IOPscience iopscience.iop.org

Home Search Collections Journals About Contact us My IOPscience

Properties of native ultrathin aluminium oxide tunnel barriers

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2003 J. Phys.: Condens. Matter 15 1733

(http://iopscience.iop.org/0953-8984/15/10/320)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 171.66.16.119
The article was downloaded on 19/05/2010 at 08:16

Please note that terms and conditions apply.

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **15** (2003) 1733–1746 PII: S0953-8984(03)57053-9

Properties of native ultrathin aluminium oxide tunnel barriers

K Gloos[1,2,4], P J Koppinen[1] and J P Pekola[1,3]

[1] Department of Physics, University of Jyväskylä, PO Box 35 (YFL), FIN-40014 Jyväskylä, Finland
[2] Nano-Science Centre, Niels Bohr Institute, Universitetsparken 5, DK-2100 Copenhagen, Denmark
[3] Low Temperature Laboratory, Helsinki University of Technology, PO Box 2200, FIN-02015 HUT, Finland

E-mail: gloos@fys.ku.dk

Received 3 December 2002
Published 3 March 2003
Online at stacks.iop.org/JPhysCM/15/1733

Abstract

We have investigated planar metal–insulator–metal tunnel junctions with aluminium oxide as the dielectricum. These oxide barriers were grown on an aluminium electrode in pure oxygen at room temperature till saturation. By applying the Simmons model we derived discrete widths of the tunnelling barrier, separated by $\Delta s \approx 0.38$ nm. This corresponds to the addition of single layers of oxygen atoms. The minimum thickness of $s_0 \approx 0.54$ nm is then due to a double layer of oxygen. We found a strong and systematic dependence of the barrier height on the barrier thickness. Breakdown fields up to 5 GV m^{-1} were reached. They decreased strongly with increasing barrier thickness. Electrical breakdown could be described by a metal–insulator like transition of the dielectric barrier due to the large density of tunnelling electrons.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Aluminum oxide (AlO_x) is one of the standard construction materials for tunnelling barriers because it can easily and reliably be formed. It is important for applications like single-electron transistors [1, 2] and Coulomb-blockade thermometers [3]. AlO_x is also a promising candidate to replace SiO_2 based gate dielectrics in miniaturized electronic circuits, see for example [4]. Although many experiments take advantage of AlO_x tunnel barriers, few reports focus on their intrinsic properties. The conventional point of view on these properties can be summarized as follows.

[4] Author to whom any correspondence should be addressed.

© 2003 IOP Publishing Ltd Printed in the UK

(i) Almost any conductance per area can be achieved by simply choosing the right oxidation pressure and time.
(ii) The typical height of the tunnel barrier is about 2 eV, but there exist experimental data that vary from below 0.1 up to 8.6 eV [5–8].
(iii) The dielectric constant ϵ of the AlO_x barrier is smaller than that of bulk Al_2O_3, which is around 4.5–8.9 at around 295 K [9]. Like the potential height, ϵ does not seem to be a well-defined property of AlO_x.
(iv) Similarly, the effective mass m of the tunnelling electrons is not well defined, since it may depend on the thickness of the barrier. Bandstructure calculations [10] gave $m/m_e \approx 0.4$ for bulk AlO_3, while ballistic electron emission spectroscopy on thin AlO_x layers showed $m/m_e \approx 0.75$ [11]. Thus one may expect the effective mass to approach that of the bulk metal electrodes the thinner the oxide layer.

Defects in a tunnelling barrier can create localized states that contribute to resonant and inelastic tunnelling [12–15]. At thick barriers, typically of width $s \gg 1$ nm, these two processes can dominate over direct tunnelling because of the strong exponential dependence of the tunnelling probability on the width of the barrier. These defects result in certain temperature dependences of the tunnel conductance, and noisy spectra with anomalies at the energy positions of the defect states. For thin barriers, however, direct tunnelling should dominate.

In view of the importance of these tunnel barriers we have investigated them in more detail. We found a systematic variation in the properties of the samples which could hint at a strong dependence of the height of ultrathin AlO_x barriers on the width s. Our interpretation is supported by resolving the layer-by-layer thickness of the saturated oxide barrier. At the same time, the electrical breakdown fields are huge. They decrease strongly for thicker barriers. We discuss several mechanisms to explain the breakdown, and find a metal–insulator like transition of the dielectric barrier as the most plausible one.

2. Theory

The most simple picture of a tunnel junction assumes an s wide rectangular barrier of height Φ_0 across which electrons tunnel, see figure 1. In a Wentzel–Kramers–Brillouin (WKB) approximation the conductance per area at zero bias and at low temperatures becomes [16]

$$g_0 = \frac{e^2\sqrt{2m\Phi_0}}{h^2 s}\exp\left(-\frac{2s}{\hbar}\sqrt{2m\Phi_0}\right). \quad (1)$$

The first order correction increases the conductance quadratically with bias voltage V as [16]

$$g(V) = g_0(1 + V^2/V_0^2) \quad (2)$$

with $V_0^2 = (4\hbar^2/e^2m)\Phi_0/s^2$. Equations (1) and (2) can be used to estimate s and Φ_0 of the junction.

