Dielectric breakdown I: A review of oxide breakdown

J.F. Verweij and J.H. Klootwijk

MESA Research Institute, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands. Tel: (+31)53-894119. Secr.: (+31)53-892644. Fax: (+31)53-341903. Tel: (+31)53-892648. Secr.: (+31)53-892644. Fax: (+31)53-341903.
E-mail: J.H.Klootwijk@el.utwente.nl

This paper gives an overview of the dielectric breakdown in thin oxide layers on silicon. First test methods are discussed, followed by their application to the estimation of the oxide lifetime. The main part of the paper is devoted to the physical background of the intrinsic breakdown. Finally, defect-related or extrinsic breakdown is discussed. Copyright © 1996 Elsevier Science Ltd.

1. Introduction

The properties of thin dielectrics remain very important for the reliability of semiconductor devices. In the first place we think of the gate oxide in integrated circuits. It is well known that defects, contamination and metal impurities are hazardous for a sufficient lifetime of the integrated circuits. Therefore, a large effort is spent on the construction of better clean rooms, in the first place for better yields but also for better reliability of the devices.

There are some trends counteracting this effort, namely, the use of a thinner gate oxide in scaled devices and larger integrated circuits, which make the devices more vulnerable to hazardous defects. Moreover, processing of gate oxides changes. For instance, a nitridation is carried out, or the oxidation is carried out at lower temperatures. Even new dielectrics are used like oxide-nitride-oxide (ONO) layers in Electrically Erasable Programmable Read Only Memory (EEPROM) devices.

All these trends make it understandable that there is a continuous interest in the properties of thin dielectrics in general and the breakdown phenomenon in particular. The latter is the subject of this paper because breakdown is the determining factor in oxide lifetime. In this paper we treat the oxide breakdown and discuss sequentially breakdown test methods, the relation between instantaneous breakdown and time-dependent dielectric breakdown (TDDB) via the concept of charge to breakdown (Q_{BD}), the oxide lifetime prediction out of Q_{BD}, the intrinsic breakdown mechanism and, finally, the extrinsic or defect-related breakdown mechanism.

2. Test methods for oxide breakdown

Different failure modes can be observed depending on the area of the oxide, the time...
frame of the observation and, especially, the applied electric field [1-4]. The dielectric strength of an oxide layer is often expressed in terms of the electric field at which the insulator is irreversibly damaged and has lost its insulating properties. There are several test methods for measuring the breakdown parameters depending on the way the stress voltage or stress current are applied. We start with the most widely known technique, namely, the voltage ramp. The voltage is ramped with a constant ramp rate $dV/dt$. While the current through the oxide is being measured, it increases towards higher voltage levels. When there is breakdown the current jumps to a high level and passes a pre-set current limit $I_{\text{limit}}$. The value of the voltage at which this happens is defined as the breakdown voltage. It should be noted that the actual current shape that can be measured is dependent on the sensitivity of the current meter. With a less sensitive instrument only the current jump is detected. Of course, this is sufficient for determining the breakdown voltage. Dividing the latter by the oxide thickness we obtain the breakdown field (provided no polysilicon depletion is necessary).

Oxide breakdown has a strong statistical nature. In this and other breakdown techniques [5], a relatively large number of test structures (i.e. capacitors) are used to find the actual failure distribution. The cumulative number of failures $F$ is plotted in a statistical plot. Figure 1 shows results obtained on capacitors with oxide thickness of 10.7 nm [1] measured with the ramping voltage technique. A particular distribution appears to be applicable to oxide breakdown, namely, extreme value statistics. Then $\ln(-\ln(1-F))$ is plotted vs. the breakdown field, linearly (or vs. the log of the breakdown time in TDDB experiments).

In the experiments of Fig. 1 capacitors with three different areas were used, namely, 0.2, 0.8 and 3.2 mm$^2$. This was used to show the validity of the extreme value statistics. It should be noted that in Fig. 1 we can clearly observe the intrinsic breakdowns and the defect-related breakdowns with a wide distribution at low and intermediate fields.

