ELSEVIER SCIENCE Ltd. [DTD 4.2.0]
JOURNAL MR ARTICLE No. 1947
PAGES 1-5 DISPATCH 7 April 2001
MR 1947
A
PROD. TYPE: FROM Disk

MICROELECTRONICS RELIABILITY

PERGAMON

Microelectronics Reliability 000 (2001) 000–000

www.elsevier.com/locate/microrel

Dielectric breakdown distributions for void containing silicon substrates

R. Falster [a,*], F. Bonoli [a], V.V. Voronkov [b]

[a] *MEMC Electronic Materials SpA, Viale Gherzi 31, 28100 Novara, Italy*
[b] *MEMC Electronic Materials SpA, Via Nazionale 59, 39012 Merano, Italy*

Abstract

Distributions of gate oxide failure in various types of silicon substrate materials have been investigated for a wide range of oxide thicknesses. Silicon substrates containing various well-characterized void distributions along with defect-free materials were tested using special low-series resistance capacitor structures. Results of both ramped field tests of variable ramp rate and constant field tests were performed and analyzed within the framework of Weibull statistics. Ramped field tests are not "time zero dielectric breakdown" tests as is commonly asserted. They can in fact be very useful in extrapolating time dependent failure. The same set of Weibull parameters can be used to describe both ramped field and constant field wearout tests if an appropriate model for the time dependent damage accumulation during the field ramp is used. There are implications for reliability predication and the burn-in screening of device populations containing such defects.

1. Introduction

Void microdefects in silicon Czochralski (CZ)-grown silicon are the origin of one of the technologically most important modes of dielectric breakdown. Most CZ-silicon wafers contain a low number ($\approx 10^5$–10^7 cm^{-3}) of small ($\approx$100–150 nm) octahedral voids [1]. These voids are the result of the precipitation of supersaturated lattice vacancies during crystal growth from the melt [2] and do not exist in CVD epitaxial silicon. Much has been learned recently about the formation of such defects and CZ crystal growth techniques have been developed which produce void free silicon known as perfect silicon [3]. They are known by a variety of names depending largely on the method of their detection: flow pattern defects, d-defects, light scattering tomography defects, crystal originated particles and even "gate oxide integrity" (GOI) defects. It is well established that void microdefects can lead to premature dielectric breakdown failure [4]. We present a short summary of the dielectric breakdown distributions produced by different populations of voids under various test and capacitor configurations. Our purpose is the development of a more rational assessment of material reliability and device burn-in criteria of such materials. Void defect distributions form a very well-controlled defect population and are themselves an interesting and useful model system for the investigation of statistical models of dielectric breakdown. Weibull-type analyses of the data are performed. By this means it is shown that data from constant field and ramped field measurements may be connected by a simple parameter set. A model for damage accumulation is developed which allows for this.

2. Experimental results

Ramped field and constant voltage and field dielectric breakdown distributions were determined for both non-void and void-containing silicon substrates with well-controlled void distributions using very low-series resistance polysilicon gate capacitor structures with front and back aluminum metallization. The dielectric breakdown response to a given population of voids depends on the oxide thickness. Various combinations of

* Corresponding author. Tel.: +39-3-2133-4394; fax: +39-3-2169-1000.
E-mail address: rfalster@memc.com (R. Falster).

0026-2714/01/$ - see front matter
PII: S0026-2714(01)00048-8

void populations and oxide thicknesses were probed using combinations of variable field ramp rate and constant field tests to build-up a complete picture of the dielectric breakdown response in time and electric field. Various capacitor areas were used in the course of these studies, but all of the data illustrated here is for 0.1 cm^{-2}.

2.1. Basic void-related breakdown distributions

Fig. 1 illustrates the most basic case and shows simple linearly ramped (at 0.5 MV/cm/s) field breakdown distributions for 3 sample groups: two different void containing (A,B) and one non-void perfect silicon group (C). The oxide thickness in this case was a thick 50 nm and the capacitor area was 0.1 cm^{-2}. The sample sizes were 830 capacitors for groups A and B and 332 for group C.

