Analysis of breakdown in ferromagnetic tunnel junctions

W. Oepts a) and H. J. Verhagen
Department of Physics and COBRA, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

R. Coehoorn
Philips Research Laboratories, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands

W. J. M. de Jonge
Department of Physics and COBRA, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

(Received 16 December 1998; accepted for publication 29 June 1999)

Due to their very thin tunnel barrier layer, magnetic tunnel junctions show dielectric breakdown at voltages of the order of 1 V. At the moment of breakdown, a highly conductive short is formed in the barrier and is visible as a hot spot. The breakdown effect is investigated by means of voltage ramp experiments on a series of nominally identical Co/Al2O3/Co tunnel junctions. The results are described in terms of a voltage dependent breakdown probability, and are further analyzed within the framework of a general model for the breakdown probability in dielectric materials, within which it is assumed that at any time the breakdown probability is independent of the (possibly time-dependent) voltage that has been previously applied. The experimental data can be described by several specific forms of the voltage breakdown probability function. A comparison with the models commonly used for describing thin film SiO2 breakdown is given, as well as suggestions for future experiments.

© 1999 American Institute of Physics.

I. INTRODUCTION

Recently, ferromagnetic tunnel junctions have emerged as a new class of magnetoresistance devices. These tunnel junctions consist of two ferromagnetic electrodes separated by a thin oxidic barrier. When the relative orientation of the magnetizations of these two electrodes changes in an applied magnetic field, a large magnetoresistance (MR) effect is found, which is the result of spin dependent tunneling. Magnetic tunnel junctions are potentially applicable in magnetic field sensors and magnetoresistive random access memories (MRAMs). Optimizing the properties of the insulating barrier is essential for successful operation of the junctions. In order to have a sufficiently high tunnel probability, the barrier thickness (d) must be below 2.5 nm, but pinhole free. Unfortunately, these very thin insulating layers suffer from dielectric breakdown. We find that when voltages above 1 V are applied, corresponding to electric fields of the order of E ≈ 1 × 10⁹ V/m, breakdown typically occurs within a few minutes or less. Recently, we have investigated breakdown voltages for a series of Co/Al2O3/Co tunnel junctions as a function of the voltage ramp speed. The results of these measurements are given in Fig. 1 for a series of junctions with varying electrode widths. For our tunnel junctions, we concluded in Ref. 2 that with increasing applied voltage the probability of breakdown per unit of time increases, and thus the lifetime at a fixed voltage decreases. We have been able to locate the position of the breakdown with a technique for visualizing hot spots on the junction surface that is described in Ref. 3 and it has led to the conclusion that breakdown occurs at only one specific location in the junction barrier. It is known that sensor elements based on anisotropic MR (AMR) and giant MR (GMR) effects, e.g., in read heads, can be damaged by electrostatic discharge, but the typical voltages leading to damage are much higher, viz., tens of volts. This implies that for magnetic tunnel junctions, voltages applied during fabrication and operation must be controlled even more strictly.

In this article we investigate the breakdown of magnetic tunnel junctions in more detail, and we will concentrate on the analysis of the experimental data as given in Fig. 1 in terms of various forms of a general statistical model describing the voltage dependence of the breakdown probability. The parameters that are derived within these various models lead to predicted voltage dependencies of the lifetime of our tunnel junctions that differ largely. Nevertheless, the analysis gives a first indication of the applicability of these junctions at lower voltages in future devices.

We will first describe the fabrication of our samples (Sec. II), followed by characterization of the barrier properties and an overview of the results of the breakdown measurements (Sec. III). In section IV we present a general statistical model for breakdown and apply two specific forms of this model to our junctions in an analysis of the experimental data obtained. In Sec. V we summarize a number of models often used for breakdown in SiO2-based capacitors and compare them with our results. In section VI possible future experiments are discussed and a summary is given.
II. FABRICATION

The magnetic tunnel junctions used for our experiments are fabricated in a cross-bar geometry by using an in situ shadow mask system. The junctions consist of Co electrodes, evaporated in an ultrahigh vacuum (UHV) Balzers UMS630 multichamber molecular beam epitaxy (MBE) system, at a base pressure in the $10^{-8}$ Pa $(10^{-10}$ Torr) range. The tunnel junction fabrication procedure consists of the following steps. First a bottom electrode of Co (20 nm) is evaporated through the shadow mask onto a liquid N$_2$ cooled insulating Si(100) substrate. Thereafter the substrate is transported to a separate UHV chamber ($P_{\text{base}}=1 \times 10^{-7}$ Pa), and without a shadow mask a thin Al layer is sputtered from a 2 in. Al target, purity 99.999%, in an Ar pressure of 0.6 Pa. The areal density of the Al atoms in the junctions was determined by chemical analysis after the measurements. First the structure was dissolved in HCl, from which the Al content was determined with inductively coupled plasma optical emission spectrometry (ICP-OES). A total of $(4.79 \pm 0.48) \times 10^{-6}$ kg/m$^2$ of Al was found, which is equivalent to an Al thickness of $1.77 \pm 0.18$ nm. Immediately after deposition the chamber is pumped down and then filled with O$_2$ (99.999% purity) to a pressure of 9.2 Pa. A dc glow discharge is ignited from a ring-shaped cathode at a voltage of $-1.6$ kV with respect to both the substrate and the UHV chamber. In 100 s the Al layer is oxidized. If all the Al were oxidized and formed into Al$_2$O$_3$, then the Al$_2$O$_3$ thickness calculated from the ICP-OES experiment would be $2.27 \pm 0.23$ nm. To conclude the fabrication process, a 80 nm Co top electrode is evaporated in the MBE system, again through a shadow mask. The 32 junctions used for the breakdown study described below were all grown in the same run, and consisted of four series of eight samples each with 50, 150, 200, and 200 μm.