In general, image forces act on the electrons as they move through the barrier from one metal electrode to the other. They distort the potential distribution of real junctions, as described by the standard Simmons model [16]. These image forces reduce both the height of the barrier as well as its effective thickness, depending on the dielectric constant ϵ. Including the bias voltage V, the full potential becomes [16]

$$\phi(x) \approx \phi_0 - eV\frac{x}{s} - 1.15\lambda\frac{s^2}{x(s-x)} \quad (3)$$

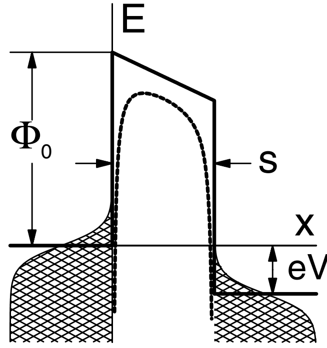

Figure 1. Rectangular tunnel barrier of width s and height Φ_0 after applying a bias voltage eV. The dashed curve shows how the barrier changes due to image forces. The shaded areas represent populated electron states at finite temperature.

with $\lambda = e^2 \ln(2)/(8\pi\epsilon\epsilon_0 s)$. The electrical tunnel current density can then be calculated in the WKB approximation as [16]

$$J = \frac{me}{2\pi^2\hbar^3} \int D(E_x, eV) \int [f(E) - f(E + eV)] \, dE_r \, dE_x \tag{4}$$

with the transmission probability

$$D(E_x, eV) = \exp\left(-\frac{\sqrt{8m}}{\hbar} \int \sqrt{\phi(x) - E_x} \, dx\right). \tag{5}$$

Here the kinetic energy $E = E_x + E_r$ of the electrons has been split into E_x and E_r due to the velocity components pointing towards and parallel to the barrier, respectively. In equation (4), $f(E)$ is the Fermi–Dirac distribution function at temperature T, and only the region of positive $\phi(x)$ contributes to the integral of equation (5). The barrier height and its width determine then, together with the dielectric constant ϵ and the effective electron mass m, the zero-bias conductance g_0 and the slope parameter V_0. Zero-bias anomalies (ZBAs) omitted so far are discussed below.

To apply the Simmons model requires that direct tunnelling dominates. Otherwise the derived junction parameters are not directly related to the height and the width of the barrier but to the tunnelling states created by defects or impurities inside the barrier.

3. Experimental details

We fabricated 99 samples on an oxidized Si substrate, using the standard two-angle shadow evaporation technique, see figure 2(a), to deposit the metal electrodes in an ultra-high vacuum (UHV) chamber at a rate of 0.3 nm s^{-1}. The samples had $N = 12$ nominally identical junctions in series, each of $A = 2 \times 2$ μm^2 cross section. Electrodes were evaporated through a 300 nm thick Si_3N_4 mask as shown in figure 2(b). A 25 μm diameter Al wire separated the mask from the substrate. By using a metal wire as a spacer we avoid organic material near the sample which could otherwise contaminate the oxide layer or the electrodes. During evaporation the temperature of the sample was not controlled. However, we believe that most of the heat from the hot metal beam was caught not by the Si substrate but by the Si_3N_4 mask in front of it. In addition, the temperature of the mask increased by not more than $\sim$100 K. The first Al layer, either pure Al or its alloys with Cu, Au or Nb, was exposed to pure oxygen

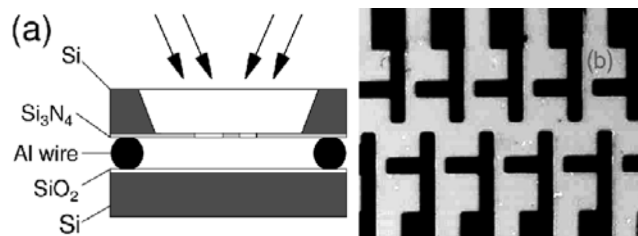

Figure 2. (a) The setup used to evaporate the metallic electrodes through a thin Si_3Ni_4 mask onto an oxidized Si substrate (schematic not to scale). Arrows mark the two different evaporation angles before and after oxidation. (b) Microscope view of a typical Si_3Ni_4 mask. The dark regions have been etched away. After evaporating the first metal layer and oxidizing it, the sample is rotated vertically to evaporate the second layer so that the 2 μm wide opposing horizontal and vertical fingers just overlap.

to form a thin insulating AlO_x barrier. For this about 5×10^4 Pa oxygen pressure was applied for 30 min to the sample in the loading chamber of the UHV system, while both sample and oxygen were kept at ambient temperature (∼295 K). Then the second electrode (Al, Cu, Nb, Au, or an alloy between Al and one of the other metals) was evaporated.

To prepare the alloys we first evaporate the Cu, Nb or Au layer. Immediately afterwards an Al layer was added at the same evaporation angle. Oxidation in the loading chamber started about 1 min after finishing the Al layer. In this way the metal surface to be oxidized consisted mainly of Al, and oxidation should result in AlO_x. However, the actual amount of impurity (Cu, Nb or Au) atoms at the Al surface due to diffusion is not known. We cannot even estimate it, since we had no access to the temperature of the electrodes during the evaporation process. Complete interdiffusion of the metals can be expected only after some long time interval. For the second electrode Al was evaporated first, and then covered by the other metal.

The total thickness of each electrode was about 90 nm. We have found no systematic dependence of the final tunnel conductance per area on oxygen pressure, oxidation time, evaporation rate, junction area, or thickness of the metal electrodes when the relevant parameters were varied by a factor of two. In this respect the experiments are quite reproducible and represent the saturation thickness of the barrier at 295 K in pure oxygen.

Both the size of each junction and the number of junctions are a compromise: the area is large enough to be reliably measured, but sufficiently small to show a resolvable Coulomb blockade signal at 4.2 K when the sample is in the liquid helium of a transport dewar. We have chosen this condition to make a reliable comparison and to minimize the turn-around time. The number of junctions N still allows sufficiently large voltages to be applied across each junction using conventional electronic equipment. But simultaneously it reduces the risk of external high voltage spikes destroying the junctions.