Two clearly separated regions of dielectric breakdown are observed: a high field and a low field distribution. The high field response was common to all sample groups – suggesting a common non-void origin of this breakdown mode. The low field response is present only in void containing groups. In the void containing samples, failures begin to be detectable at about 2.2 MV/cm. The numbers rise until a saturation of failures of this type is reached for the two groups. This saturation corresponds to the exhaustion of the number of voids intersecting the wafer surface. At this point all of the voids present have resulted in failure. Groups A and B contain a void density of 1×10^6 and 0.4×10^6 cm^{-3} respectively. The void density is controlled by varying the cooling rate of the crystal [2]. Group B was from a more slowly cooled crystal. The data of Fig. 1 illustrate that groups A and B contain saturated GOI defect densities of 15 and 6 cm^{-2} respectively. Assuming average void sizes of about 150 nm [1] (we neglect the consumption of silicon by the growing oxide), this corresponds well with an estimate of the area density of voids intersecting the sample surface. The fact that the two void populations show similar distributions of failure at low fields prior to saturation has implications about the void-size dependence of dielectric failure distributions. This will be discussed elsewhere.

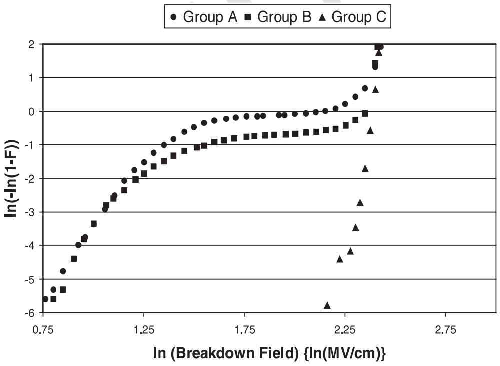

Fig. 1. Typical Weibull plots of the dielectric breakdown distributions of three sample groups: two different void containing (A,B) and one non-void material (C). Results are plotted against the natural logarithm of electric field – for reasons of Weibull analysis.

2.2. Effect of oxide thickness

At larger oxide thicknesses, the two modes of breakdown, void and non-void are widely separated in electric field. This separation becomes less with thinner oxides.

Fig. 2 illustrates the results of tests made on the void population of group A for several oxide thicknesses between 5 and 50 nm.

From the data of Fig. 2, it can be seen that the main effect of oxide thickness on the distribution of dielectric failure resulting from a given population of voids is to shift to higher values the average field at which the voids are "activated" – that is, converted into a "GOI defect". The saturated density of the ultimately activated voids remains the same regardless of oxide thickness. That is to say all voids eventually cause dielectric failure. We note in passing, however, that the rate at which the average field increases with decreasing oxide thickness is less than $1/t_{ox}$. This means that the average voltage at which voids activate decreases with decreasing oxide thickness. This may have implications for device V_{dd} scaling and burn-in screening for this failure mechanism.

2.3. Time dependence of void-related failure

Data from such ramped field tests are often referred to as "time zero dielectric breakdown". They are nothing of the sort. Ramped field tests implicitly contain time elements, and thus useful information, through the component of damage built-up during the ramp-up in

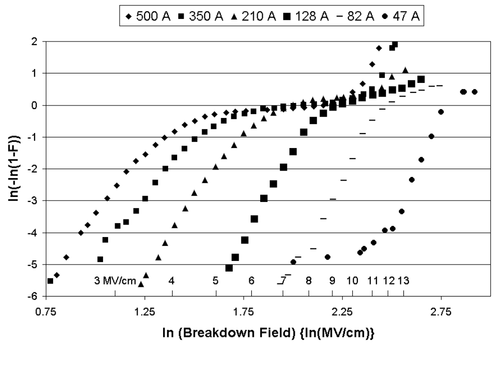

Fig. 2. Breakdown distributions for void-containing sample group A at various oxide thicknesses.

