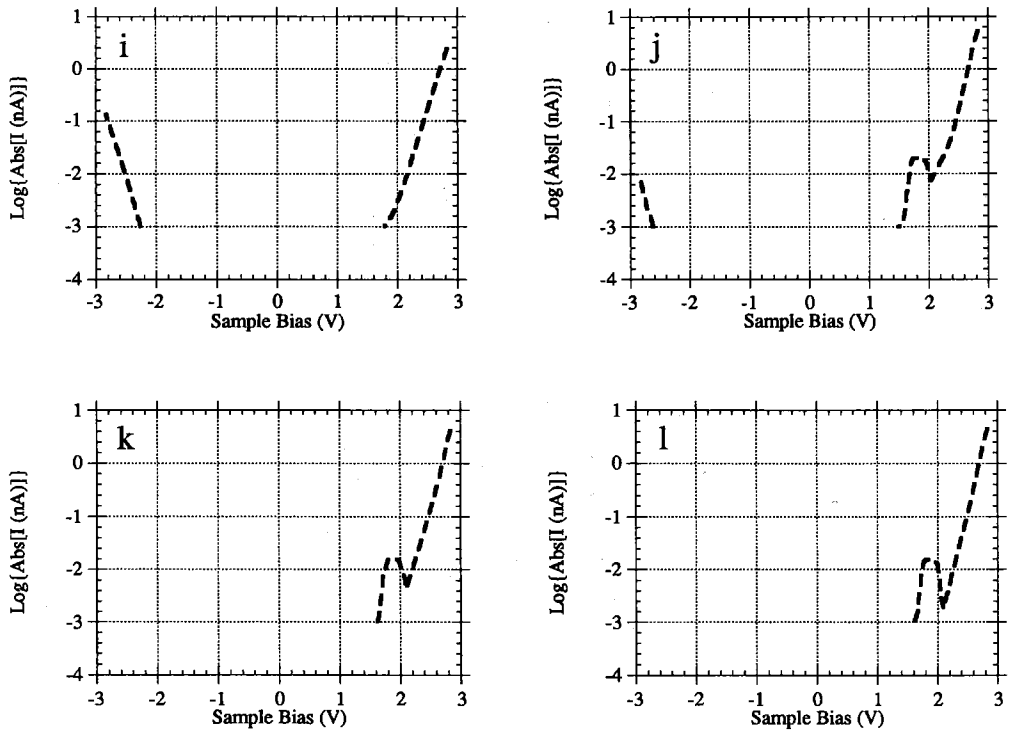

Fig. 2 (continued).

to this journal [12] with the qualitative interpretation that spatially localized electronic states at the edge of mesas caused changes in the tunneling spectra. The justification for this interpretation was a peak in the $(dI/dV)/(I/V)$ spectrum at -2.0 V whose intensity grew as the distance from the edge of the mesa diminished. In this paper, an attempt to quantify the interpretation of this phenomenon using a computational model is made and, thus, only selected characteristic spectra are presented. Fig. 4a shows a spectrum from a flat region of the O-terminated surface, which shares the same characteristic features as the spectrum recorded at the edge of the mesa on the Zn-terminated surface (Fig. 3c). The tunneling spectrum representative of those recorded at an edge of a mesa on the O-terminated surface (Fig. 4b) is characterized by symmetric inflections in the positive and negative bias regions (at $V_a = -1.8$ V and $V_a = 1.6$ V). The tunneling spectrum characteristic of flat regions of the surface following exposure to O_2 is shown in Fig. 4c. In this spectrum, the current in the negative bias region is similar to that in the spectrum of the untreated flat region on the O-terminated surface (Fig. 4a), but the magnitude of the positive bias current is significantly reduced. In addition, there is a NDC feature at $V_a = 2$ V. The theory that will serve as a basis for the detailed analysis of the tunneling spectra presented in Section 5 is described next.

4. Theoretical calculations

4.1. General

The tunnel current calculations are based on a model initially presented by Bono and Good [6] and later extended by Feenstra and Stroscio [7,8]. In this model, the donor band is assumed to be empty and the electrons are assumed to fill the conduction band