From the sheet resistance of the shorted junctions, obtained either by omitting the oxidation process or when the tunnel barriers have been short-circuited by a large bias voltage, we estimate a residual electrical resistivity of the evaporated metals of 4–10 $\mu\Omega$ cm at 4.2 K and a resistance ratio between 295 and 4.2 K of around 1.5, indicating good quality of the electrodes, both for the pure metals and the alloys.

The differential conductance $\mathrm{d}I/\mathrm{d}V$ was recorded by applying a small low-frequency (∼100 Hz) ac voltage $\mathrm{d}V$ superposed on the bias voltage V to the junction and detecting the current modulation $\mathrm{d}I$ with the help of a lock-in amplifier. Results are always presented as the voltage drop per junction and the conductance per area A of the junctions $g = (1/A)\,\mathrm{d}I/\mathrm{d}V$.

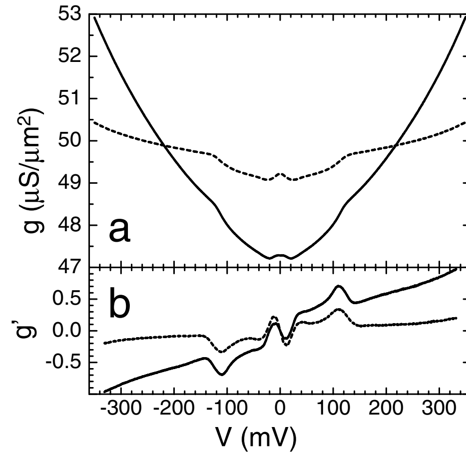

Figure 3. (a) Differential conductance $g = (1/A)\,\mathrm{d}I/\mathrm{d}V$ versus voltage drop V for two typical tunnel junctions at $T = 4.2$ K. Solid curve, Al–oxide–Cu; dashed curve, Al–oxide–Au (not symmetrized original data). (b) Second derivative $g' = \mathrm{d}g/\mathrm{d}V$ of the two spectra in arbitrary units.

4. Results: barrier width and height

Figure 3 shows typical $\mathrm{d}I(V)/\mathrm{d}V$ spectra of an Al–oxide–Cu and an Al–oxide–Au tunnel contact at $T = 4.2$ K. Both have the same characteristic ZBAs. The essential difference between them is the strength of the V^2-dependence at larger voltages. It is basically this difference which indicates the thinner (Au contact) and the thicker (Cu contact) barrier.

Since most of our junctions have normal electrodes (at $T = 4.2$ K), we rely on several observations to ensure quantum tunnelling.

(i) The typically ~15% reduction of the conductance on cooling from 295 to 4.2 K [17], which has been proposed recently as a good tunnelling indicator [18].

(ii) The V^2-dependence of the conductance. This increase in conductance with increasing bias voltage is typical for tunnel junctions. We have observed the transition from tunnelling to metallic-like behaviour when the zero-bias conductance drops below about $g_{0,c} \approx 5$ mS μm^{-2}.

(iii) Several ZBAs appear at lower voltages. They can be attributed to inelastic electron–phonon scattering in the barrier [8]. We believe that the well-pronounced characteristic 110 meV anomaly, which we have regularly observed at thin barriers with $s \leqslant 1.0$ nm, indicates high quality, defect free tunnel junctions.

(iv) Several Al–oxide–Nb junctions, prepared at the same conditions as the other samples, showed the expected superconducting anomalies if finite lifetime broadening is taken into account.

(v) The known shape of the Coulomb blockade spectrum, see below.

The partial Coulomb blockade at $T = 4.2$ K can be resolved as long as the background ZBA has a conductance maximum at $V = 0$. Figure 4 shows that the spectra normalized to the smoothed background fit well to the expected shape due to the Coulomb blockade. This also indicates that the $N = 12$ junctions are rather identical. In both cases the width of the anomaly [3] $V_{1/2} = 5.44\,k_BT/e$ derived from the fit curve corresponds to a nominal temperature of 4.4(3) K, in good agreement with the experimental 4.2 K. Because of the large

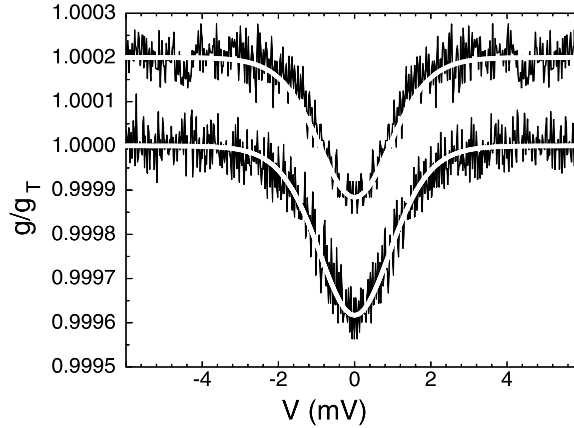

Figure 4. Coulomb-blockade anomaly around zero bias of the two tunnel junctions of figure 3. The conductance g has been normalized to the smooth background conductance g_T. Solid white curves are theoretical predictions. From the size of the anomalies we estimate junction capacitances of about 0.22 and 0.18 pF and dielectric constants of $\epsilon \approx 5.6$ and 3.1 for the Al–oxide–Cu (top, vertically displaced) and the Al–oxide–Au (bottom) sample, respectively.

contact area the Coulomb blockade anomaly is quite small. Nevertheless, its relative size can be used to derive the capacitance of the junctions C [3]. From this we estimate an average dielectric constant of $\epsilon \approx 4$ for the parallel-plate geometry with $C = \epsilon\epsilon_0 A/s$.