The ratio of aluminum to oxygen in the Al$_2$O$_3$ barrier was determined afterwards by scanning Auger measurements, in which a depth profile was obtained at the area of a junction by sputter etching of the sample. The peak position of the Al signal provided no indication of the presence of metallic Al, which is expected when all the Al is oxidized. During sputter etching of the barrier, the highest accuracy in determining the Al to O ratio is obtained when the middle of the barrier is reached. At this location an O/Al ratio of 1.6 ± 0.2 was found, indicating that the barrier is stoichiometric within the experimental uncertainty. In addition, the presence of O was found in the upper few layers of the Co bottom electrode, which can be explained by slight overoxidation of the barrier.

III. EXPERIMENTAL RESULTS

The junctions fabricated with the above described process have typical resistances of the order of 100 kΩ. The resistance of the tunnel junctions with the same area is observed to be equal within 10%. The junction resistance was observed to vary less strongly with the junction area than expected: the resistance of a 200×50 μm$^2$ junction is only 2 times lower than a 50×50 μm$^2$ junction. A possible explanation is a higher conductivity at the two edges of the junction with the bottom electrode. With an atomic force microscope we investigated the width of the edges of the electrode stripes and found that the edge zones of our evaporated electrodes typically have a width of 10 μm. Within the edge zones the thickness of the electrodes increases gradually from zero to the nominal value of the electrode thickness. When we assume that at the two 10 μm broad edge zones of the junction area with the bottom electrodes, where the oxide is grown on a sloping underlayer, the tunnel conductance per unit area is larger than that for the remaining surface, a ratio of a factor of 6 can explain the observed scaling of the resistance. The difference in resistivity can be the result of a difference in surface roughness of the bottom electrode at the edge. This suggests that the contribution to the total tunnel current from the two edge zones is quite significant.

In Fig. 2 a typical $I$–$V$ characteristic of a tunnel junction is shown. Initially the current increases linearly with the voltage, and at larger voltages an exponential increase is seen, which is in agreement with Simmons’ theory of

![FIG. 1. Breakdown voltages of a series of identical Co/Al$_2$O$_3$/Co tunnel junctions, measured with different voltage ramp speeds.](Image)

![FIG. 2. Current voltage characteristic of a Co/Al$_2$O$_3$/Co junction. After an initial linear dependence, the current increases exponentially until at 1.17 V a breakdown occurred. The inset shows the time dependence of the tunnel current during a 1 h voltage stress experiment with $V_{\text{bias}}=600$ mV.](Image)
tunneling. Using Simmons’ model we obtained for this junction values for the barrier height $\phi = 1.0$ eV and barrier thickness width $d = 2.21 \pm 0.01$ nm. The value of the thickness is quite comparable with the thickness ($2.27 \pm 0.23$ nm) estimated in Sec. II on the basis of the measured Al areal density and assuming the formation of stoichiometric Al$_2$O$_3$. The measurement shown in Fig. 2 was carried out by ramping the voltage at a ramp speed of $8.3 \times 10^{-4}$ V/s. At a voltage of 1.17 V breakdown of the junction occurred. The inset of Fig. 2 shows the current through a junction during a 1 h voltage stress experiment with the bottom electrode positively biased. A small increase in the current with time is found. By rapid measurement of the $I$–$V$ characteristic directly after a stress experiment, barrier parameters that are slightly different from their initial values were found. For the particular example shown, a short circuit of at least 5 min was required before the fit parameters returned to (within the error of the fit procedure) their original values. This reversible change of conductance during a stress experiment might be explained by the diffusion of ions in the barrier, for instance, of impurity atoms or of excess Al$_{3+}$ ions. As a result of this process, the shape of the barrier changes slightly, leading to an increase of the conductance. The same phenomenon was seen at negative polarity of the stress voltage.

No observation of current creep has been reported for magnetic tunnel junctions before, but it has been observed for Al/Al$_2$O$_3$/Al junctions and Al/Al$_2$O$_3$/Pb junctions. We have no direct evidence that this creep process is related to the probability of breakdown. Further discussion of this effect, however, is beyond the scope of this article.

The coercivities of the two Co electrodes of the junctions investigated were both approximately 50 Oe, which resulted in a magnetoresistance of a few percent. This is rather low compared with the MR ratio of 18% of identically fabricated Co/Al$_2$O$_3$/Co$_{50}$Fe$_{50}$ junctions, within which the two magnetic layers have strongly different coercivities. However, since the structure and thickness of the Al$_2$O$_3$ layers in the Co/Al$_2$O$_3$/Co junctions employed in the present study are identical to those in the Co/Al$_2$O$_3$/Co$_{50}$Fe$_{50}$ junctions mentioned, we regard our experimental data obtained on the breakdown effect as representative for junctions in which tunneling is strongly spin dependent.

Studying breakdown of dielectric thin films and tunnel junctions can be accomplished with the use of various measurement methods. Recently a detailed review was given by Martin et al. In the most straightforward way, the time to breakdown can be measured in an experiment in which a constant bias voltage (or current) is applied to the junction. This method has the disadvantage of the a priori unknown time to breakdown, which can exceed several days when measuring at low stress voltages, making these measurements less convenient. Therefore, breakdown is most often studied by a voltage ramp experiment, in which the applied voltage increases monotonically (often linearly) with time, and the breakdown voltage is measured. In view of the nonlinear $I$–$V$ characteristic, this method needs a feedback system in order to control the rate of voltage increase. In the actual experiment, the voltage is therefore ramped in the form of a succession of small voltage steps. A comparable experiment can be carried out with ramped applied current instead of a ramped bias voltage. In our experiments we used measurement methods with a constant bias voltage, a ramped voltage (with feedback) and a ramped current.