A second technique consists of the application of a constant voltage and measuring the time to breakdown. This is the TDDB. An example of recent results, in this case on inter-poly oxide, is shown in Fig. 2. This is taken from [2]. Again, a pre-set current limit serves for defining the breakdown. This limit should be adapted to the same current density level for different area capacitors. The data points in Fig. 2 are obtained by voltage stressing a 22 nm inter-poly oxide at six different voltages. The solid lines through the distributions are obtained by predictions based on ramped-voltage test results taking trapping in the oxide into account. The model used for this is outside the scope of this paper. The results in Fig. 2 are shown here primarily as an illustration of TDDB, but also to show that this breakdown again has a strong dependence on electric field.

Forcing a current through an oxide layer can cause a breakdown. It appears that this breakdown is related to breakdown in practical use conditions but that the time to breakdown can be considerably shorter than in the constant voltage technique. The first experiments reported [6] show that the time to breakdown $t_{\text{BD}}$...
times the stress current $I$ is approximately constant. This led Wolters and coworkers [1, 4] to the introduction of the concept of the charge to breakdown $Q_{BD}$, which is $P_{f_{BD}}$, or better $Q_{BD} = \int_{0}^{T_{BD}} I(t) \, dt$, as the actual measure of dielectric strength. It appears that also this phenomenon can also be observed in the ramped or constant-voltage stressing. This is described below in more detail. It should be noted that the oxides of Wolters and coworkers had thicknesses in the range 11 nm to 100 nm.

3. Relation between instantaneous breakdown and TDDB

The relation between instantaneous or field breakdown and TDDB will be discussed with the aid of Fig. 3 and the following [1]. We start with Fig. 3 which shows a breakdown distribution in an extreme value plot. We see the defect-related part which was relatively high and the intrinsic part starting at a field as high as 13 MV/cm. Capacitors out of the same lot were stressed in a constant voltage test (TDDB) at field values as indicated by the arrows in the figure ($E = 10.8$-$11.6$ MV/cm). The results of this TDDB experiment are given in Fig. 4, again in an extreme value statistics plot. It can be seen clearly that the plots consist of two parts, one at relatively short times with a small slope and one at the end of the stress period with a steep slope. All curves have their bend at about the same defect density. Moreover, this defect density is the same as that found in Fig. 4. This strongly suggests that the distribution in Fig. 4 with the small slope is defect-related and the distribu-
Fig. 3. Breakdown distribution of capacitors with a relatively high oxide defect density. The arrows indicate the electric fields at which time-dependent tests were carried out on oxides from the same batch. Reproduced from [1] with permission.

Fig. 4. Breakdown distributions of the time-dependent tests mentioned in Fig. 3. Reproduced from [1] with permission.

These results indicate that the measurement of breakdown in the classical ramped field technique already gives information on the behaviour to be expected in lifetime experiments, i.e. in TDDB. The relation between the two kind of experiments comes from the concept of charge to breakdown, $Q_{BD}$. In both the ramped field and constant field techniques the breakdown occurs when the value of $Q_{BD}$ has been reached.

More evidence comes from measuring the current through the oxide in a constant voltage/bias stress and calculating the $Q_{BD}$ at the breakdown time [4]. This is compared with $Q_{BD}$ in a constant current stress. The result is plotted in
Fig. 5. Comparison of breakdown distributions obtained in a constant current stress and in a constant voltage/bias stress. Reproduced from [4] with permission.

Fig. 5. It can be seen that exactly the same intrinsic distribution is found and about the same defect-related distribution. The differences at low $Q_{bd}$ are mainly due to the accuracy of the equipment at short times (and/or at low currents).