Surprisingly, most junctions with thin barriers had a quite symmetric bias-dependent conductance, like that in figure 3, even when the electrodes are of different metals. This implies rather symmetric tunnel barriers. On the other hand, even symmetric Al–AlO_x–Al tunnel junctions have been reported to have asymmetric spectra [6].

We always symmetrized the spectra to obtain the average barrier heights. Plotting the differential conductance versus V^2 reveals a linear relationship above the 110 meV anomaly from which V_0 can be extracted (the Al–oxide–Au samples had a very small V^2 contribution, making the analysis more difficult). Fitting the spectra above the ZBA using the Simmons model and the average $\epsilon = 4$ allows us to determine s and Φ_0. This fit was restricted to bias voltages below $|V| \leqslant 0.3$ V. Usually the fit parameters are good enough to describe the $I(V)$ characteristic up to the breakdown voltage, see figure 9 below. This also confirms that in the low-voltage range electrons can be described by a single effective mass. The final result depends only weakly on the absolute value of ϵ if this is varied by $\pm 50\%$, and the complete analysis indicates a temperature dependence of ϵ and/or Φ_0. For this analysis we have also assumed that the tunnelling electrons have the bare electron mass $m = m_e$ inside the barrier. In the worst case scenario, $m = 0.4\ m_e$, according to band-structure calculations for bulk AlO_3 [10], would result in slightly (about 5%) smaller barrier heights and about 25% larger widths. Thin AlO_x barriers have $m \approx 0.75\ m_e$ [11], so that these corrections can be safely neglected.

The asymmetry $\Delta\Phi$ of the barriers can now be estimated using the Simmons model as generalized by Brinkman *et al* [19]. In this model the conductance minimum is not at zero bias but at

$$V_{min} \approx 65\big(\text{mV nm eV}^{-1/2}\big)\Delta\Phi/s\sqrt{\Phi_0}. \tag{6}$$

Our experimental data typically have $V_{min} \leqslant 5$ mV. Combining this with the experimental s and Φ_0 the asymmetry amounts to not more than $\Delta\Phi \approx 0.1$–0.2 eV. Thus the relative asymmetry is small, at least at large Φ_0.

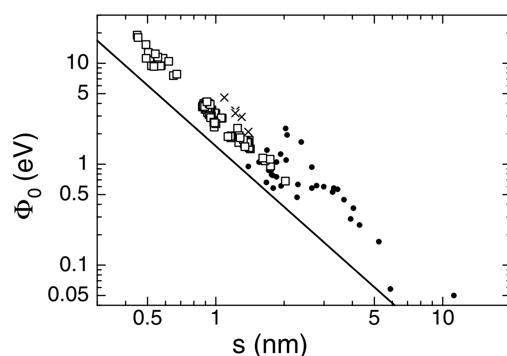

Figure 5. Barrier height Φ_0 versus width s of AlO_x tunnelling barriers. The solid line represents equation (7). Solid circles are the data of [5] for sputter-deposited Al_2O_3 barriers. Crosses are averaged data from [6] for thermally or plasma-discharge oxidized Al.

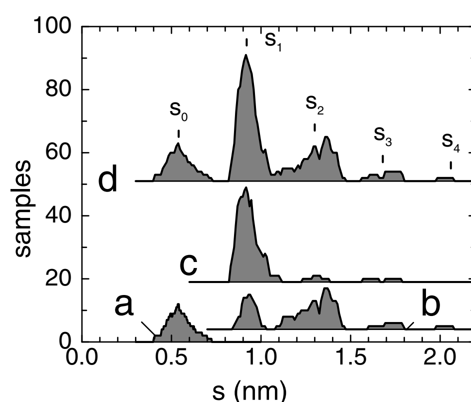

Figure 6. Histogram of the number of samples falling inside a 0.1 nm wide interval around a barrier thickness s. A total of 99 samples have been measured. Trace (a) represents the Al–oxide–Au samples, trace (b) belongs to (M, Al)–oxide–M and (M, Al)–oxide–(Al, M) samples (M = Al, Cu, Au or Nb), and trace (c) to the Al–oxide–M samples. Trace (d) is the sum of all samples. Traces (b), (c) and (d) are vertically displaced. The position of the peaks marked s_0, s_1, s_2, s_3, s_4 are equally spaced with $\Delta s = 0.38$ nm.

Figure 5 summarizes our main findings. First, the barrier height depends very strongly on the thickness s of the barrier like $\Phi_0 \approx 2.5$ eV s^{-2}(nm). These data clearly demonstrate that Φ_0 is neither related to the band gap of bulk Al_2O_3 nor to the work function of the different metals. Second, most surprisingly data points accumulate at certain values of the thickness s_i ($i = 0, 1, 2, 3$ and 4), following a regular pattern with a $\Delta s \approx 0.38(5)$ nm spacing as shown in the histogram in figure 6. Such a spacing could be expected for a homogeneously grown oxide layer, built up basically by adding one layer of oxygen atoms to an already existing one (the Al atoms are much smaller than the O atoms). The smallest observed thickness $s_0 \approx 0.54(5)$ nm corresponds then to two oxygen layers as proposed theoretically in [20]. This minimum thickness is also consistent with recent observations of stable 0.6 nm thick Al_2O_3 films grown on a Si substrate [4] and 0.59 nm Al_2O_3 films on Ta [21].