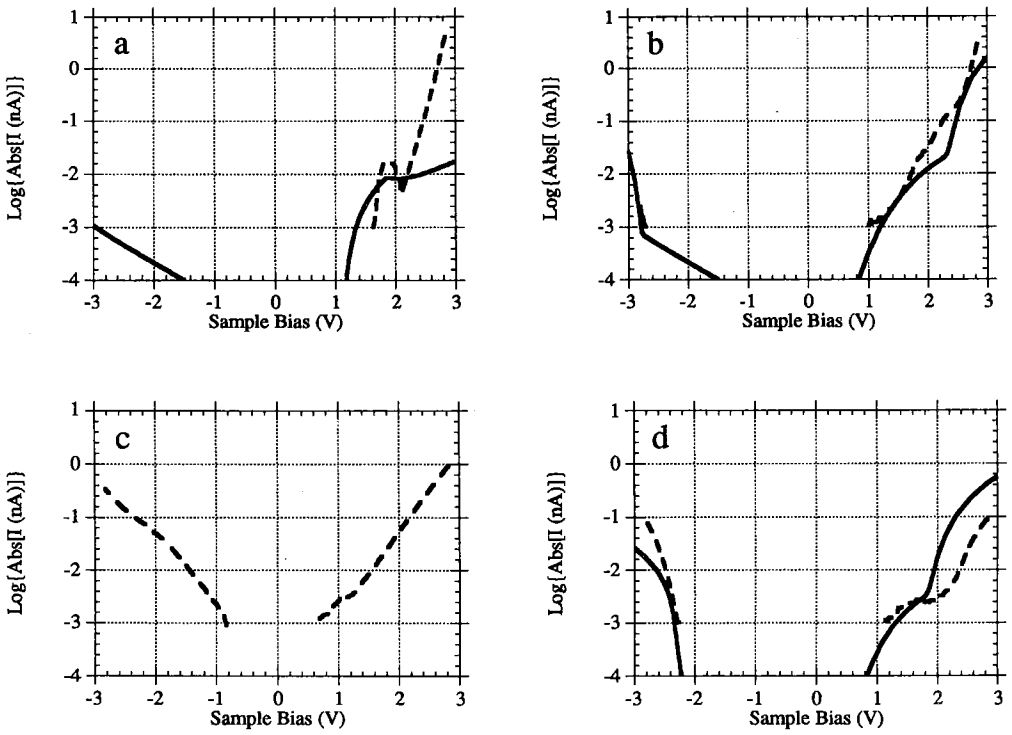

Fig. 3. Selected tunneling spectra from the Zn-terminated surface. Solid lines are simulations and dashed lines are data. (a) Tunneling spectrum from a flat region far from the mesa in Fig. 2a. (b) Tunneling spectrum from a region near the mesa in Fig. 2a. (c) Tunneling spectrum from the edge of the mesa in Fig. 2a. (d) Tunneling spectrum taken from a flat region far from a mesa for the O_2-dosed Zn-terminated surface.

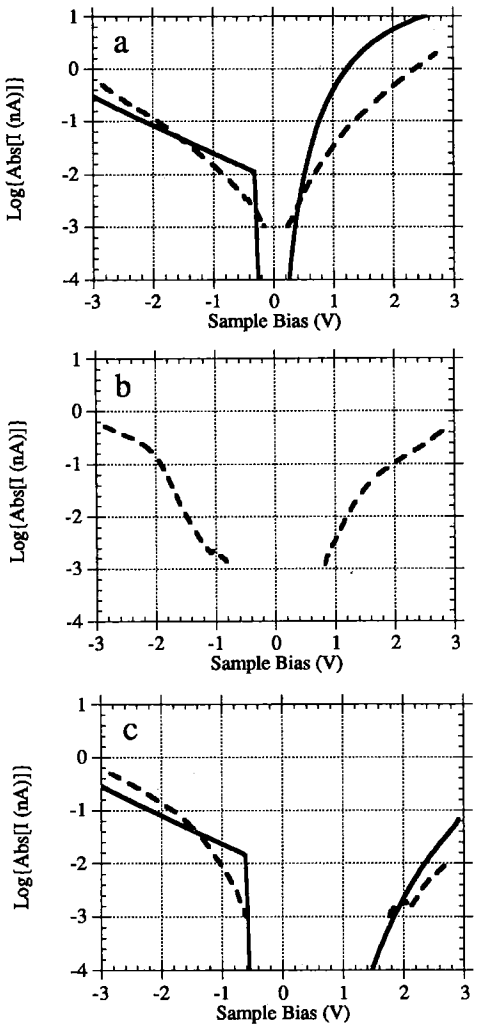

Fig. 4. Selected tunneling spectra from the O-terminated surface. Solid lines are simulations and dashed lines are data. (a) Tunneling spectrum from a region far from a mesa. (b) Tunneling spectrum taken on an edge of a mesa. (c) Tunneling spectrum taken from a flat region far from a mesa for the O_2-dosed O-terminated surface.

as if the sample were at absolute zero. This places the Fermi level at an energy,

$$\left(\frac{3\pi^2 n_d}{2\sqrt{2}}\right)^{2/3} \frac{\hbar^2}{m_c^* e},$$

above the conduction band (CB) minimum, where n_d is the average donor concentration, $\hbar$ is Planck's constant, m_c^* is the ZnO CB effective mass, and e is the elementary charge. The tunnel current is $I = AJ$, where A is the tip area, and J is the current density, given by Eq. (2) in Feenstra and Stroscio [7]:

$$J_{\substack{\mathrm{cb}\\\mathrm{vb}}} = \frac{me}{2\pi^2\hbar^3}\int_{E_f - eV_a}^{E_F} \mathrm{d}E\,\Theta\left[\pm\left(E - E_{\substack{\mathrm{cb}\\\mathrm{vb}}}\right)\right] \times \int_{E(1\mp\alpha_{\substack{\mathrm{cb}\\\mathrm{vb}}})\pm\alpha_{\substack{\mathrm{cb}\\\mathrm{vb}}}E_{\substack{\mathrm{cb}\\\mathrm{vb}}}}^{E} \mathrm{d}E_z\, D_{\substack{\mathrm{cb}\\\mathrm{vb}}}(E_z),$$

where $D_{\substack{\mathrm{cb}\\\mathrm{vb}}}(E_z)$ is the total transmission probability.

$$D_{\substack{\mathrm{cb}\\\mathrm{vb}}}(E_z) = D_v(E_z) D_{\substack{\mathrm{cb-S}\\\mathrm{vb-S}}}(E_z),$$

where D_v and $D_{\substack{\mathrm{cb-S}\\\mathrm{vb-S}}}$ are the transmission coefficients for an electron to penetrate the vacuum barrier between the tip and the sample (given by Eq. (4) in Feenstra and Stroscio [7]) and the semiconductor barrier due to band bending, respectively.

$$D_{\substack{\mathrm{cb-S}\\\mathrm{vb-S}}}(E_z) = \exp\left[-2\sqrt{2m_{\substack{\mathrm{cb}\\\mathrm{vb}}}^*/\hbar^2}\int_0^{z_b} \mathrm{d}z\sqrt{E_{\substack{\mathrm{cb}\\\mathrm{vb}}}(z) - E_z}\right], \quad (1)$$

where $E_{\substack{\mathrm{cb}\\\mathrm{vb}}}(z)$ is the band edge as a function of distance from the surface, and z_b is the classical turning point inside the semiconductor. Eq. (1) is reproduced here because this coefficient may be difficult to calculate directly. The complexities of the calculation are discussed and a simplified procedure is outlined in the Appendix.