These results prove that the information on the breakdown distributions is virtually independent of the actual measurement technique. It also means that at any electric field a current is flowing through the oxide layer, which was supposed to be an insulator. The question is: what is the actual conduction mechanism? The answer is: injection through the oxide-silicon energy barrier, described as Fowler-Nordheim tunneling [7]. The details of this mechanism are described below in the discussion on the breakdown mechanism. The property of this injection is that it is so extremely low at operating voltages that an oxide layer can be considered as a very good insulator.

4. Oxide lifetime prediction

The use of the $Q_{bd}$ concept in reliability predictions strongly depends on the behaviour of $Q_{bd}$ as a function of the stress current density. A result of this behaviour is plotted in Fig. 6, taken from [8]. It should be noted that the points represent averages at the different current densities. These measurements were carried out on
It can be seen that in Fig. 6, below a stress current density of 0.05 A/cm², the value of $Q_{BD}$ is virtually constant and has a value of about 20 C/cm². This can be used for intrinsic reliability estimates of gate oxide (see below) and the switching degradation in floating gate EEPROM. A constancy of $Q_{BD}$ would allow a simple estimate of the time to breakdown under use via eq. (1):

$$t_{BD} = \frac{Q_{BD}}{J_u}$$

where $J_u$ is the current density under use conditions. However, there are two complications for the use of $Q_{BD}$ in these cases. The first is the sharp drop in $Q_{BD}$ at high stress current densities, which can be ascribed to inhomogeneous injection. This has been proved by a study of the homogeneity of the charge injection with the aid of photoemission [8]. Figure 6 shows the results on photoemission, or more specifically the emission ratio between the edge, $m_e$, and centre, $m_c$, of the test capacitors as a function of the injection current density. It can be seen that at the current density where $Q_{BD}$ starts to drop then the value of $m_e/m_c$ goes up from one, indicative of inhomogeneous injection. Moreover, it indicates that the values of $Q_{BD}$ at these high injection levels are not relevant for reliability prediction!

The second complication is a much more serious one. It is the phenomenon that $Q_{BD}$ is not constant for oxide layers as thin as 10 nm, but that it increases with decreasing oxide voltage (i.e. with decreasing stress current densities) [9]. Therefore, for thin oxides the extrapolation from $Q_{BD}$ in stress to time-to-breakdown in practical use needs a particular model. This could be the anode hole injection model as presented by the authors in [9]. The physical background of this and other models will be discussed in the next sections.

5. Intrinsic breakdown mechanism

The actual mechanism of the destructive oxide breakdown is rather complicated [10], and some aspects are still of speculative nature. Nevertheless, model descriptions have been given [11-13] in terms of acceleration factors for temperature and electric field during tests. Of course, the model parameters are translated into physical quantities if possible. In this paper we concentrate on the physical mechanism of breakdown. We may discern a number of sequential phases in the degradation process leading to the final breakdown. The initial phases form the clearer and unambiguous parts of the mechanism [10]. In the previous sections we have seen that a certain amount of charge, namely $Q_{BD}$, has to be transported before breakdown occurs. As the oxide is an insulator, which means that it has an extremely low concentration of free carriers, this implies that the charge carriers, in order to reach $Q_{BD}$, have to be injected. It is generally accepted that this injection is Fowler-Nordheim [7]. This
Fig. 7. Energy band diagram of the system silicon (Si), silicon oxide (SiO₂), metal (M), during electron (e) injection from the silicon. Reproduced from [5] with permission.

is illustrated in Fig. 7, which shows the energy band diagram of oxide (SiO₂) between a metal electrode (M, positive) and a silicon substrate (Si). The quantities $E_c$ and $E_v$ designate the bottom of the conduction band and the top of the valence band, respectively.