Note that only Al–oxide–Au junctions were of s_0 type, all other samples with pure Al base electrodes, like Al–oxide–Al and Al–oxide–Cu, were mainly of s_1 type, as shown in figure 6. Samples with Al alloys as a base electrode, like $Cu_{0.5}Al_{0.5}$, were mostly of s_2 type,

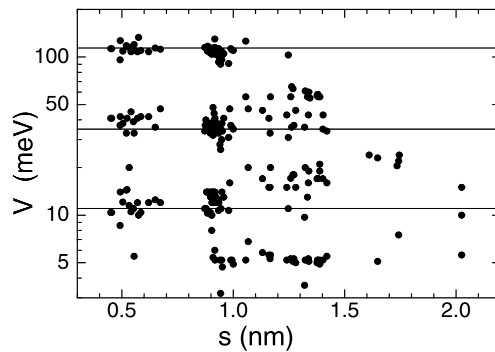

Figure 7. Position of the ZBAs V versus width s of AlO_x tunnelling barriers. Horizontal lines are guide to the eyes.

but a considerable fraction of them were of s_1 and s_3 type. Obviously, thicker barriers are more easily formed on Al alloys with their degraded (irregular) surface structure than on pure Al. On the other hand, the preference of the Au samples to form very thin barriers could result from the fact that Au is difficult to oxidize, unlike Al or Cu. After oxidation, the top-most layer is probably chemisorbed oxygen. When evaporating Al or Cu as the second electrode, these oxygen atoms form bonds with the metals, but desorb easily when Au is evaporated.

The narrow s_0 and s_1 peaks in figure 6 directly demonstrate that the thin barriers of samples with a pure Al base electrode are quite homogeneous. The s_2 peak, on the other hand, is considerably broader. This indicates then less homogeneous thick barriers of the Al alloy samples, which could even result in samples with an average thickness between s_1 and s_2.

We emphasize here that the observed discreteness of the barrier width s with reasonable absolute values strongly supports our method of deriving the width. In addition, a correct width s can only result when the barrier height Φ_0 has also been determined correctly.

The thin oxide layers with s_0 and s_1, in particular, have a well-pronounced 110 meV ZBA as shown in figure 3. This anomaly characterizes the surface (or longitudinal) phonons of crystalline Al_2O_3 (sapphire), see for example [22]. Thus in these samples the oxygen ions are fully bound in a regular crystal lattice and not only chemisorbed. Chemisorption may be the case for the thicker barriers. For them the 110 meV anomaly is strongly suppressed and an anomaly at around 50 meV appears. The position of the various ZBAs is summarized in figure 7. A second reason for this suppression of the 110 meV anomaly could be inhomogeneities in the barrier. This is consistent with observing slightly noisier spectra at those junctions, that could indicate defects in the barrier. With those defects, the barrier appears to be thicker on applying the Simmons model. Consequently, the apparent barrier height would be reduced. However, those samples with the thicker (s_2, s_3, and s_4) barriers show the same relative change of zero-bias conductance as the samples with thinner (s_0 and s_1) barriers when they are cooled from room temperature to 4.2 K. This points to direct tunnelling as the dominating process for charge transport even across those thicker barriers.

Like the potential height, the position of the ZBAs depends systematically on the thickness s. The histograms in figure 8 show that each thickness s_i has a preferred set of ZBAs.

Several other experiments either contradict or support our findings. Lau and Coleman [8] investigated metal–oxide–metal tunnel junctions with thermally and plasma-discharge grown AlO_x barriers while Barner and Ruggiero [7] sputter-deposited Al_2O_3 tunnel barriers. Figure 5 shows that in both cases the absolute values as well as the thickness dependence of the barrier

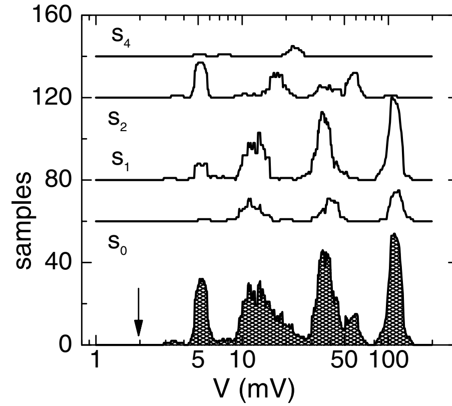

Figure 8. Histogram of the number of samples with ZBAs falling inside a $\pm 10\%$ wide interval around a voltage V. Different traces are separated as a function of barrier thickness and independent of the electrode material. The arrow marks the half width $V_{1/2}$ of the Coulomb blockade anomaly at 4.2 K.

height $\Phi_0(s)$ agree well with our data. In particular the latter report extends our data set to significantly larger s. Note that those data have been analysed like ours using Simmons model.

However, experiments on metal–oxide–semiconductor junctions gave completely different results: ballistic electron emission spectroscopy on AlO_x found 3.90(3) eV high barriers for 8 nm oxide thickness [23] and 1.2 eV for 0.6–1.5 nm oxide thickness [11], respectively. Internal photoemission gave $\Phi_0 = 3.25(8)$ eV for 5–15 nm wide barriers [27]. According to [23] there is almost no difference in the barrier height when AlO_x is replaced by SiO_2. But according to [27] the numbers are different, namely 4.25(5) and 3.25(8) eV for SiO_2 and AlO_x, respectively. Only our s_1-type junctions have barriers of this magnitude.

The different results for the different kinds of setups could indicate that boundary conditions matter. The experimental $\Phi_0(s)$ of our samples in figure 5 marks roughly the kinetic energy

$$E = \frac{h^2}{2ms^2} \approx \frac{1.5\ \text{eV}}{s^2\ (\text{nm})} \tag{7}$$

at which the de Broglie wavelength of the electrons becomes smaller than the width s of the barriers. One could then speculate whether the energy Φ_0 is simply determined by an electronic resonance phenomenon inside the barrier.