The eight fixed parameters used in these calculations included the electron affinity of ZnO = 4.2 eV [9,10], the work function of the Pt tip = 5.65 eV [13], the relative dielectric constant of ZnO = 8.5 [9,10], the temperature = room temperature, the energy gap of ZnO = 3.2 eV [14], and the tip area = (25 Å)^2. Also, by assuming parabolic bands, the effective masses used are 0.3 and 0.7 ($(0.3^{3/2} + 0.6^{3/2})^{2/3}$) for the conduction band and valance band of ZnO, respectively [14]. Of these fixed parameters, all but the tip area are accepted physical constants. The estimation of the tip area represents an educated guess based on step widths in the topographic images. The effect of the tip area in the calculations is to scale the magnitude of the current, thus these

calculations are assumed valid only to an overall multiplicative factor.

Although the four remaining parameters were allowed to range over physically realistic values, they were constrained by the condition that each set remain as consistent as possible from simulation to simulation. These parameters are the donor concentration, the sample–tip separation (s), the surface charge density (σ), and the positions of spatially localized surface gap states (E_{gs}, where E_{gs} is the energy difference between the CB minimum and the gap state maximum). The donor concentration, which was fixed for all spectra recorded on the same surface, was chosen so that the magnitude of current observed for negative biases from the flat regions on both the Zn- and O-terminated surfaces (Figs. 3a and 4a) was reproduced with reasonable accuracy. It was necessary to use a different donor concentration for each of the two surfaces to explain the observed spectra. This is dictated by the current at $V_a = 3$ V being larger for the Zn than for the O surface, yet in the negative bias region the reverse is true. This cannot be attributed to a change in the sample–tip separation or to any other scaling factor. A consistent value for the donor concentration was determined to be 10^{17} cm^{-3} (6×10^{17} cm^{-3}) for the Zn (O) surface. These values are within reasonable bounds, especially when one considers that defect concentrations are known to be altered by thermal treatments [9,10]. In addition, it has been reported that the O surface has surface states in the gap, which may increase the carrier concentration for this surface [3]. The sample–tip separations, which simply scale the magnitude of the simulated currents, were assumed to be between 6 and 8 Å. The last two parameters (the surface charge density and the energetic positions of occupied states in the gap) were manipulated to reproduce the observed trends. The significance of these parameters and their influence on the tunneling spectra are discussed in Section 5.

Consideration of the dynamic tip-induced band bending leads to the prediction that electrons will accumulate on the surface of the semiconductor at negative biases; however, such a prediction significantly overestimates the current caused by the donor electrons [7]. As a result, the calculations were carried out using the approximation that electrons are unable to relax into states near the surface and below the CB minimum of the semiconductor on the time scale of the experiment. Thus, the calculation presented here allows the bands to bend according to the Fermi–Dirac statistics, but the electrons are assumed not to fill states below the CB minimum. The consequences of this approximation on the simulations are addressed in Section 5.

The modeling component of this study was carried out with the goal of finding a consistent set of parameters which fall within acceptable physical bounds and can be used to explain the characteristic features of the spectrum. No attempt was made to find a set of "best-fit" parameters and the absolute values of the reported numbers are considered to be less significant than the variation in values of the parameters in different simulations. The primary advantages of the simulations are that they allow one to test postulated models and they allow the characteristic features of the spectra to be understood in terms of physically intuitive parameters.

4.2. Negative differential conductance

Some aspects of the NDC phenomenon can be reproduced using a model that includes an empty donor band (DB) 40 meV wide and 10 meV below the CB minimum as shown schematically in Fig. 5. The model allows for the maximum change in current to occur during the NDC period by setting the transmission coefficient to tunnel into states in the

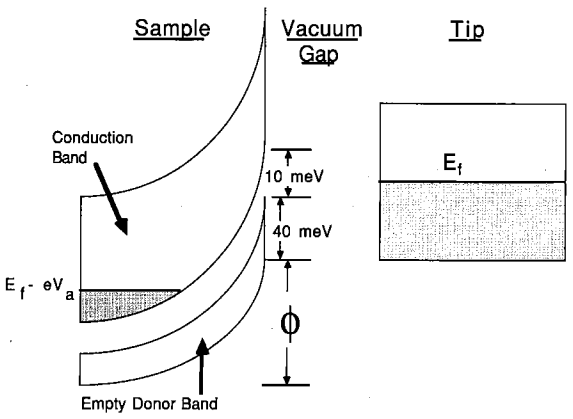

Fig. 5. Schematic diagram showing tunneling into the regions below the CB edge which are bent up above the Fermi level of the sample.