In Fig. 7 the upper arrow at the left-hand side indicates the injection of electrons from the silicon into the SiO₂ conduction band by Fowler-Nordheim tunnelling. This is a tunnelling through the energy barrier between the Si and SiO₂ conduction band and may give rise to a measurable electric current. The equation for the current density $J$ is

$$J = AE^2 \exp \left( -\frac{B}{E} \right)$$

(2)

where $E$ is the oxide field and $A$ and $B$ constants not dependent on this field. For the next phase in the breakdown process there are several possibilities. After injection the electrons travel through the conduction band. They may gain energy from the electric field, but a consideration of the scattering processes together with experimental results leads to the conclusion that the energy gain is low, but still sufficient to break bonds and to create traps. For a recent discussion on the different scattering mechanisms we refer to [14]. What kind of traps are then generated? Arnold et al. [14] believe that this is the liberation of hydrogen by the energetic electrons. Although this is a possibility there is yet no firm proof. It should be noted that the liberation of hydrogen, which is bonded in the SiO₂ lattice during the processing, can be liberated by reaction with holes [15].

This brings us to the next possibility: the anode hole injection. At the metal side (anode side) of the oxide the electron drops down to the Fermi level, $E_F$, and may donate its energy of at least 3.1 eV (see Fig. 7) to the lattice at the SiO₂-M interface. This energy is sufficient to break an Si-O bond, hence in this way the damage also starts [3]. Also, the energy may be donated to an electron-hole at the anode side so that it gets sufficient energy to cross the SiO₂-M barrier and to be injected into the SiO₂ valence band. This anode hole injection has already been mentioned above as serving for the physical background of the breakdown model of Schuegraf and Hu [9]. In a recent publication DiMaria [16] showed that anode hole injection is important, however only for gate voltages larger than $\approx 7.6$ V, while Schuegraf and Hu used their model for quantitatively describing breakdown down to 2.5 V in very thin oxides. In this respect we should mention the experiments of Gao [17]. He found that the value of $Q_{BD}$ is influenced by the cathode material (and not by the anode material), i.e. $Q_{BD}$ increases with decreasing barrier at the cathode side of the SiO₂ system. This seems to oppose the anode hole injection as a dominant mechanism in breakdown. However, we should mention that Gao did not check for homogeneity of the electron injection at the cathode when Au and Al were the injecting electrodes (with the larger barriers in comparison with polysilicon). We suspect that the inhomogeneity can be quite large in this way, leading to low values of $Q_{BD}$. Anyhow, we may conclude that more investigations are needed to settle the controversies about the anode hole injection. After the injection holes are transported through
the oxide, although at a much lower speed compared with electrons. Part of them are also trapped in the oxide at the Si-SiO₂ interface where they may stimulate the electron injection [12]. Another possibility is the removal of an electron from Si-H or SiO-H bonds leading to the liberation of hydrogen. The holes may even remove electrons from particular Si-O bonds upon hole trapping, in this way loosening the SiO₂ network [10]. It should be noted that the SiO₂ is a non-crystalline layer with many strained bonds that are vulnerable for breaking and defect creation. Concerning the creation of traps it should be noted that holes are able to create traps at the Si-SiO₂ interface [15]. However, Kerber and Schwalke [18] have concluded that the mechanisms of interface trap creation and of breakdown are different. This is based on the observed differences in dependence on injected charge and temperature. In our opinion this should not be too surprising, because interface traps are measured when they are homogeneously distributed, while the (final phase of) breakdown has a strong inhomogeneous nature.