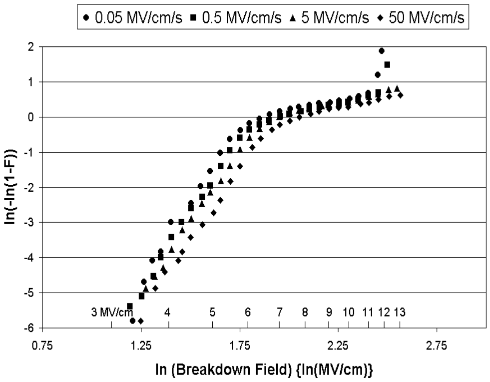

Fig. 3. Variable ramp rate measurements performed on sample group A.

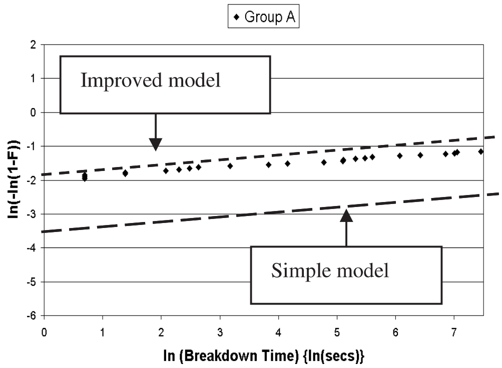

Fig. 4. Constant voltage (10 V) failure distribution (dots) for void group A using 21 nm oxides. Predicted distributions from the data of Fig. 3 using a simple and improved model for damage accumulation.

field, just as wearout tests contain information on the stress-level dependence. That this is so is easily seen in the void case by simply varying the ramp rate of ramped field tests on a given population of voids. Fig. 3 shows the breakdown distributions for the void population of group A tested using 21 nm oxides and field ramps ranging from 0.05–50 MV/cm/s.

Properly analyzed, ramped field tests can provide a wealth of very useful information. In order to assess the reliability implications and burn-in strategies for dealing with a particular mode of breakdown, usable information on both the field and time distribution of failure is required. Such tests can yield information on both over a very wide range of electric field; it is essentially a parallel measurement capable of accessing field and time information even at very small stress levels. Serial wearout tests at variable stress levels can be time consuming in the extreme.

We analyze such data within the usual two-parameter Weibull distribution which we write as:

$$1 - F = \exp(-Ct^a E^b) \qquad (1)$$

From this we get the useful linear Weibull equation:

$$\ln(-\ln(1-F)) = a\ln(t) + b\ln(E) + \ln C \qquad (2)$$

This power law formalism gives a representation of the increase in effective defect density over time, t, at some fixed value of electric field E; in other words, it is used to describe the results of constant stress tests. The slopes of a Weibull plot for constant stress yields the time parameter a. Fig. 4 shows an example of data from such a measurement performed on the void-containing group A with an oxide thickness of 21 nm and an applied voltage of 10 V. The slope of the best fit line to the data is about 0.15. The dashed lines are discussed later.

Constant stress is but a special case of the more general problem. In order to model the more general case of time dependent stress we need to have a model of how "damage" is built-up at defect sites. For the case of a linearly ramped field, the most straightforward approach to Ref. [5] is to simply integrate Eq. (1).

$$1 - F = \exp\left(-C\int_0^t at'^{(a-1)}E^b \, dt'\right) \qquad (3)$$

For a linear ramp $t = E/(dE/dt)$, and the expression becomes

$$1 - F = \exp(-C[a/(a+b)]E^{a+b}(dE/dt)^{-a}) \qquad (4)$$

The Weibull plot for this is:

$$\ln(-\ln(1-F)) = (a+b)\ln E - a\ln(dE/dt) + \ln C + \ln(a/(a+b)) \qquad (5)$$

Thus the vertical shift in a ramped field Weibull plot resulting from a change in the ramp rate should be equal to a times the logarithm of the ratio of the two ramp rates. Applying this to the data of Fig. 3 and a value for "a" from ramped field testing is indeed found to be very similar to that of the constant stress test. The most basic feature of the time dependence can thus be found in ramped stress tests. However, this simple analysis incorrectly predicts the build-up of damage during the ramp. Using the parameters extracted from Fig. 2 and Eq. (5) and plotting the result on the constant stress data of Fig. 4 shows that the "time zero" fails are grossly underestimated using this simple model.