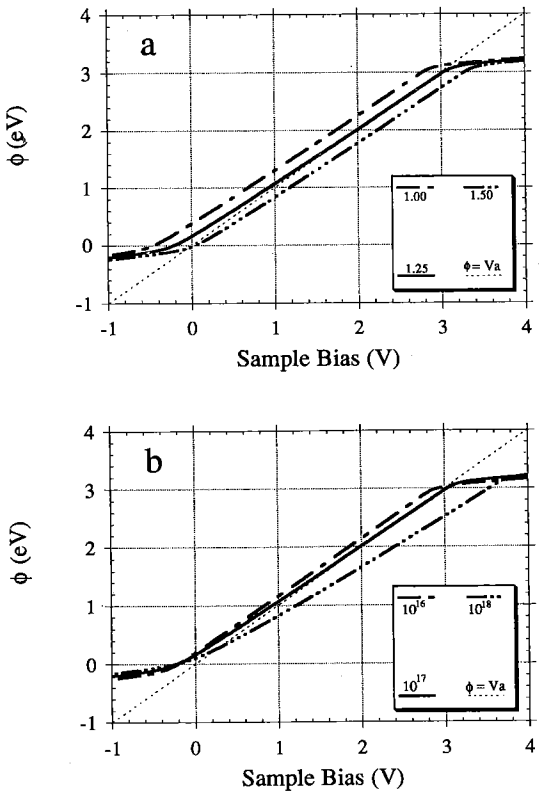

Fig. 6. Calculated total potential drop across the sample (ϕ) as a function of the total applied sample bias. (a) Three curves calculated with varying scaled surface charge density and labeled: $\sigma_s = 1.50$ eV, $\sigma_s = 1.25$ eV, and $\sigma_s = 1.00$ eV. (b) Three curves calculated with varying donor concentration and labeled: $n_d = 10^{16}$ cm^{-3}, $n_d = 10^{17}$ cm^{-3}, and $n_d = 10^{18}$ cm^{-3}.

conduction band that are beneath the surface ($D_{db\text{-}cb}$) to zero, the transmission coefficient to tunnel into DB states at the surface (D_{db}) to unity, and the transmission coefficient to tunnel into the DB states that are beneath the surface ($D_{db\text{-}vb}$) is zero. The mechanism by which this narrow band-gap (10 meV) creates the energetically much wider NDC feature (300 meV) can be understood in terms of tip-induced band bending.

4.3. Tip-induced band bending

The relationship between ϕ (potential across sample) and V_a shown in Fig. 6 was obtained with the standard technique [7]. Satisfying the electric displacement boundary condition yields the bias across the sample (at zero applied bias) as being directly dependent on the work function of the tip (ϕ_m), less the ZnO electron affinity (X), and less the properly scaled surface charge density ($s\sigma/\epsilon_0$) on the sample (i.e., $\phi_m - X - (s\sigma/\epsilon_0)$), where ϵ_0 is the permittivity of free space. When ϕ_m is 5.65 eV (Pt work function) and σ is zero, the bands are bent up because of the 1.45 eV difference between the work function and the electron affinity. As a result, no current may flow from the tip to the sample until tip-induced inversion occurs. For the cases considered here, current flow would not have occurred within the experimentally probed range of applied positive biases. Because current was observed at much lower biases, one must assume that the tip work function is smaller (perhaps due to contamination) and/or there is a positive surface charge due to surface defects. Under the assumption that both factors contribute, ϕ_m is set nominally at 5.65 eV, and $s\sigma/\epsilon_0$ is set at the value required to simulate the data, then treated as an upper limit.

Earlier measurements of the work function of cleaved and annealed ZnO polar surfaces demonstrate that all of these surfaces possess some surface charge [3]. If $s\sigma/\epsilon_0$ is set to 1.45 eV and s is set to 8 Å, a positive surface charge density of approximately 10^{13} cm^{-2} is obtained. Considering that the density of surface unit cells on ZnO is approximately 10^{15} cm^{-2}, this result would indicate that 1% of the surface cells contain a positive charge. This charge, which could be caused by singly and doubly ionized Zn interstitial or oxygen vacancy defects, is not unreasonable considering the observed topographic structure of the surface. However, $s\sigma/\epsilon_0$ is treated only as an upper limit, any reduction in the tip work function would significantly lower the value of the surface charge density. For this reason; the absolute value of σ is considered less significant than the variation in σ.

The best way to obtain the most insight in the calculated tunneling spectra is to consider two critical points in the relation of ϕ and V_a. The first is where $\phi = 0$. This point defines the onset of accumulation and the point at which electrons from occupied sample states are easily detected. The second critical point is determined by the condition $\phi = V_a$, where electrons from the tip are easily detected. Using a donor concentration of 10^{17} cm^{-3} and a

sample–tip separation of 8 Å, ϕ is plotted against V_a for three different properly scaled surface charge densities, σ_s (defined as $s\sigma/\epsilon_0$), in Fig. 6a. For identification of the second critical point, a dotted line representing $\phi = V_a$ is also shown. For $\sigma_s = 1.50$ eV, $\phi(V_a)$ is always lower than $\phi = V_a$ for positive biases. This indicates that electrons are free to flow into the sample for all positive biases. For $\sigma_s = 1.00$ eV, $\phi(V_a)$ is always higher than $\phi = V_a$ for positive biases less than 3 V. This indicates that electrons are never able to freely flow into the sample (they must tunnel through the near surface region of the semiconductor). For $\sigma_s = 1.25$ eV, current does not appreciably flow into the sample until $V_a > 1.5$ V. In addition, this curve is approximately coincident with $\phi = V_a$ for a small voltage range. It is precisely this short period of coincidence that is providing the large 300 meV NDC feature when there is only a 10 meV gap.