A third possibility for trap creation is impact ionization. When accelerated electrons in the SiO₂ have sufficient energy they may create electron-hole pairs via band-band transition of electrons. For a review of this scattering process and its probability in comparison with other scattering processes we refer to the literature [14]. These authors state that impact ionization is possible but only at fields above 7 MV/cm and in oxides with thicknesses above 25 nm. Moreover, they believe that the created hole is detrimental in the breakdown process because of the creation of traps. In a recent publication [19], time to breakdown measurements are published (reproduced in Fig. 8), where at fields as high as 12 MV/cm the experimental values are explained in terms of impact ionization by energetic carriers in the SiO₂ conduction band. The model is based on the work of DiMaria and coworkers [20] on trapping phenomena, also at these high fields. However, some critical remarks have to be made. In the first place it seems difficult to decide that two mechanisms are needed to explain the experimental results in Fig. 8, taking into account the spread in the data points. In the second place the breakdown and trapping results may suffer from the inhomogeneous charge injection at high current levels, i.e. at high fields, as is discussed with the aid of the data in Fig. 6. We refer also to [21]. Regardless of whether impact ionization or inhomogeneous charge injection or both cause the drop of QBD, the data from the high current/voltage region are not relevant for extrapolations to practical use conditions at lower fields.

A fourth possibility for trap creation in the breakdown is by field emission of electrons inside the SiO₂ layer [22, 23]. This mechanism serves as the physical basis for a statistical description of breakdown that goes back to the work of Suhe et al.[24]. In this description the capacitor area is split up into cells in which traps are generated at random. Two parameters are needed to describe a time-to-breakdown distribution, namely a critical cell area and a critical number of traps in a cell that is needed...
to trigger breakdown. This model description is able to fit the experimental results on oxides in the range 2.5 nm to 22 nm published by Schuegraf and Hu [9] and for which these authors have a different physical model (see above). Although we cannot exclude this mechanism of internal field emission, we think it is an unlikely process in the breakdown, because Poorter and Wolters [25] have shown that breakdown can occur at very low fields when there is another way of electron injection into the oxide. Moreover, the parameters in the statistical description have not yet been verified in independent measurements.

This brings us to the final phase in the breakdown process, the destructive rupture of the oxide. In the preceding phase we have the creation of traps by one or more mechanisms, as discussed above. These traps or broken bonds form a connecting chain of defects that eventually result in a low ohmic path. This can be detected by measuring the leakage current through the oxide after a high field stress. DiMaria [26] has measured these stress-induced leakage currents in thin oxides (in the range 3.7 nm to 5.5 nm) as a function of stress voltage. It was found that the onset of the leakage currents occurs at the same voltage of 5 V as the creation of (neutral) electron traps, but below the voltage at which anode hole injection becomes measurable [16]. DiMaria concludes that neutral traps are formed by hot electrons and that these traps form stepping stones for the tunnelling carriers contributing to the leakage current. However, it should be noted that in these thin oxides breakdown can occur at voltages below 5 V and which is still of intrinsic nature. This proves that the relation between stress-induced leakages and breakdown is not a trivial one. More investigations are needed here. Nevertheless, we believe that the formation of a low ohmic path is a step in the breakdown process. The capacitor is then discharged through this path, resulting in the final rupture [27].

In summary, the following steps can be seen leading to intrinsic breakdown:

- electron injection (at cathode);
- transport through oxide, some trap creation;
- energy dissipation at anode, breaking of bonds starting at anode;
- hole injection, transport, trap creation;
- formation of low ohmic path;
- discharge of capacitor.

6. Defect-related breakdown

So far we have described the intrinsic breakdown mechanism. The question arises: what is the mechanism of defect-related breakdown? Very probably this is caused by irregularities in the oxide layer (dust particles) or at the Si-SiO₂ and SiO₂-M interfaces (spikes). Another possibility is the presence of metal contamination in the SiO₂ that lowers the energy barriers, in this way increasing the current at low fields.

It appears that all these effects can be modelled by a certain oxide thinning [28]. This is sketched in Fig. 9. The oxide thickness \( X_{ox} \) is, at the defect or weak spot, reduced to \( X_{eff} \).