Weibull statistics make no claim to model any particular physical effect. They simply describe data in a convenient and useful way. In this spirit, it is found that the Weibull formalism that does indeed describe the data of the two types of test results together is a slight modification of Eq. (4), namely:

$$1 - F = \exp(-C[a/(a+b)]^a E^{a+b}(dE/dt)^{-a}) \qquad (6)$$

The only difference between this expression and that of Eq. (4) is the exponent "a" in the $[a/(a+b)]^a$ part. Using this formalism together with the data of Fig. 3 yields excellent fits to constant stress tests. An example of this is shown of Fig. 4 ("improved model").

This is a very useful equation to have. The parameters for any void population (at whatever temperature is of interest) can be extracted from ramped field tests. Eqs. (1) and (6) then form the basis of a reliability extrapolation.

What is now left to do is to derive an expression for damage build-up during field ramps which matches the experimental observations of constant stress measurements. We do this simply by working backwards from Eq. (6), to find the kinetic equation for damage build-up necessary to produce it.

2.4. Damage build-up during ramped field stressing

The fundamental quantity in this problem is the probability $F_v(E)$ for an interfacial void to have already caused a breakdown by the time the field has reached a particular value E. More generally, the relevant quantity it is not E itself but some other quantity which depends on the whole 'field history' $E(t)$. The probability F_v is also dependent on the average void diameter D and oxide thickness w.

At a particular moment in time some voids, (the fraction $1 - F_v$) are still safe (sub-critical) while all the others, (the fraction F_v), have already caused a breakdown. The surface density of 'bad' voids is ρF_v where ρ is the total surface density of voids (ρ equals DN_v where N_v is the bulk density of voids). The probability for a capacitor, of area A_{goi}, not to have any bad voids (and so not to have failed yet) is $\exp(-\rho F_v A_{goi})$. Accordingly, the probability of a capacitor breakdown already having happened by a given point in time, due to the presence of one or more bad voids, is:

$$F = 1 - \exp(-\rho F_v A_{goi}) \tag{7}$$

This quantity, F, is just what is measured in GOI tests and processed using the Weibull formalism (F_v is proportional to $E^b t^a$) for the fixed field case.

The process of the accumulation of oxide damage under applied field can be formally described by a kinetic equation leading to the Weibull expression at constant E. Following the Weibull path, the equation for a 'damage amount' W should be of the power law type

$$\mathrm{d}W/\mathrm{d}t = qE^d/W^g \tag{8}$$

where q is a constant prefactor. The rate of damaging is strongly enhanced by the field but strongly retarded by already accumulated damage W.

Adopting for the moment the formal kinetic equation (8) and integrating it by shifting the W^g factor to the left-hand side one obtains the damage W as a function of the field history:

$$W = p\left[\int E^d\,\mathrm{d}t\right]^a \tag{9}$$

$a = 1/(g+1)$ and is a very small exponent – in this case. The prefactor p is a combination of the previously introduced parameters: $p = (q(g+1))^a$.

For constant field the damage is

$$W = pE^b t^a \tag{10}$$

where $b = da$. For the ramped field case:

$$W = p(\mathrm{d}E/\mathrm{d}t)^{-a}[a/(a+b)]^a E^{a+b} \tag{11}$$

The interfacial void geometry is strongly scattered due to the variable offset of the intersecting plane with the distributed voids in the bulk of the wafer. The ability of a void to cause a breakdown at particular damage W is strongly scattered too. We can introduce the distribution function of voids over the critical damage W^* causing a breakdown: $Q(W^*)\,\mathrm{d}W^*$ is the probability of the critical damage to be within $\mathrm{d}W^*$. This means that the breakdown event occurs when W is within the $\mathrm{d}W^*$ interval. The quantity we are interested in, the cumulative breakdown probability F_v, is then expressed as

$$F_v = \int Q(W^*)\,\mathrm{d}W^* \tag{12}$$

where the integration is from 0 to the actual value of the damage, W. We do not know what is the distribution function $Q(W^*)$ is. The simplest assumption is that it is a broad function – in comparison to the actual value of W – so that we can replace $Q(W^*)$ by just a constant, $Q(0)$ in the integral (12). Then F_v is proportional to W, and the expressions (10) and (11) are applicable to F_v, with a different prefactor only. For constant field the damage is

$$F_v = rE^b t^a \quad \text{with } b = da \tag{13}$$

and for ramped field it is

$$F_v = r(\mathrm{d}E/\mathrm{d}t)^{-a}[a/(a+b)]^a E^{a+b} \tag{14}$$

with $r = pQ(0)$.