While the surface charge density plays a critical role in the shape of the ϕ versus V_a curve, the donor concentration is also important [15]. Using $\sigma_s = 1.25$ eV and $s = 8$ Å, the effect of three different donor concentrations on ϕ versus V_a are shown in Fig. 6b. If the donor concentration is 10^{18} cm^{-3}, the slope is much less than unity and current will flow for small biases. The curve labeled 10^{17} cm^{-3} is the same as that labeled $\sigma_s = 1.25$ eV shown in Fig. 6a. For much lower donor concentrations, 10^{16} cm^{-3}, the curve has a slope very close to unity and, as a result, may be approximately coincident for a larger bias range than is the 10^{17} cm^{-3} curve. Perhaps the most important consequence of the voltage-dependent band bending is that the applied bias is not equivalent to the energy from the CB minimum in the sample. For example, if the slope of $\phi(V_a)$ is exactly unity, the tip would only sample one energy level in the crystal for all applied biases (until inversion). This coincidence period, which is responsible for the 300 meV width of the NDC (despite the 10 meV gap used in the simulation), is very sensitive to both the surface charge density (σ) and the local donor concentration (n_d).

Tunneling spectra calculated with the same parameters as those for Fig. 6a are shown in Fig. 7a. The two critical points on each ϕ versus V_a curve are easily identified as the current onset voltages at positive and negative biases. The increase in current

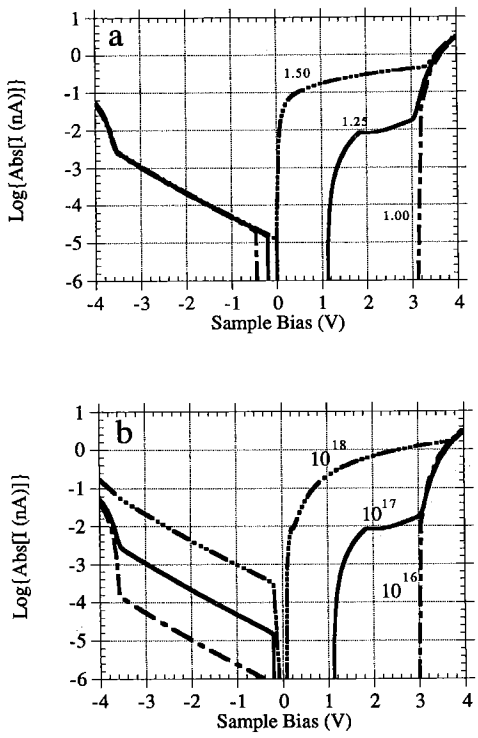

Fig. 7. Calculated current voltage curves. (a) Three spectra calculated with varying scaled surface charge density and labeled: $\sigma_s = 1.50$ eV, $\sigma_s = 1.25$ eV, and $\sigma_s = 1.00$ eV. (b) Three spectra calculated with varying donor concentration and labeled: $n_d = 10^{16}$ cm^{-3}, $n_d = 10^{17}$ cm^{-3}, and $n_d = 10^{18}$ cm^{-3}.

at $V_a = -3.5$ V is due to electrons tunneling from the valence band of ZnO to the tip. The increase in current at $V_a = 3.2$ V is related to the bend in the ϕ versus V_a curve (Fig. 6a) which is due to inversion. The NDC is clearly observed in the curve with $\sigma_s = 1.25$ eV (i.e., the seemingly flat response from 1.9 to 2.2 V), whereas it is not evident in the other two curves. When $\sigma_s = 1.50$ eV the coincidence is completely lacking between the ϕ versus V_a and the $\phi = V_a$ curves for positive biases. In the case where $\sigma_s = 1.00$ eV the coincidence is too abrupt and it occurred after the more dominant inversion.

Tunneling spectra calculated with the same parameters as those used in Fig. 6b are shown in Fig. 7b to examine the effect of varying carrier concentration. Again, the two critical points on each ϕ versus V_a curve are easily identified as the current onset voltages at positive and negative biases. The curve

labeled $n_d = 10^{17}$ cm^{-3} is the same as that labeled $\sigma_s = 1.25$ eV in Fig. 7a. For a higher donor concentration, $n_d = 10^{18}$ cm^{-3}, a short NDC feature occurs at $V_a = 0.2$ V, whereas for a lower concentration, $n_d \doteq 10^{16}$ cm^{-3}, the NDC feature does not exist. Again, the appearance of the NDC feature is related to the coincidence of the ϕ versus V_a and $\phi = V_a$ curves. The NDC features are sensitive to both the surface charge density and the donor concentration.

5. Discussion

Within the framework of the model used to calculate the tunneling current, differences among the observations necessarily imply variations in parameters such as the position of occupied states in the band-gap region or the surface charge density. The Zn-terminated surface is described first and then the O-terminated surface will be discussed.

5.1. Zn-terminated surface

The characteristic features of the spectra from the Zn-terminated surface, far from an edge of the mesa, are the absence of current below 1.6 V and the NDC feature centered at 2.0 V. Physical phenomena that can account for the observed NDC have been described by others [16,17]. One possibility is that there is a narrow band of electronic states separated by an energy gap from the bottom of the conduction band (i.e., a varying density of states (DOS) model). The simulated results shown in Fig. 3a were computed assuming that a 40 meV wide donor band exists 10 meV from the conduction band minimum, that there is a surface charge density of 8.6×10^{12} cm^{-2}, and that the sample–tip separation was 8 Å. The origin of this donor band is assumed to be a population of ionized and localized states associated with point defects. Note that the position of this band is consistent with the known energy levels of singly ionized oxygen vacancies or Zn interstitials [18].

The presence of this band produces the observed NDC for the following reason. At zero bias, both the donor band and the conduction band edge are bent above the Fermi level. As V_a increases, the Fermi level of the metal rises and, when it reaches the empty donor states, electrons begin to tunnel into the sample. Once the Fermi level of the tip is above the donor band, the current decreases because of the combined effects of the absence of new states into which to tunnel and the continued increase in the tunneling barrier. Finally, when the Fermi level reaches the conduction band edge, the current begins to increase again. As explained earlier, the expanded width of the NDC feature is a result of the tip-induced band bending; a 300 meV range of applied bias samples only 10 meV in the sample.