Then, the application of a voltage \( V_{ox} \) gives a field \( V_{ox}/X_{eff} \). Consequently, the time to breakdown, \( t_{BD} \), of an oxide sample is characterized by the thickness, \( X_{eff} \), of its weakest spot:

\[
t_{BD} = \tau_0 \exp \left( \frac{G X_{eff}}{V_{ox}} \right)
\]

The quantity \( G \) is the field acceleration factor. The values of \( X_{eff} \) have a certain distribution depending on the properties of the defects. This explains the observed large distribution in the breakdowns. Recently, Schuegraf and Hu [9] rediscussed this model. From their experimental results on the very thin oxides they concluded that the model of eq. (3) is able to describe the extrinsic part of the breakdown distribution of these oxides. This implies that the time to breakdown \( t_{BD} \) is an exponential function of the
Fig. 9. Model for defects in oxides in terms of oxides in terms of oxide thinning. $X_{\text{eff}}$ is the effective oxide thickness at a weak spot in an oxide with overall thickness $X_{\text{ox}}$. Reproduced from [28]. Copyright © 1988 IEEE.

inverse oxide field. This is understandable in view of eq. (2) describing the injected electron current in the oxide, which is also an exponential function of the inverse oxide field.

However, the other part of the breakdown distribution, namely the intrinsic part, has a different field dependence. Suehle et al. [29] and Prendergast et al. [30] found a simple exponential dependence of $f_{BD}$ on oxide field (E model). The same was found by Dumin et al. [22] and Scott et al. [23] by working on the intrinsic breakdown data of [9]. It has been pointed out [30] that evaluation of the proper field dependence model is possible only by including data on intrinsic breakdown at oxide fields below 7 MV/cm, which implies relatively long testing times and testing to 100% failures. It has been suggested, based on this dependence and on the temperature dependence, that the mechanism for extrinsic breakdown is different from that for intrinsic breakdown [30]. Hence it is not a difference in injection condition, i.e. local injection at oxide thinning or other irregularity vs. homogeneous injection, with in both cases a similar sequence of steps leading to breakdown. Contrary to this we have to conclude that these steps may be different for the extrinsic breakdown from the ones described in Section 4. Anyhow, more experiments are needed to find the actual mechanism.

7. Conclusions

We have reviewed several methods for the testing of breakdown in oxides. It is shown how the different test methods reveal the same distributions: an extrinsic and an intrinsic one. The difficulty is the lifetime prediction. For the thicker oxides (>20 nm) the charge to breakdown, $Q_{BD}$, measured at low stress current densities, offers possibilities for lifetime prediction. This is not possible for oxides below 20 nm, because $Q_{BD}$ becomes a strong function of the electric field. Then determination of the field acceleration factor and the use of a proper model is necessary for extrapolation to practical use conditions.

Concerning the physical mechanism of dielectric breakdown in oxides, we conclude that several steps are operative in the total breakdown process. After the first step of electron injection, trap creation follows in the oxide. This can occur by accelerated electrons or by anode hole injection or, less likely, by holes from impact ionization. Anyhow, the holes appear to be detrimental in the process of trap creation and bond breaking. We feel that there is a certain convergence of this part of the mechanism based on the elaborate studies on injection, transport and scattering of charge carriers in the oxide.

The trap creation and bondbreaking lead to a conductive path in the oxide layer through which the capacitor is discharged and which leads to the final rupture.

The background of the defect-related breakdown is not at all clear. Possibly it starts with an inhomogeneous injection of electrons, but the following steps differ from the intrinsic break-
downs because of differences in the observed field and temperature dependence.

Acknowledgements

The first author (J.F. Verweij) thanks the organizers of the 6th ESPRIT Workshop on Growth and Characterization of Thin Dielectrics for Microelectronics for the invitation to present an invited talk at this conference. The contents of this presentation form the basis of this paper. Moreover, the authors thank the eight referees of this paper for valuable comments which contribute greatly to the final contents of the paper. The support of the Foundation for Fundamental Research on Matter and the Netherlands Technology Foundation is gratefully acknowledged. Finally, the careful typing of the manuscript has been carried out by Mrs. Margie Rhemrev.

References