At a perfect metal–oxide interface the energy bands align far away from the interface according to the work function of the metal, the electronegativity of the oxide and excess charges in the oxide or at the metal–oxide interface. In the ideal case, Al–Al_2O_3–Al tunnel junctions have typical barriers of a few electronvolts, like our samples with widths $s = s_1$ and s_2. The most obvious way to change the barrier height is, then, to have excess charge carriers inside the barrier, as discussed for example in [24, 25].

The maximum possible barrier height is limited by the band gap of AlO_x. If we assume the same band gap as that of crystalline Al_2O_3 [26], then an upper bound of about 9 eV results. This agrees fairly well with the average Φ_0 of the $s = s_0$ thin tunnel barriers. Such a strongly enhanced barrier could be caused by excess cations in the oxide layer, like Al atoms not bound into the crystal lattice. However, to raise the valence band by several electronvolts would require that basically all Al atoms of the oxide are ionized. In addition, mutual stress between the oxide lattice and the metal host is strongest at the thinnest barriers. This stress could change

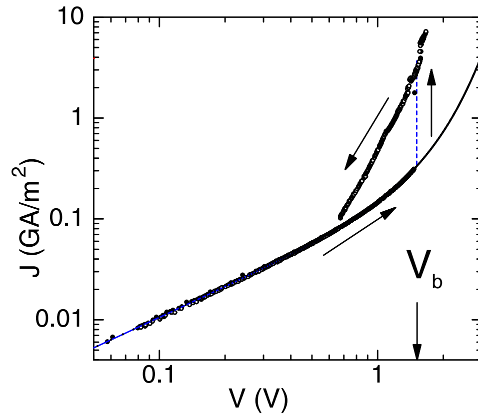

Figure 9. Current density J versus bias voltage V of a Al–AlO_x–Al tunnel junction. The solid curve represents the Simmons model equation (4) with barrier height $\Phi_0 = 3.52$ eV and width $s = 0.912$ nm. On increasing the bias voltage the current density jumps, as indicated by the dashed line, from the lower to the upper branch, defining the breakdown voltage V_b. This was one of the few samples that survived the breakdown, recovering the lower branch again when the bias voltage was lowered.

the lattice parameters and, thus, possibly increase the band gap beyond the 9 eV value. But our samples with thin $s = s_0$ barriers also show the well-pronounced 110 meV phonon anomaly like bulk Al_2O_3. Therefore any effect from lattice distortion should be small.

Likewise, the tunnel barrier could be lowered by negative excess charges, unbound (or weakly bound) oxygen ions. The suppression of the 110 meV phonon anomaly at tunnel junctions with the thicker $s = s_3$ and s_4 widths already indicates that the crystal lattice softens, although it gives no information on the sign of the excess charge.

The systematic $\Phi_0(s)$ dependence leads to a preferred transparency of the tunnel junctions for each saturation thickness which, in this series of experiments, covered roughly two decades in conductance per area. A wider range of transparencies was difficult to achieve by just varying the thickness of the barriers.

5. Breakdown field

We have tried to verify independently the potential height by measuring the current–voltage characteristics to higher voltages. Figure 9 shows that at the breakdown voltage V_b the tunnel current rises dramatically. Usually the junctions are destroyed in this event, resulting in either a short ($\sim$75% of the junctions) or an open ($\sim$20%) circuit. A short circuit means that at least one direct metallic path through the barrier has been created. In this case the current is limited by the lead resistance of $\sim$20 Ω/junction. An open circuit could indicate a broken lead since the current density through the leads is about 20 times larger than that through the junctions. Few ($\sim$5%) of the junctions survived, as shown in figure 9. Its $I(V)$ could be described as if there are two different current-carrying states. While at low voltages direct tunnelling dominates, the second branch with a larger current density becomes favourable at around V_b. The current density of the second branch is probably limited by the lead resistance, so that the intrinsic change of conductance due to the breakdown could not be measured. The transition between both seems to be non-continuous, and may be triggered by external noise. On lowering the voltage, the second state is conserved at first, until the original $I(V)$ branch is reached again.

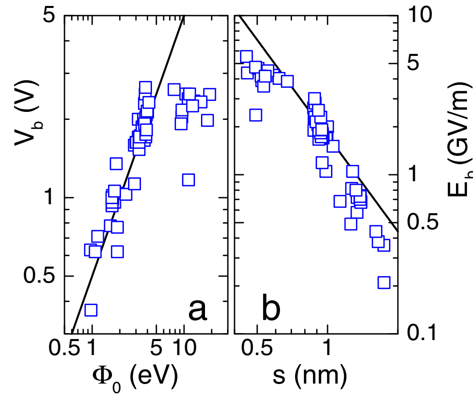

Figure 10. (a) Breakdown voltage V_b versus barrier height Φ_0 and (b) breakdown field E_b versus barrier thickness s. The solid lines are $V_b = 0.5\,\Phi_0/e$ and the metal–insulator critical field discussed in the text $E_c = 1.6$ (GV m^{-1})$/s^2$ (nm), respectively.

Figure 10(a) summarizes the experimental breakdown voltages V_b. They are smaller than the ones expected for field emission. Up to about $\Phi_0 = 4$ eV these breakdown voltages are closely correlated to the potential height, again supporting the magnitude of Φ_0 derived from the spectra at low bias voltages. The larger Φ_0 above 4 eV requires additional considerations because of the obvious trend of V_b to saturate at around 2.5 V (the Al–oxide–Au samples had only slightly, but systematically, larger V_b than the Al–oxide–Cu or Al–oxide–Al samples). Either these large barrier heights are completely wrong or we face a new effect. The latter could be the case because of the huge electrical breakdown fields $E_b = V_b/s$, approaching $\sim$5 GV m^{-1} at small s, see figure 10(b). Compared to this, typical literature data for AlO_x junctions have much lower breakdown fields, $E_b \approx 0.1$–1 GV m^{-1} at most [28]. A breakdown field which decreases with increasing gap between the electrodes is a well known, yet not fully understood, phenomenon in vacuum high-voltage insulation [29]. The 2.5 V upper bound of V_b could be readily explained if there was local heating or by the damage done due to the mechanical stress exerted by the electrical field.