We observed that all junctions show breakdown within several minutes or less when a $V_{\text{bias}}$ of 1 V or higher is applied. After breakdown, the $I$–$V$ characteristic is found to be almost ohmic. The resistance of junctions after breakdown is typically of the order of 10–100 $\Omega$, although negative resistances were also measured. The latter observation is an artifact of the cross-bar geometry, in which negative resistances can arise. In the case of a square junction a negative resistance is measured when the junction resistance is less than one fourth of the lead sheet resistance. Stress measurements at a constant voltage of 1 V showed a time to breakdown of the order of minutes. Junctions grown in the same run and having the same junction area showed a comparable lifetime when biased at the same $V_{\text{bias}}$. Bioising the junction at a slightly lower voltage (e.g., 800 mV) led to an increase in the lifetime to several days or longer. The decrease of the resistance after breakdown can be understood as the formation of a microscopic ohmic short in the barrier at the moment of breakdown. In some cases it was found that the junction resistance returned to its previous value after a breakdown event. This suggests that the formation of a short is not always irreversible, and that after a while a second degradation process step leads to a sharp decrease of the local conduction (for instance, by burning away the spot with high conduction, thus repairing the junction). In the literature the latter process is often referred to as a ‘‘self healing breakdown.’’ We note that this effect was only seen in current ramped experiments, in which directly after breakdown the voltage across the junction drops.

The effect of breakdown on the junction magnetoresistance ratio was investigated with junctions with a Co$_{50}$Fe$_{50}$ top electrode. These junctions were fabricated using the same procedure as that described in Sec. II, but with a thinner barrier. The barrier thickness was 1.3 nm, as determined with Simmons’ theory. In most junctions, breakdown led to the loss of the magnetoresistance effect, which can be understood when after breakdown transport is mainly due to the ohmic conduction of the short instead of spin dependent tunneling. In one case, breakdown in a voltage ramp experiment led to a decrease of the junction resistance at low bias from 1.8 k$\Omega$ to 275 $\Omega$, and a decrease of the MR ratio from 8.2% to 1.2%, as is seen in Fig. 3. From the observation of a bias dependence we conclude that the small remaining MR effect is still a tunnel MR effect. By assuming that upon breakdown a short is formed with a low resistance in parallel with the still undamaged remainder of the tunnel junction ($R = 1.8$ k$\Omega$), we derive a short resistance of 325 $\Omega$. For an otherwise perfect sample with such a short, having no MR, one expects that the MR ratio after breakdown is 1.2%, precisely equal to the measured value. If the short could be treated as a ballistic point contact (diameter $d$ much smaller than the electron mean free path $l$), one would estimate that its diameter is of the order of 1 nm using the expression $R = 4\rho l/\pi d^2$ with $\rho = 10$ $\mu$cm and $l = 5.0$ nm. This would then, once again, suggest that the process of breakdown is
Indeed a very microscopic event. However, we stress that we actually do not know the nature of the shorts formed. Since the $I-V$ characteristic of the junction was monitored during breakdown we are able to estimate the power dissipation shortly after the moment of breakdown. At the moment of breakdown (at 1.24 V and 1.29 mA), lowering of the junction’s resistance led to a ratio of the junction resistance and circuit resistance that was not anticipated by the voltage feedback circuit, resulting in a voltage applied over the junction of 0.88 V during the next step. The current was 3.07 mA, and the power dissipation at the breakdown event can thus be estimated to be 2.7 mW. If we assume that this power is completely dissipated at the location of the short within a volume with a diameter equal to the electron mean free path of Co, we calculate a local increase of temperature of 430 °C [using the thermal conductivity of Co (100 W/mK) at room temperature, which decreases only a little at temperatures of a few hundred °C]. Although this temperature increase is not sufficient for melting of the materials involved, further structural damage to the barrier can be expected. Annealing experiments of magnetic tunnel junctions have shown that above 210 °C the junction (magnetoresistance) drops severely, indicating that irreversible processes take place in or close to the junction barrier at this temperature.\(^1\)^

Several Co/Al\(_2\)O\(_3\)/Co junctions were inspected with a scanning electron microscope after breakdown, but no visible sign of damage was found, as is expected when breakdown is a very microscopic event. The location of the short was, however, visualized with the use of a thin liquid crystal film deposited on the junction.\(^2,3\)^ With this method we observed that breakdown almost always occurred at a single location in the junction barrier, confirming the above mentioned assumption of the formation of only one short at the moment of breakdown. By examining the location of the breakdown in a series of junctions we found that in almost one half of all junctions the breakdown event occurred at the edge of the junction surface, while the remaining events that took place were randomly distributed over the junction surface. The preference of breakdown at the edges is a second indication of the aforementioned assumption of a higher tunnel probability at the edges.

For voltage ramped experiments with a small ramp speed, we noted that the occurrence of breakdown led to a smaller resistance decrease upon breakdown than in experiments with a large ramp speed. In experiments with a large $dV/dt$, the detection of a sudden increase of the current to a value above a certain limit is sufficient to identify the moment of breakdown. In experiments with a small $dV/dt$ the junction current has to be monitored more precisely to enable one to identify breakdown as a sudden increase in the ratio of the measured current during two succeeding steps of the ramp. For metal–oxide–semiconductor (MOS) capacitors containing ultrathin (\(d<5\) nm) SiO\(_2\) layers, a so-called “soft” breakdown phenomenon has been reported.\(^14-16\) In this case the increase of conductance after breakdown is much smaller than for breakdown in thicker layers. In our study the moment of breakdown is defined as the first sudden (but in some cases small) increase in the current, the same as was proposed by Depas et al.\(^15\)

In Fig. 1 we have already shown the results of a large number of measurements of the time to breakdown upon ramping the applied voltage from $V=0$ with a constant $dV/dt$, as obtained for four series of Co/Al\(_2\)O\(_3\)/Co junctions with different bottom electrode widths that were fabricated in the same run (see Sec. II). From Fig. 1 it is seen that $V_{bd}$ depends on the ramp speed. In some cases we found tunnel junctions with very low resistances and no MR, or junctions with a lower resistance which showed breakdown on a much shorter time scale than neighboring tunnel junctions. These atypical junctions, as well as the junctions that broke down during handling, have not been used for the study discussed below.