Within this description, both Weibull parameters a and b are not elementary constants but combinations of the damage exponents d and g entering the initial kinetic equation (8).

2.5. Some general comments on testing

The above discussion has laid the foundation for a rational assessment of void related reliability risk. Let us now make a few general comments. The first is to note that (because of the very weak time dependence of the void mode) the extrapolated time dependence to operating conditions of void-related failures (or reliability risk) has very little to do with the actual void defect density. This is the number that comes from the failure percentage in the plateau in the breakdown data, the "saturated" defect density. Complete conversion of voids into "GOI defects" happens only at high stress or very slowly over very long nearly geological periods of time. It is the initial tail of the distribution that is important. The important part of the data from a reliability point of view actually lies in the rising part of the Weibull plot, more specifically in its intersection with the intended operating point.

Be that as it may, it is however, with the saturated value or some other high field value – and not the field distribution – that most testing and material evaluation generally focussed on. Constant stress or t_{db} or Q_{bd} tests are often applied to this kind of problem (and metal contamination effects are another type of problem to which the present discussion applies). Such tests are nearly always performed at rather high stress levels, for obvious reasons: only here are measurable signals generally to be found. The only piece of information gleaned from such measurements is, at best, the void density – at worst some rather arbitrary number somewhere in the middle of the distribution. Ramped test results are often screened by simple, arbitrary assignments of "A or B or C mode" failures in which "defect density" is defined at a fixed field level (say 8 MV/cm for "B" mode). This also misses what is actually the most important information in the distribution. Burn-in screening based on such principles can easily result in large numbers of perfectly good parts of the distribution being screened out.

3. Conclusions

Voids resulting from the agglomeration of vacancies during the growth of silicon crystals are often a major component of the dielectric breakdown response of capacitors produced on polished CZ-silicon substrates. The details of the dielectric breakdown response of void containing CZ-silicon wafers has been investigated using specially developed tests procedures which allow for the reliable collection of data over a very wide range of fields.

Unlike many other modes of dielectric breakdown, the voids which are the source of this particular response are finite in number density and thus saturation is observed in the breakdown response with increasing field and/or time. This corresponds to the point at which all the voids intersecting the oxide/silicon interface having been "activated", resulting in a breakdown event. The average "activation" field for a particular distribution of voids increases with decreasing oxide thickness. The rate of this increase is, however, not fast enough to maintain at least a constant "activation" voltage with decreasing oxide thickness. The saturated or total density of voids in a sample set plays nearly no role whatsoever in the extrapolated reliability resulting from a given population of voids.

Voids exhibit a distinctive mode of dielectric breakdown characterized by certain field and time parameters. These can be extracted from Weibull analyses of the dielectric breakdown response of both ramped field and constant stress tests, provided a correct model of the build-up of damage is used in making the connection between constant and variable stress testing.

Used and analyzed properly, the use of variable ramp rate field stressing (in conjunction with constant stress tests) can be a powerful tool for reliability extrapolation – including temperature dependencies. This is true not only for void distributions – the high-field mode can also be extrapolated in this way if proper damage build-up models are produced for this mode of breakdown too. Series resistance is a major experimental issue when attempting this however. Populations with multiple modes can be easily handled with such an approach; the use of probe capacitors of variable area are useful in this regard. Our experiences with this will be reported elsewhere.

References

[1] Ueki T, Itsumi M, Takeda T. Jpn J Appl Phys 1997;36:1781.
[2] Voronkov VV, Falster R. J Cryst Growth 1998;194:76.
[3] Falster R, Voronkov VV. MRS Bull 2000;25:25.
[4] Park JG, Lee GS, Park JM, Chon SM, Chung HK. ECS Proc 1999;99(1):324.
[5] Falster R. J Appl Phys 1989;66:3355.