The effects of local heating are difficult to evaluate. Although the dissipated power per junction area $U_b I_b/A \approx 0.2$–2 GW m^{-2} is huge, the breakdown does not depend systematically on the power, making local heating as the limiting mechanism unlikely. Even though the samples are in liquid helium at 4.2 K, at V_b their local temperature is certainly higher. The enhanced temperature enhances the tunnel current because the Fermi–Dirac distribution of the electrons in the metal electrodes broadens (the dielectric constant of the barrier and the barrier height probably also depend on temperature). This temperature dependence would enhance the current. But this is not supported by our experiments.

The mechanical stress at the barrier $\sigma = \epsilon\epsilon_0 E^2/8\pi$ [30] reaches 30 MPa at the smallest widths. Because of this mechanical stress, a defect inside the barrier could start to move, enhancing the field locally. In a chain reaction, neighbouring defects could then also move. This would explain the sudden rise of current. But bulk Al_2O_3 has a much larger yield strength of around 4 GPa [31]. The actual yield strength of our AlO_x barriers and their interfaces with the metal electrodes could be lower, depending on their quality. We believe that our barriers have a rather high quality, as indicated by the large E_b and also by observing the inelastic phonon modes of the barrier. The latter are suppressed when s increases, possibly indicating a lower quality caused by weaker Al–O bonds. This would coincide with the reduced E_b. In

this case, however, we would expect a larger scattering of the $E_b(s)$ data and not the observed systematic behaviour. Thus mechanical stress also fails to explain the reduced current density below V_b.

We have observed the transition from tunnelling to metallic-like behaviour at zero-bias conductances of around $g_{0,c} \approx 5$ mS μm^{-2}. Conventionally, this metallic behaviour is explained by pinholes in the barrier, leading to a direct metallic path between the metal electrodes. But there could be an alternative explanation. A closer look at equation (4) reveals that this equation describes only the net current density. Even at zero bias, current flows from one electrode to the other only to be exactly cancelled by the current flowing in the reverse direction. Therefore the net density of tunnelling electrons consist of the sum of the densities for the two opposite flow directions which can be quite large at small conductances. The zero-bias current density in one direction is [16]

$$J_0 = \frac{\hbar}{es}\sqrt{\frac{\Phi_0}{2m}}g_0. \quad (8)$$

Including image forces requires knowledge of the width and the barrier height, but their main effect is already contained in g_0. The density of tunnelling electrons $2J_0/ev_F$ can be quite large at large g_0. Our samples have a critical current density of (within an order of magnitude) $\sim$1 mA μm^{-2}. This, in turn, sets the critical zero-bias conductance to $g_{0,c} \approx 2$ mS μm^{-2}, coinciding well with our experimental data. Although we cannot independently verify that those metallic junctions have good tunnel barriers free of pinholes, it is quite plausible that in these cases the barrier is suppressed due to the large number of tunnelling electrons.

The electron density n is directly related to the tunnel current density $J = env$. In the metal electrodes electrons move at the Fermi velocity $v = v_F \approx 1.5 \times 10^6$ m s^{-1} which also defines the electron velocity in the thin barrier. The experimental critical current density results then in a critical electron density of $n_c \approx 4 \times 10^{21}$ m^{-3}.

Metal–insulator transitions are a well known phenomenon in the solid state [32]. They are usually observed in disordered systems with a low density of charge carriers like Ge:As or Si:P. There the (As or P) impurities in their (Ge or Si) host material either provide electrons as charge carriers or the empty states into which these electrons can hop. The spatial extension of their wavefunction or the localization radius a fixes the critical density n_c of the transition. The Mott criterium states [32] $n_c \approx (0.25/a)^3$. Typical experimental n_c range from $\sim 10^{20}$ to 10^{28} m^{-3} [33]. This comprises our above estimate.

Doping is the standard method of varying the electron density n and, thus, to drive an insulator or an insulating semiconductor into a metallic state, both in bulk samples and in thin films [32]. In a tunnelling barrier, defects can create localized states that contribute to resonant and inelastic tunnelling [14, 15]. At thick barriers these two processes can dominate over direct tunnelling because of the strong exponential dependence of the tunnelling probability on the width s of the barrier (thick means typically $s \gg 1$ nm). By increasing the number of defects it should be possible to drive an insulating barrier into a metallic state, although no such effects have been reported.

In thin films the situation is different because direct tunnelling should dominate. The field effect has been used to alter electrostatically the density of charge carriers in a thin semiconducting layer [32]. It is then possible to increase the carrier concentration sufficiently to induce a metal–insulator transition, for example in indium oxide [34], which can be recorded by the transport properties in the plane of the layer, perpendicular to the applied field.

Using the standard Mott formalism [32], the critical conductivity at n_c amounts to about $\sigma_c \approx 10^{-2}\, e^2/\hbar a$. In conjunction with the Mott criterium the critical field of the metal–insulator

transition becomes then

$$E_c \approx 1.76\hbar v_F / e a^2. \tag{9}$$

Once E_c is exceeded, the tunnel conductance should rise dramatically. This identifies E_c as the breakdown field E_b. Within the dielectric barrier electrons are localized to within the barrier width s which is therefore a natural substitute for the localization length a. The final result $E_c \approx 1.6$ (GV m^{-1})$/s^2$ (nm) fits our experimental E_b in figure 10 very well.