IV. ANALYSIS

Although breakdown of plasma oxidized Al\(_2\)O\(_3\)-based capacitors has been reported earlier,\(^17\) a microscopic model describing the process leading to breakdown has not yet been developed. In contrast, various mechanisms of breakdown across SiO\(_2\) in MOS capacitors have been proposed, and all are able to describe the measured data in some area of the parameter space. However, in spite of the fact that this subject has already been investigated for more than three decades, no consensus among various groups concerning the physical mechanism has been reached. This indicates that the various possible physical mechanisms behind breakdown are difficult to distinguish. Here in Sec. IV, we will give a general mathematical method to describe dielectric breakdown, and will apply two different models for the breakdown probability density to our measured data. In Sec. V we will discuss our results in terms of the existing models developed for SiO\(_2\) breakdown.

Breakdown can be described as a statistical process, and a full study requires the investigation of sufficiently large ensembles of nominally identical systems. We consider breakdown experiments that are carried out by applying at
time $t=0$ a time-independent or a time-dependent voltage (or current) stress to all systems within the ensemble at constant external conditions (such as the temperature). The result can be represented in terms of the fraction $F(t)$ of the ensemble that has shown breakdown at a certain time $t$. The breakdown rate $f(t_0) = (dF/dt)|_{t=t_0}$ is the fractional number of junctions that show breakdown with a unit time interval around a time $t_0$. We define the breakdown probability density $p(t)$ as:

$$\rho(t) = \frac{1}{1-F(t)}f(t) = \frac{1}{1-F(t)} \frac{dF}{dt}, \quad (1)$$

i.e., $p(t_0)dt$ is the probability that a junction that has not yet shown breakdown at $t=t_0$ shows breakdown in the time interval $(t_0; t_0+dt)$. In the case of a time independent breakdown probability density, $p(t)=p$, one obtains $1-F = \exp(-p)$. The fraction of nonfailed junctions thus decays exponentially with time. The mean lifetime ($\tau_{1/2}$) of the junctions of an ensemble can be defined as the time at which 50% of the junctions have experienced breakdown. When $p(t)=p$ we have

$$\tau_{1/2} = \ln(2) \frac{1}{p}. \quad (2)$$

Measured functions $F(t)$ for SiO$_2$-based capacitors at constant voltage $V$ have suggested that breakdown occurring in a very early stage of the experiment is often caused by a different physical mechanism than breakdown during a later stage; see, e.g., Refs. 18 and 19. Investigators have identified these mechanisms as extrinsic and intrinsic failures, respectively. Extrinsic failures are defect related and can, in principle, be minimized by improving factors like substrate quality and the class of cleanroom dust. Intrinsic failures are inherently related to the physical properties of the oxide and the oxide/electrode interface and to the statistical variation of their structure and local composition. In breakdown experiments of large ensembles intrinsic failures are found to dominate after certain times, when most extrinsic failures have already taken place. With decreasing capacitor area the relative importance of intrinsic breakdown increases.\textsuperscript{19} Our experiments were not carried out for large ensembles. Instead, a number of nominally identical samples was subjected to a range of voltage stress conditions (ramp rates). Nevertheless, we have (as already mentioned in Sec. III) observed that a small fraction of our junctions showed breakdown within a strikingly short time after the beginning of the voltage ramp experiments. We regard these events as truly extrinsic, and have excluded them from the data displayed in Fig. 1. Future experiments based on much larger sets of nominally identical junctions should be carried out in order to improve the statistical basis for a discrimination between “intrinsic” and “extrinsic” in our junctions.

For SiO$_2$-based capacitors the processes that have been proposed as the microscopic mechanism leading to breakdown can be divided into two categories: processes which lead to a gradual change of the atomic or electronic structure of the oxide (“wearing of the oxide”), followed finally by breakdown, and processes which occur as a single sudden event. In the first case the breakdown probability density $p(t_0)$ at time $t=t_0$ depends on the voltage-time history, $V(t)$ for $0<t<t_0$, of the junction. In the latter case the structure of the oxide and the oxide/electrode interface is essentially identical to the structure at the beginning of the stress experiment, and the breakdown probability density at time $t=t_0$ is independent of the voltage-time history of the junction: it only depends on the voltage $V(t_0)$. In the remainder of Sec. IV we will develop equations for breakdown with a probability density which is not explicitly time dependent (no wearing), but only implicitly, viz., as a result of the time dependent voltage in a voltage ramp experiment: $p(t)=p[V(t)]$. Let us assume that $p(t)$ increases monotonically with time, reflecting a monotonic increase of the breakdown probability density with increasing voltage. We then expect that there is a certain voltage, $V_{\text{max}}$, at which the breakdown rate is maximal as function of time, i.e., $df(t)/dt=0$. $V_{\text{max}}$, which depends on the ramp speed $dV/dt$, can be calculated by solving

$$\frac{df(t)}{dt} = \frac{dp(V)}{dt} (1-F-[p(V)]^2(1-F)=0, \quad (3)$$

which follows directly from Eq. (1) leading to

$$\frac{dp(V)}{dt} \bigg|_{V=V_{\text{max}}} = \frac{dV}{dt} \frac{dp(V)}{dV} \bigg|_{V=V_{\text{max}}} = [p(V_{\text{max}})]^2. \quad (4)$$

Solving Eq. (4) leads to an expression of $V_{\text{max}}$ as function of the ramp speed $dV/dt$.