Although this overly simplified model correctly predicts the width of the NDC feature, it fails to predict the magnitude of current reduction. This is especially noteworthy considering that the transmission probabilities were set to maximize the decrease in current. The model also fails to predict the rapid increase in current following the minimum. Although parameters could be selected to reproduce the latter effect, they would not have been consistent with the parameters required to model the other spectra from this surface (Figs. 3b and 3d). One reason for these failures might lie in the properties of the donor band itself. Because it is a collection of states caused by randomly positioned defects with a range of energies, hopping among these states is likely to occur at a similar or slower rate than tunneling to the surface [19]. The total tunneling probability will, therefore, depend on how far into the crystal the electron must tunnel to find a defect state at a similar energy. Thus, the donor band states may populate more rapidly than they depopulate through a hopping mechanism, causing a larger decrease in current during the NDC period.

Continuing with the Zn-terminated surface data, the I–V spectra change systematically as the distance from the edge of the mesa decreases; a trend illustrated in Figs. 2c–2f. The most noticeable change is that as the edge of the mesa is approached, current begins to be detected in the negative bias region of the spectrum. Within this model, this implies that there are occupied states from which to tunnel and that the energies of these states rise as the edge of the mesa is approached. This situation was modeled by assuming the presence of occupied states above the valence band edge. For example, a calculation which assumes the presence of occupied states at 2.5 eV below the Fermi level, a surface charge density of 8.7×10^{12} cm^{-2}, and a sample–tip separation of

8 Å, reproduces the characteristic elements of the spectrum in Fig. 3b. For this example, the onset of current for both negative and positive biases are well-correlated with the data. In addition, the gap states at 2.5 eV simultaneously provide the proper current surge at negative biases and the inflection in the positive bias region due to inversion.

The origin of these states in the gap are thought to be point defects on the surface and in the near-surface region. Considering that surface edges are expected to act as sources of point defects, that are likely to be produced during the high temperature in vacu thermal treatments, one would expect a defect distribution near these edges which increases as the edge is approached. Because these states occur only in the vicinity of the edge of the mesas (which represents about 1% of the surface), it is reasonable to assume that they would not be detected by other spectroscopic techniques that average over the entire surface. Note an alternative model one could create, would be to use a bias-dependent surface charge density, such that, the bands are locally pinned in the upward direction and the increase in current at $V_a = -2.7$ V in Fig. 3b is due to the valence band. Even though either model could be used, the authors are using the local surface gap states model to explain the data throughout this paper.

There is a discontinuous change in the spectra when the tip reaches the edge of the mesa. A representative spectrum, shown in Fig. 3c, is characterized by a continuously increasing current in all regions of the spectrum. The topographic plot (Fig. 2b) indicates that the edge of the mesa is formed by an inclined facet that is not perpendicular to the (0001) surfaces above and below. Although it is not possible to specify the index of this facet (the inclination is almost certainly determined by a convolution of the tip and sample geometry), it is possible to say that the tip is probing something other than the Zn-terminated surface and it is interesting to note that this has the same characteristic features as are observed in spectra of the O-terminated surface (see, for example, Fig. 4a). The simulation in Fig. 4a uses a larger surface charge density than is needed for Fig. 3b. Thus, the I–V simulation requires a larger surface charge density near the edge of the mesa (Fig. 3c). This is evidence of patch charging on the surface, which manifests itself in the simulations as a change in the surface charge density from spot to spot [20].

Finally, the effect of O_2 adsorption on the tunneling spectra is considered. The characteristic features of the representative spectrum are the surge in current at -2.2 V and the inflection at 2.2 V. This spectrum was simulated using a model that was similar to that used to model the spectrum near the edge of the mesa. Specifically, occupied surface states were included 2.0 eV below the conduction band edge, the sample–tip separation was 9 Å, and the surface charge density was 7.8×10^{12} cm^{-2} (this implies a 0.8×10^{12} cm^{-2} decrease in surface charge density from the untreated surface). The presence of the gap states simultaneously creates the current onset in the negative bias region of the spectrum and the inflection in current in the positive bias region of the spectrum. One intuitive aspect of this model is that the reduction in the surface charge density (with respect to the clean surface) and the addition of states in the gap is consistent with models describing the adsorption of O_2 on ZnO [18].

It is noteworthy that the theory systematically predicts the occurrence of inversion at a lower bias than is observed experimentally. Feenstra and Stroscio [7] suggest that this phenomenon is caused by the use of Fermi–Dirac statistics to calculate ϕ versus V_a, which assumes that the system is in equilibrium. Inversion is predicted when the amount of band bending is equal to the energy of the gap state. However, for low carrier concentration and large band gap samples, the gap states would be many depletion widths ($\sqrt{[2\epsilon_s\phi_0/(n_d e^2)]} = 1000$'s Å) from the CB states at the same energy [21], where ϵ_s is the dielectric constant of the semiconductor, ϕ_0 is the amount of band bending, and n_d is the donor concentration. Under such conditions, the system is not in equilibrium and the point of inversion is shifted to higher biases (to a point at which the gap states are positioned only 10's Å from the CB states).

An additional factor in the case of ZnO is its piezoelectric response properties. As a portion of the total applied bias is dropped across the sample the crystal length will change. The effect of this strain on the I–V spectra is to continuously increase the sample–tip separation distance as the bias is ramped from 3 to -3 V. Using the d_{33} piezoelectric coefficient of 0.12 Å/V the effect is small, but its effect

does move the calculated I–V spectra closer to the experimental result.