An additional argument in favour of a metal–insulator transition might be the switching behaviour of the $I(V)$ characteristics of the few junctions that survived the breakdown. This switching between two $I(V)$ branches is similar to that found for chalcogenide alloy glass or VO_2, as reviewed and discussed by Mott and Davis [32].

6. Summary

We have resolved the discrete saturation thickness of native AlO_x barriers. We also found a strong dependence of the barrier height on the width of the barrier which could be explained by positive or negative excess charges in the oxide. Otherwise the possible influence of defects inside the barrier on our results can be discarded for several reasons: the relative change of zero-bias conductance on cooling down from room temperature is practically independent of barrier thickness, the breakdown fields are huge, and the barrier thickness has discrete values. The latter fact confirms the excellent quality of our tunnel barriers with a very small number of defects that could act as weak spots. The mechanism for electrical breakdown could be a metal–insulator transition of the dielectric barrier due to the large density of tunnelling electrons.

Acknowledgments

We thank K Hansen, N Kim and K Likharev for discussions. This work has been supported by the Academy of Finland under the Finnish Centre of Excellence Programme 2000–2005 (Project No 44875, Nuclear and Condensed Matter Programme at JYFL) and by the European Commission under the Measurement and Testing activity of the Competitive and Sustainable Growth Programme (contract G6RD-CT-1999-00119).

References

[1] Averin D V and Likharev K K 1991 *Mesoscopic Phenomena in Solids* ed B L Altshuler, P A Lee and R A Webb (Amsterdam: Elsevier) p 173
[2] Grabert H and Devoret M H (ed) 1992 *Single Charge Tunneling* (New York: Plenum)
[3] Pekola J P, Hirvi K P, Kauppinen J P and Paalanen M A 1994 *Phys. Rev. Lett.* **73** 2903
[4] Kundu M, Miyata N and Ichikawa M 2001 *Appl. Phys. Lett.* **78** 1517
[5] Gundlach K H and Hölzl J 1971 *Surf. Sci.* **27** 125
[6] Kadlec J and Gundlach K H 1975 *Solid State. Commun.* **16** 621
[7] Barner J B and Ruggiero S T 1989 *Phys. Rev.* B **39** 2060
[8] Lau J C and Coleman R V 1981 *Phys. Rev.* B **24** 2985
[9] Bolz R E and Tuve G L (ed) 1983 *CRC Handbook of Tables for Applied Engineering Science* (Boca Raton, FL: Chemical Rubber Company Press)
[10] Xu Yong-Nian and Ching W Y 1991 *Phys. Rev.* B **43** 4461
[11] Rippard W H, Perrella A C, Albert F J and Buhrman R A 2002 *Phys. Rev. Lett.* **88** 046805
[12] Naito M and Beasley M R 1987 *Phys. Rev.* B **35** 2548
[13] Glazman L I and Matveev K A 1988 *Sov. Phys.–JETP* **67** 1276
Glazman L I and Matveev K A 1988 *Zh. Eksp. Teor. Fiz.* **94** 332

[14] Ephron D, Xu Y and Beasley M R 1992 *Phys. Rev. Lett.* **69** 3112
[15] Bahlouli H, Matveev K A, Ephron D and Beasley M R 1994 *Phys. Rev.* B **49** 14496
[16] Simmons J G 1963 *J. Appl. Phys.* **34** 1793
[17] Gloos K, Poikolainen R S and Pekola J P 2000 *Appl. Phys. Lett.* **77** 2915
[18] Jönsson-Åkerman B J, Escudero R, Leighton C, Kim S, Schuller I K and Rabson D A 2000 *Appl. Phys. Lett.* **77** 1870
[19] Brinkman W F, Dynes R C and Rowell J M 1970 *J. Appl. Phys.* **41** 1915
[20] Jennison D R, Verdozzi C, Schultz P A and Sears M P 1999 *Phys. Rev.* B **59** R15605
[21] Chen P J and Goodman D W 1994 *Surf. Sci.* **312** L767
[22] Popova I, Zhukov V and Yates J T Jr 2000 *J. Appl. Phys.* **87** 8143
[23] Ludeke R, Cuberes M T and Cartier E 2000 *Appl. Phys. Lett.* **76** 2886
[24] Schäfer J and Adkins C J 1991 *J. Phys.: Condens. Matter* **3** 2907
[25] Jeurgens L P H, Sloof W G, Tichelaar F D and Mittemejer E J 2002 *Surf. Sci.* **506** 313
[26] Bortz M L and French R H 1989 *Appl. Phys. Lett.* **55** 1955
[27] Afanas'ev V V, Houssa M, Stesmans A and Heyns M M 2001 *J. Appl. Phys.* **78** 3073
[28] Bhatia C S, Guthmiller G and Spool A M 1989 *J. Vac. Sci. Technol.* A **7** 1298
[29] Diamond W T 1998 *J. Vac. Sci. Technol.* A **16** 707
[30] Landau L L and Lifshitz E M 1960 *Electrodynamics of Continuous Media* (Oxford: Pergamon) p 36
[31] Meade C and Jeanloz R 1990 *Phys. Rev.* B **42** 2532
[32] Mott N F and Davis E A 1979 *Electron Processes in Non-Crystalline Materials* 2nd edn (Oxford: Clarendon)
[33] Edwards P P and Sienko M J 1978 *Phys. Rev.* B **17** 2575
[34] Adkins C J, Hussain T and Ahmad N 1993 *J. Phys.: Condens. Matter* **5** 6647
[35] Abeles B, Pinch H L and Gittleman J L 1975 *Phys. Rev. Lett.* **35** 247