In the literature on SiO$_2$ breakdown two expressions for the electric field ($E=V/d$) dependent breakdown probability density are frequently used, viz., the so-called $E$ model:\textsuperscript{20}

$$p(t) = A \exp \left( \frac{E(t)}{B} \right), \quad (5)$$

and the so-called $1/E$ model:\textsuperscript{21}

$$p(t) = C \exp \left( -\frac{D}{E(t)} \right). \quad (6)$$

Both models have been reported to give a good description of measured breakdown data within a limited field interval. Within the $1/E$ model [Eq. (6)] the factor $C$ is sometimes considered to be field dependent. We will neglect this possible complication here, and return to this issue in Sec. V, when we discuss the physics behind these two models. For the $E$ model $f(t)$ can be expressed analytically as

$$F(t) = 1 - \exp \left( -p(t)B' \left( \frac{dV}{dt} \right)^{-1} + AB' \frac{dV}{dt} \right), \quad (7)$$

with $B'=B/d$ and

$$p(t) = A \exp \left( \frac{dV}{dt} \frac{t}{d} B' \right). \quad (8)$$

From Eq. (4) the voltage $V_{\text{max}}$ at which $f(t)$ peaks is given by

$$V_{\text{max}} = B' \ln \left( \frac{dV}{dt} \frac{1}{AB'} \right). \quad (9)$$
For the 1/E model $F(t)$ and $V_{\text{max}}$ cannot be expressed in a closed analytical form. In Fig. 4 we have plotted our data and the results of fits with the two models for the 50 $\mu$m bottom electrode data. We have chosen to discuss the series with 50 $\mu$m electrodes, which contains the most data points over the largest range of voltage ramp speed. However, fits of the models to the series with other bottom electrode widths were also possible. The fits of these models resulted in the following values of the parameters defined in Eqs. (5) and (6): $A = 6.3 \pm 0.5 \times 10^{-17}$ s$^{-1}$ and $B/d = 0.035 \pm 0.002$ V, or $C = 5.6 \pm 0.7 \times 10^{9}$ s$^{-1}$ and $D/d = 31 \pm 2$ V. We note that the errors given are the errors of the fit to the data points. Both models, however, are based on statistics and therefore larger scatter of the data points may be expected. This point will be discussed further on. In Fig. 5 both $F(t)$ and $f(t)$ for a voltage ramp experiment are plotted as a function of voltage for the case of the E model with A and $B'$ as given above and $dV/dt = 1 \times 10^{-6}$ V/s. The function $f(t)$ has an asymmetric peak shape. Note that in this example $F(t)$, the fraction of broken down junctions, is more than 0.5 at the moment at which $V_{\text{max}}$ is reached. In order to obtain an estimation of the (statistical) scattering in experimental $V_{\text{bd}}$ data when investigating a large ensemble, the width of the peak of the breakdown rate $f(t)$ can be used. The width of this peak is centered at $V = V_{\text{max}}$, and can be characterized by the distance between the two inflection points, which are the solutions of the following equation:

$$\frac{d^2 f(t)}{dt^2} = \frac{d^2 p(V)}{dV^2} \left(\frac{dV}{dt}\right)^2 - 3 p(V) \frac{dp(V)}{dV} \frac{dV}{dt} + [p(V)]^2 = 0.$$  

(10)

The area between the two inflection points of $f(t)$ is found to be independent of parameters $A$, $B$, and $dV/dt$, and is equal to 61% of the total area. For the 1/E model the parameter values that lead to good fits of our experimental data (see below) lead to peak shapes that are quite similar to those for the E model. We have derived the area between the inflection points numerically: for the parameters that describe our experimental data it is approximately 65%. We conclude that for both models the intervals between the inflection points of $f(t)$ give a good estimation of the expected variation (scattering) of the experimentally measured $V_{\text{bd}}$ values. In Fig. 4 the dashed lines represent the calculated voltage at the inflection points of the fitted $f(t)$ of the 50 $\mu$m wide bottom electrode junctions with both models. These experimental data points fall in the area between these lines, indicating that scattering of the data is still well described by both the E model [Eq. (5)] and the 1/E model (Eq. 6). The prefactors $A$ and $C$ in Eqs. (5) and (6) are proportional to the junction area if the breakdown probability is independent of the location on the junction. As a result, $V_{\text{max}}$ is expected to vary with a change of junction area $S$ with a factor $n$ as

$$V_{\text{max}} \left(\frac{dV}{dt}\right) = V_{\text{max}} \left(\frac{S}{n \frac{dV}{dt}}\right).$$  

(11)

From Fig. 4 we note that the $V_{\text{bd}}$ values do not significantly decrease with increasing junction area, although it is expected that a factor of 4 in area would be significant enough to be visible in the data. However, due to the observed large breakdown probability at the bottom edge region, the influence of a change of area might not be distinguishable. We have insufficient information on the exact area dependent breakdown probability to derive a scaling relation such as that discussed in Sec. III for the resistance.

Although both models can describe the experimental data very well, a large difference between the extrapolated lifetimes at lower $V_{\text{bias}}$ between both models is found. In Fig. 6 the calculated lifetime curves for both models are shown, including an estimation for the statistical variation. At $V_{\text{bias}} = 0.5$ V we find extrapolated lifetimes of 217 years or $10^9$ years for the E model and the 1/E model, respectively. Both values suggest that the applicability of these junctions at low bias is not hindered by short lifetimes. We stress that these
values are still only an indication due to the uncertainty of the applicability of these models for breakdown of our junction. In order to distinguish more clearly between the two models, either fast (ramp rate ≥ 5 V/s) or long duration (more than 1 year) measurements or experiments on a much larger number of samples must be carried out.