5.2. O-terminated surface

The spectrum in Fig. 4a, which is representative of those recorded on the O-terminated surface, far from an edge of the mesa. It is assumed that the continuously increasing current in the positive bias region of the spectrum is caused by tunneling into the Zn-derived conduction band states, and that the continuously increasing, but slightly lower current at negative biases is caused by tunneling out of states populated by donor electrons also in Zn-derived states. The relatively high current levels at low positive biases suggest that the expected tip-induced band bending is reduced from that on the Zn surface. Thus, the characteristic features of the spectrum were simulated using a model that assumed a sample–tip separation of 6 Å and a surface charge density of 1.1×10^{13} cm^{-2}.

The representative spectrum from an edge of the mesa on the O-terminated surface, shown in Fig. 4b, has features reminiscent of the spectrum far from the edge of the mesa; however, there are two significant differences between these two spectra: the current levels in the positive bias region of the spectrum are lower and the sharp increase in current at -1.2 V is followed by an inflection at -2 V. This increase in the slope of the curve, followed by a decrease in slope at higher negative bias, is responsible for the peak in the differential conductance spectrum $(\mathrm{d}I/\mathrm{d}V/(I/V))$ that was identified in an earlier report [12]. It was demonstrated that this peak is spatially correlated with the edge of the mesa and suggested that it is due to electronic states associated with this geometric feature. Within the framework of the model used to simulate these spectra, the sharp increase in slope at -1.2 V is reproduced by assuming the presence of additional occupied states associated with the edge of the mesa (consider, for example, the sharp rise in current in the negative bias region of the spectrum from the O_2-treated Zn-terminated surface in Fig. 3d).

The most important difference between the spectrum from a flat region of the untreated surface (Fig. 4a) and the spectrum from a flat region of the surface after exposure to O_2 (Fig. 4c) is that the current is much lower (and has a higher onset voltage) in the positive bias region of the spectrum after exposure. This suggests that in this case, there is greater upward band bending, a phenomenon simulated by reducing the surface charge density (relative to the untreated surface). As can be seen in Fig. 3c, a calculation based on the assumption that the sample–tip separation is 6 Å and the surface charge density is 7.8×10^{12} cm^{-2} accounts well for the experimental observations (this implies a 3.2×10^{12} cm^{-2} decrease in the surface charge density from the untreated surface). Upward band bending simultaneously increases the biases (positive and negative) at which electrons tunnel to occupied and unoccupied band levels. Considering that a likely defect on the vacuum annealed O-terminated surface that could account for the surface charge density is an O vacancy, it is reasonable to assume that the interaction of these defects with adsorbed O_2 (an electrophilic species) would act to reduce the surface charge density. Furthermore, if the added negative charge on the surface reduced the local mobile charge concentration (nearer to the value on the Zn-terminated surface) it is possible that the NDC could reappear, which might account for the small NDC feature observed in Fig. 4c.

The simulations in Figs. 4a and 4c predict an exponential relationship between the tunneling current and the applied bias once accumulation is reached. The observed data, on the other hand, exhibit some curvature in this region. The absence of any curvature in the simulated data is, as mentioned earlier, a result of the approximation that during accumulation, electrons do not fill any states below the CB minimum. By this data, this seems to be an over-approximation. However, Feenstra and Stroscio showed that even more serious deviations occur if it is assumed that electrons fill all of the available states [7]. For the interpretation of these tunneling spectra, the most realistic model probably lies between these two approximations.

Another approximation leads to the systematic overestimation of the current in the positive bias region of the spectra in Figs. 4a and 4c. The positive bias current is caused by tunneling into the Zn-derived conduction band states. Because the Zn ions are below the surface, electrons tunnel through both the vacuum barrier and through a plane of negatively

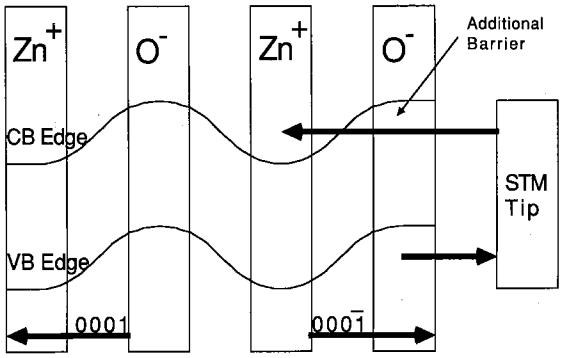

Fig. 8. Schematic diagram of valence band and conduction band edges along the *c*-axis for a ZnO sample. The additional barrier for tunneling to the Zn-derived states on the O surface is shown. No added barrier exists for tunneling to O-derived states on the O surface.

charged O-ions. Visualizing the crystal as an alternating plane of positive and negative charge, an oscillating potential exists (see Fig. 8) which defines the additional barrier through which electrons must tunnel to arrive at the plane just below the surface. Calculation of the transmission coefficient through this barrier is complicated by the complexity of the local potential (which is determined by both the layer of ions and the surface charge density) and that the WKB approximation is not valid for such an abrupt barrier. Although this effect was not included in the calculations, a qualitative consequence of this barrier is that it would reduce the current in the conduction band region of the spectrum and improve the agreement between the simulation and the experiment.