V. MICROSCOPIC MODELS OF BREAKDOWN

The physical background of the models leading to Eqs. (5) and (6) will now be explained, as will other models proposed for breakdown across SiO$_2$. The corresponding threshold voltages above which the breakdown effect occurs are discussed (if appropriate), and the possible relevance to breakdown across Al$_2$O$_3$ is given. When comparing the observed breakdown in our Al$_2$O$_3$ junctions with models of breakdown commonly used for SiO$_2$ films it is important to evaluate the validity of the SiO$_2$ models in the thickness range of our films. In ultrathin films (i.e., less than 3 nm) the applied voltage $V_a$ will be limited to a few volts due to breakdown. When $eV_a$ is smaller than the barrier height ($\phi$) direct electron tunneling occurs, as described by Simmons’ theory. Breakdown upon direct tunneling across ultrathin SiO$_2$ films has not yet been explored in much detail, since SiO$_2$ layer thicknesses in MOS capacitors have only recently become thinner than 3 nm. The voltages applied in breakdown studies for 15–40 nm SiO$_2$ thin films will be much larger than $\phi/e$, and lead to Fowler-Nordheim tunneling. The difference between both types of tunneling is schematically shown in Fig. 7. In the Fowler-Nordheim tunneling regime, electrons enter the conduction band of the dielectric and can subsequently lose part of their energy due to inelastic scattering processes.

Due to the higher energy of the tunneling electrons at higher applied voltages, the processes leading to breakdown can differ from those at lower voltages. Certain processes have threshold voltages below which they will not occur. The highest threshold is for electron energies of the order of the insulating band gap (i.e., for SiO$_2$ 9 eV or larger), when breakdown can be triggered by electrons ionizing atoms in the barrier resulting in the occurrence of an electron cascade in the barrier. This phenomenon was first described by Von Hippel in 1936. For amorphous Al$_2$O$_3$ the energy gap is not well known, but most likely it is in the range of 7–9 eV. Electrons with such high excitation energies do not occur in our ultra thin films. Therefore this process can be excluded with certainty.

The three models that follow in Secs. V A, V B, and V C are different with respect to the physical model involved, the threshold voltage and the oxide thickness range for which they are relevant. An important issue that distinguishes different physical models is whether the breakdown probability is dependent (due to wear out) or independent of the stress $(V,t)$ history.

A. $Q_{bd}$ model

When tunneled electrons have sufficient energy to damage the oxide locally, so-called wearing of the oxide may result. For the case of tunneling across SiO$_2$ strong experimental support in favor of the wearing mechanism has been obtained by Wolters and van der Schoot, who observed that for 8–40 nm SiO$_2$-based capacitors under certain conditions the breakdown probability density depends on the total amount of tunneled electrons, the (electron) charge to breakdown, $Q_{bd}$, which is the electron current density integrated until breakdown. The quantity $Q_{bd}$ (averaged over a large ensemble) was found to be independent of the type of experiment employed (e.g., constant voltage or current, or ramped voltage or current). The authors presented evidence that damage to the SiO$_2$ lattice due to energy dissipation of the tunneled electrons starts at the positive electrode, and gradually forms a low resistance path through the oxide, finally leading to breakdown. The energy available for creating damage to the oxide is argued to be independent of the applied voltage or current density because the tunneled electrons always lose most of their energy ($>3$ eV) when they enter the anode. It is crucial to note that breakdown of these
relatively thick layers occurs upon tunneling in the Fowler-Nordheim regime, in which the electrons tunnel through a triangular barrier, and subsequently into the conduction band of the oxide [Fig. 7(b)]. The assumption that leads to the \( Q_{bd} \) model is that the inelastic mean free path of these electrons is relatively short and the barrier height at the anode is relatively large. In Fig. 8 we have plotted the measured total passed charge at breakdown for our Co/Al\(_2\)O\(_3\)/Co junctions with 50 \( \mu \)m bottom electrodes, the breakdown voltages of which are given in Fig. 4. It is seen that for our junctions a constant \( Q_{bd} \) is not found, but that \( Q_{bd} \) decreases with higher voltages (and current densities). It has been found for SiO\(_2\)-based systems that the constant \( Q_{bd} \) model is valid within a good approximation for low current densities \( J \) at the moment of breakdown, but that above a critical current density, \( J_{cr} \), \( Q_{bd} \) decreases rapidly with increasing \( J \). This was explained from the (independent) observation that for \( J > J_{cr} \) the oxide contains a significant fraction of occupied traps (relatively localized electron states which at \( V = 0 \) are empty because their energy is higher than \( E_F \), but whose occupation increases with increasing \( J \)). The trapped electronic charge modifies the originally relatively flat equipotential planes, leading to confinement of the injected electron current to regions with a low space charge density, thereby enhancing the local current density and decreasing \( Q_{bd} \). We have no indication that the failure of the constant-\( Q_{bd} \) model to our junction breakdown could be related to the occurrence of a current density exceeding a certain critical value, \( J_{cr} \), above which the density of occupied trap states is very high. In such a case the conductivity would be observed to change an order of magnitude with time in a constant applied voltage experiment,\(^{18}\) a phenomenon that we did not observe.

### B. Anode hole injection model

The results of breakdown experiments for structures containing thinner (typically 3–15 nm) SiO\(_2\) layers have not been found to be in agreement within the constant-\( Q_{bd} \) model. Instead, it has been possible to interpret the experimental data obtained within the so-called anode hole injection model [Eq. (6)], developed by Schuegraf and Hu,\(^{25}\) which leads to the \( 1/E \) model. In the anode hole injection model it is assumed that the energy of the electrons arriving at the anode (which again is assumed to be equal to or only slightly larger than the barrier height) is used to excite deep-valence band electrons to a state above the Fermi level, thereby creating a hole in the anode which can tunnel into the oxide. This may generate an electron trap in the oxide, leading to an enhanced local current density. The experimental possibility of carrier separation in SiO\(_2\) devices enabled the determination of the hole current during a stress experiment. This led to the observation that, for a given oxide thickness, the hole charge to breakdown, \( Q_{bd,h} \), is a constant in an experiment at constant applied voltage, supporting the anode hole injection model, whereas the electron charge to breakdown, \( Q_{bd,e} \), decreases with applied voltage. The hole fluence is smaller than the current density, but it increases with the applied voltage: not every electron arriving at the anode will create a hole that forms a trap, but the probability of hole tunneling increases with the applied voltage.\(^{19}\) When hole tunneling from the anode is in the Fowler-Nordheim regime, the hole tunnel current is proportional to \( E^2 \exp (-D/E) \), with \( D \) a parameter independent of the electric field, \( E \). Since the density of holes in the anode is proportional with the electron current \( J(E) \), one expects that the breakdown probability density is given by

\[
p(E) = C J(E) E^2 \exp \left( -\frac{D}{E} \right),
\]

where \( C \) is a constant, if the creation of only one trap state is sufficient for triggering breakdown. In that case wearing of the oxide plays no role, and \( p \) does not depend explicitly on \( t \). Evidence for this has been obtained by Degraeve et al.\(^{26}\) for the case of ultrathin (\( d < 2 \) nm oxides). On the other hand, for thicker oxides the same authors concluded that a critical trap density has to be created first before the damage formed leads to breakdown. Numerical (Monte Carlo) studies then yield the time-dependent breakdown probability at constant field or voltage.\(^{26,27}\) In many analyses of breakdown experiments the restrictions mentioned concerning the validity of Eq. (12) have been disregarded. In addition, in practical applications of the model the field dependence of the prefactor of the exponential function in Eq. (12) is often neglected. In that case Eq. (12) reduces to Eq. (6), which is commonly referred to as the \( 1/E \) model. Although this seems a rather crude approximation, it must be remembered that the formula is only applied to data obtained in a restricted field range, in which variations due to the exponential factor dominate variations in \( p(E) \). This was confirmed by a fit of our experimental data to the functional form of \( p(E) \) given by Eq. (12), from which we obtained \( D/d = 29.2 \pm 2.4 \) V. However, if we insert a current density \( J(E) \) as determined with a fit to our experimentally current density of the form \( J = f_1 \exp(f_2 V) \), we still obtained a good fit, but with \( D/d = 11.8 \pm 0.5 \) V or with \( D/d = 22.5 \pm 15 \) V if this current density was directly inserted into Eq. (6) (thus without the \( E^2 \) term).
In spite of the good fits obtained using both approaches, we cannot find strong physical arguments in favor of the $1/E$ model if applied to our junctions. First, hole tunneling is not expected to be in the Fowler-Nordheim regime, since $\phi_h = E_g - \phi_p > \phi_p > V_a$ (assuming a band gap $E_g$ for Al$_2$O$_3$ of $\sim 7-9$ eV and using a barrier height for electrons of $\phi_e \approx 1$ eV). This invalidates the derivation of the exponential factor that is used in Eq. (12). Second, this estimate shows that the energy of the holes created (which is at most equal to the energy $eV$ of the tunneled electrons) is much less than $\phi_h$, so that direct hole tunneling through the barrier becomes extremely unlikely. In fact, for the same reason DiMaria et al. have argued that for SiO$_2$-based systems hole tunneling does not occur for applied voltages over the barrier below a certain threshold value, $V_{th} \approx 5$ V.\footnote{28}

The anode hole injection model for SiO$_2$ has been extended in order to describe the effect of breakdown at relatively low voltages due to release of hydrogen (H$^+$) into the oxide as the result of electrons that arrive with an excess energy at the anode.\footnote{23} In SiO$_2$-based devices hydrogen is present due to a hydrogen passivation step. The fabrication of our Al$_2$O$_3$ barriers does not include such a step. We do not have any indication that this mechanism is applicable to our junctions.

**C. E model**

The E model relates breakdown to the field induced distortion of atomic bonds in the oxidic barrier. It is often alternatively referred to as the thermochemical model. Based on thermodynamic free energy considerations, a quantitative thermodynamic model for breakdown for SiO$_2$-based capacitors was developed by McPherson and Mogul\footnote{20} and by Kimura.\footnote{20} In the case of both amorphous or crystalline SiO$_2$, a Si atom is surrounded by a tetrahedron formed by four oxygen atoms. However, during the growth of amorphous SiO$_2$ or in the presence of an applied external field, the Si-O bond angles can be distorted. If the distortion leads to bond angles above or below a critical value, the oxygen atom will be displaced and a Si–Si bond is formed. This defect is believed to be the precursor for breakdown, which will occur when either one Si–Si bond is broken or when a certain critical fraction of broken bonds is reached. Although the authors do not make clear whether the breakdown rate determining mechanism is the breaking of a Si–O–Si bond (with a very high bond energy) or the breaking of the much weaker Si–Si bond, McPherson and Mogul have derived an expression for the breakdown probability based upon the breaking of a Si–Si bond.\footnote{20} The authors start by stating that the local field acting on atoms in the oxide will be equal to the externally applied field, enhanced by a contribution due to the polarization of the dielectric: $E_{loc} = (1 + L \chi) E$, in which $L$ is the Lorentz factor ($L = 1/3$ for cubic point symmetry), and $\chi$ is the electric susceptibility. The Si–Si pair is part of a structural fragment in the network of the form O$_3$=Si–Si=O$_3$. The two Si=O$_3$ dipoles which form this fragment have antiparallel dipole moments, $\pm p$. In the presence of an electric field parallel to the Si–Si bond, i.e., parallel and antiparallel to the two dipole moments, the contribution from the two dipoles to the total energy is decreased and increased, respectively, by an energy $p \cdot E_{loc}$. This energy helps to lower the activation barrier for collapse of the antiparallel dipole, resulting in a broken Si–Si bond, and also resulting in the creation of a localized electronic defect which gives rise to dielectric breakdown.\footnote{20} The average time to breakdown due to this process is given by:\footnote{20}