6. Conclusions

There are several features which characterize the electronic structures of the two distinct polar surfaces of ZnO. Spatially localized surface gap states on the Zn-terminated surface, are necessary to explain sudden current increases at negative biases. The local surface gap states simultaneously yield observed inflections in the positive bias region and the simulations show that the tip-induced inversion of the ZnO bands causes these inflections. In addition, the surface charge density appears to vary across the edge of the mesas (evidence for patch charging at the edge of the mesa). On both surfaces the influence of oxygen dosing was modeled by increasing the amount of negative charge on the surface. Finally, NDC is observed on the flat regions of the Zn surface (not spatially localized). Our model assumed that NDC was a result of the band bending that maintains coincidence with the Fermi level of the tip. The NDC was found to be sensitive to the local donor concentration and to the local surface charge density. The change in the magnitude of the current that occurs during the NDC period is, however, not accounted for with the variable DOS model. An electron-trapping model that may reduce the total number of final states appears necessary to explain the magnitude of change in current during the NDC period.

Appendix

The transmission coefficient through the semiconductor is given by:

$$D_{\substack{\mathrm{cb-S}\\ \mathrm{vb-S}}}(E_z) = \exp\left[-2\sqrt{2m^*_{\substack{\mathrm{cb}\\ \mathrm{vb}}}/\hbar^2}\int_0^{z_b} \mathrm{d}z\sqrt{E_{\substack{\mathrm{cb}\\ \mathrm{vb}}}(z)-E_z}\right],$$

and can be a very difficult to calculate. For the conduction band, one can write $E_{\mathrm{cb}}(z) = E_{\mathrm{cb}}(z=\infty) + \phi(z)$, where $\phi(z)$ must satisfy Poisson's equation ($\phi''(z) = \rho(z)/(e\epsilon_0)$, $\phi(0) = \phi$, $\phi'(\infty) = 0$). Described here is a simplified technique for calculating this transmission coefficient. Since $\rho(z)$ only depends on $\phi(z)$, one can transform Poisson's equation to: $\int \mathrm{d}\phi'\phi' = \int \mathrm{d}\phi\rho(\phi)/(e\epsilon_0) \equiv P(\phi)$. This can be transformed into the following: $\mathrm{d}z = \mathrm{d}\phi/[\pm\sqrt{2P(\phi)}]$. Using this to change variables, one obtains the transmission coefficient:

$$D_{\mathrm{cbs}}(E_z) = \exp\left\{\pm 2\sqrt{(2m^*_{\mathrm{cb}}/\hbar^2)} \times \int_{\phi(z=0)}^{\phi(z=z_b)} \mathrm{d}\phi\sqrt{\frac{[E_{\mathrm{cb}}(z=\infty)+\phi-E_z]}{2P(\phi)}}\right\}.$$

This can now be easily integrated with the limits of integration simply determined by observing that $\phi(z=0)$ is given in the ϕ versus V_a curve and that $\phi(z=z_b)$ is simply related to the tunneling energy.

Acknowledgments

The authors gratefully acknowledge J. Vohs both for providing the crystal used in this study and for many useful discussions regarding these experiments. In addition, the authors are grateful to R.M. Feenstra, E.J. Mele, and C.L. Kane for insightful discussions relating to the simulations presented in the text. This work was funded by IBM, the National Science Foundation under grant DMR 8819885, and the Department of Energy under grant DEFG02-90ER-45428.

References

[1] A.N. Mariano and R.E. Hanneman, J. Appl. Phys. 34 (1963) 384.
[2] G. Heiland and P. Kunstmann, Surf. Sci. 13 (1969) 72.
[3] H. Moormann, D. Kohl and G. Heiland, Surf. Sci. 80 (1979) 261.
[4] I. Ivanov and J. Pollmann, Phys. Rev. B 24 (1981) 7275.
[5] W. Göpel, J. Pollmann, I. Ivanov and B. Reihl, Phys. Rev. B 26 (1982) 3144.
[6] J. Bono and R.H. Good, Jr., Surf. Sci. 175 (1986) 415.
[7] R.M. Feenstra and J.A. Stroscio, J. Vac. Sci. Technol. B 5 (1987) 923.
[8] J.A. Stroscio and R.M. Feenstra, J. Vac. Sci. Technol. B 6 (1988) 1472.
[9] G. Heiland, E. Mollow and F. Stöckmann, Advances in Solid State Physics, Vol. 8 (Academic Press, New York, 1959) p. 19.
[10] W. Göpel and U. Lampe, Phys. Rev. B 22 (1980) 6447.
[11] H. van Hove and R. Leysen, Phys. Status Solidi (a) 9 (1972) 361.
[12] G.S. Rohrer and D.A. Bonnell, Surf. Sci. 247 (1991) L195.
[13] D.E. Eastment, in: CRC Handbook of Chemistry and Physics, Ed. R.C. Weast (CRC Press, Boca Raton, FL, 1988) E-78.
[14] E. Mollwo, in: Landolt–Börnstein Series, Group 3, Vol. 17b.
[15] R. Maboudian, K. Pond, V. Bressler-Hill, M. Wassermeier, P.M. Petroff, G.A.D. Briggs and W.H. Weinberg, Surf. Sci. 275 (1992) L662.
[16] M.G. Youngquist and J.D. Baldeschwieler, J. Vac. Sci. Technol. B 9 (1991) 1083.
[17] In-Whan Lyo and P. Avouris, Science 245 (1989) 1369.
[18] W. Göpel, J. Vac. Sci. Techol. 16 (1979) 1229.
[19] N.F. Mott and E.A. Davis, Electronic Processes in Non-Crystalline Materials (Clarendon Press, Oxford, 1971).
[20] N.A. Burnham, R.J. Colton and H.M. Pollock, Phys. Rev. Lett. 69 (1992) 144.
[21] S.M. Sze, Physics of Semiconductor Devices (Wiley-Interscience, New York, 1969).