$$\tau(E) = a \exp \left( \frac{\Delta H - p(1 + L \chi) E}{k_BT} \right)$$

$$= a \exp \left( \frac{\Delta H}{k_BT} \right) \exp(-\gamma E),$$

where $a$ is a constant, and $\gamma$ the so-called field acceleration factor. Assuming that the breakdown probability is not explicitly time dependent (no wearout) one obtains Eq. (5) with $A = \exp(-\Delta H/k_BT)/a$ and $B = 1/\gamma$. A good fit of the SiO$_2$ ($d = 10.0$ nm) breakdown data obtained by Kimura\footnote{20} was found with this model, and at room temperature the experimentally obtained value of $\gamma$ is $3.25 \times 10^{-8}$ m/V. Also, the temperature dependence of $\gamma$ could be explained well,\footnote{20} which provides strong support in favor of this model. From a fit to our experimental data, we found (Sec. IV) $B/d = 0.035$ V. When we assume a barrier thickness of $2.2$ nm we find $\gamma = 1/B = 6.3 \times 10^{-8}$ m/V. This value is approximately a factor of 2 higher than the value obtained for SiO$_2$ from the Kimura data. We would like to note that when making such a comparison at least three points concerning the difference between amorphous Al$_2$O$_3$ and SiO$_2$ should be considered.

(i) The electric susceptibility of Al$_2$O$_3$ ($\chi \approx 7.0$) is more than two times higher than for SiO$_2$ ($\chi \approx 2.9$), leading to a higher polarization and local field, and to, therefore (all other factors remaining the same), a higher value of $\gamma$.

(ii) Bonds in Al$_2$O$_3$ are more ionic (less covalent) than in SiO$_2$. One implication is that the effective ionic charge of Al is potentially larger than that of Si, which leads to higher dipole moments in the Al$_2$O$_3$ structure and a higher value of $\gamma$.

(iii) The structure of amorphous Al$_2$O$_3$ is more complex compared with that of SiO$_2$: the short range order has been described as being similar to that in $\gamma$-Fe$_2$O$_3$, i.e., a spinel structure in which the oxygen atoms form a close packed lattice and in which the cations fill tetrahedral sites as well as 66% of the available octahedral sites. In this structure various local environments of the Al ions occur.\footnote{30} For amorphous Al$_2$O$_3$ it has been found that the fraction of Al ions with a tetrahedral coordination is even larger than that in crystalline $\gamma$-Al$_2$O$_3$. Migration of the Al could probably occur via the octahedral vacancies.

An important additional consideration in favor of the applicability of this model is the fact that there is no threshold value, so it can already be applied at the low voltages at which we observed breakdown. We therefore believe that this model could be applicable to our Al$_2$O$_3$ layers.

**VI. CONCLUDING REMARKS AND SUMMARY**

We have compared our experimental data with models proposed for SiO$_2$ breakdown. Good fits are possible in both
the $E$ model and the $1/E$ model for the breakdown probability density. However, only the mechanism that leads to the $E$ model, field induced atomic displacements of atoms in the oxide, seems to provide a plausible physical basis for the breakdown effect observed. The present study is clearly limited in several respects, and extensions in many directions are required in order to clarify the breakdown mechanism and to be able to relate the breakdown properties with the oxide and interface structure, thickness, and composition. One of the unsolved issues is wearout. Although we have not obtained evidence of the occurrence of wearout before breakdown that would lead to an explicit time dependence of the breakdown probability, it is of interest to study this issue in dedicated experiments, e.g., in experiments of large ensembles under constant voltage stress. For this purpose frequency resolved electronic (current or voltage) noise measurements are also expected to be valuable, since small structural changes in the oxide might lead to a different noise spectrum, especially for the $1/f^2$ contribution to the noise spectral density. For thin SiO$_2$ layers, “soft” breakdown has been reported when low fields are applied, suggesting that small structural changes do take place during low voltage stress. It is expected that these small changes will have a much larger effect on the total junction (magneto)resistance when the area is reduced compared with the area of our junctions by microstructuring, which will be the case in practical applications. In order to distinguish between the $E$ and $1/E$ models, measurements should be extended to a larger voltage range than that used in the study, and the statistical uncertainty of each data point should be decreased by the use of larger ensembles. Also, studying the temperature dependence of the breakdown probability will provide tests of the applicability of either model. Finally, using lithographically defined junctions, the areal dependence of breakdown should be studied in order to be able to discriminate between extrinsic and intrinsic breakdown and in order to study edge effects.

In summary, we have investigated the dielectric breakdown of successfully fabricated magnetic tunnel junctions with a plasma oxidized Al$_2$O$_3$ barrier. In junctions with barriers of 2.0 nm and thinner, almost immediate breakdown is observed when voltages above 1 V are applied. At the moment of breakdown a single short is created in the barrier, leading to a low resistive path in the barrier. The exact nature and mechanism for creation of this short is still uncertain. The time-dependent breakdown probability is voltage dependent, and increases with the applied voltage. This dependence is analyzed in the framework of a general model for breakdown with a probability that is not explicitly time dependent. With this model it is shown that ramp rate dependence of the breakdown voltage as observed in voltage ramp experiments can be described using different expressions (the $E$ and $1/E$ models), which have been proposed earlier for SiO$_2$ breakdown, and which have been obtained on the basis of different physical models. Suggestions have been given for experiments from which the possible physical mechanisms can be distinguished more easily. Extrapolation of the lifetime curves of our junctions to realistic low operation voltages ($V_{\text{bias}}<0.5$ V), as obtained from fits to the experimental data, suggest that, if accidental peak voltages outside this region can be avoided, breakdown will not be a limiting factor upon applying these junctions in sensor or MRAM devices.

**ACKNOWLEDGMENTS**

The authors would like to thank D. R. Wolters for valuable discussions. This work was supported by the European Community ESPRIT Research project, Novel Magnetic Nanodevices of Artificially Layered Materials (NM).

---