GATE-CONTROLLED SURFACE BREAKDOWN IN SILICON $p$--$n$ JUNCTIONS

by H. C. de GRAAFF

Abstract

In integrated circuits and MOS transistors made in silicon by the planar technique, the situation often arises that a metal electrode on top of the oxide layer passes over an underlying $p$--$n$ junction. In that case the breakdown voltage ($V_{BR}$) of the $p$--$n$ junction can be modulated by the voltage ($V_g$) of the metal electrode. In this paper measurements on MOS transistors are given, which show that in the region where the surface is depleted, the breakdown voltage is given by $V_{BR} = V_g + \text{constant}$. The constant is about proportional to the square root of the oxide thickness. A theoretical model is proposed which leads to good agreement with the measurements.

1. Introduction

Garrett and Brattain\(^1\) have shown that charges present at the surface of a $p$--$n$ junction can alter the electric field in the depletion region because of field lines terminating at these surface charges instead of at ionized impurities in the silicon itself. The same is true for a metal-gate electrode on top of the oxide layer covering the silicon surface\(^2\)\textsuperscript{--}\textsuperscript{5}). Hence the gate voltage can influence the breakdown voltage of the $p$--$n$ junction.

In this paper the relation between breakdown voltage and gate voltage is studied experimentally as well as theoretically. The influence of substrate doping and oxide thickness is investigated. In contrast to the result reported by Grove et al.\(^5\), it is found that within wide limits the substrate doping and oxide thickness has no influence on the slope of the $V_{BR}$\textsuperscript{--}$V_g$ curve.

The experiments were carried out on MOS transistors, but it must be stressed that the same phenomenon is encountered in integrated circuits, where metal strips pass over diffused isolation walls. This effects limits in MOS transistors the maximum permissible drain voltage and in integrated circuits the reverse voltage of the isolation wall.

2. Measurements

MOS transistors with a circular geometry were used in the experiments (see fig. 1). The substrate was $p$-type of various dopings, the source and drain contacts were heavily doped $n$-type, the gate electrode was aluminium. One side of the junction (regions A in fig. 1) is not overlapped by the gate, making bulk breakdown in this region independent of gate voltage. In this way a clear distinction can be observed between surface breakdown and bulk breakdown. The breakdown voltage was measured with the substrate grounded and source
Fig. 1. Cross-section of the circular geometry of the MOS transistor. A is the junction region, not covered by the metal gate electrode, B is the covered region.

and drain interconnected, and at a constant current. Because all transistors used in the experiments showed very sharp breakdown characteristics, the value of the constant breakdown current is of little importance. It was usually 10 µA. The source-drain distance was 17 µm except in one case, where 400 µm was taken. The latter did not, however, show a different behaviour.

In some cases the gate capacitance was measured as a function of gate voltage with the reverse voltage of the p–n junction as a parameter. The measuring circuits are given in fig. 2. In fig. 3 the results are given for various substrate dopings. The oxide thickness $x_0$ was 0.2 µm. In fig. 3a it can be seen, that, at high positive gate voltages, the breakdown voltage is independent of $V_g$. Breakdown occurs in the region near the source, which is not overlapped by the gate (region A in fig. 1). When the gate voltage is decreased below a certain value, $V_{BR}$ also decreases. Breakdown now occurs in the regions which are covered by the gate electrode (marked B in fig. 1). The slope $dV_{BR}/dV_g$ equals unity, until $V_{BR}$ tends to saturate at negative values of $V_g$.

From the $C-V$ curves shown in fig. 3a, it can be deduced that at the right of the line a–a the surface is inverted and at the left of the line b–b there is

Fig. 2. Circuits for measuring (a): $V_{BR} = f(V_g)$ and (b): $C = f(V_g)$. 
Fig. 3. The breakdown voltage $V_{BR}$ and the gate capacitance $C$ as a function of the gate voltage $V_g$ for various substrate dopings $N_a$. The oxide thickness $x_o = 0.2 \, \mu m$, the source–drain distance $L_{SD} = 17 \, \mu m$. The lines a–a give the onset of inversion, the lines b–b the onset of accumulation.
accumulation. In between these two lines the surface is depleted. Thus the linear part of the $V_{BR}-V_g$ curve refers to a situation in which the surface is depleted. At negative gate voltages, where the breakdown voltage tends to saturation, the surface is in accumulation.

In fig. 3b the situation is the same, except for the substrate doping, which is much higher now. This lowers the breakdown voltage in the region where the gate electrode is absent (region A in fig. 1). As a consequence the breakdown is shifted to the gate-controlled region (region B in fig. 1) at a much lower value of $V_g$. In fact the linear part of the $V_{BR}-V_g$ curve has disappeared almost completely, leaving only the “accumulated surface” portion.

In fig. 3c the substrate is quite heavily doped. Throughout nearly the whole $V_g$ range breakdown occurs in the region A of fig. 1 and the gate voltage has no influence.

Figure 4 shows the influence of the oxide thickness. The substrate doping is

![Graph](image-url)

Fig. 4. The breakdown voltage $V_{BR}$ as a function of gate voltage $V_g$ for various oxide thicknesses; $L_{SD} = 400 \mu m$. 
not too high in this case in order to have clearly distinguishable linear parts in the $V_{BR} - V_g$ curves. In fig. 4a the different oxide thicknesses ($x_0$) were obtained by different durations of the thermal oxidation process. This gives different values for the radius of the junction curvature.

In fig. 4b all units have had the same oxidation time as the 1-μm sample. The other $x_0$ values were obtained by means of chemical etching.

From fig. 4 it follows that the breakdown voltage at $V_g = 0$ varies approximately as the square root of the oxide thickness $x_0$. This is illustrated in fig. 5.

![Graph showing $V_{BR}$ vs $x_0$](image)

**Fig. 5.** The breakdown voltage $V_{BR}$ as a function of the oxide thickness $x_0$ at $V_g = 0$ V.

3. Theory

From the measurements it has become clear that the $V_{BR} - V_g$ curves can be divided into three different parts, namely a high, positive-gate-voltage region, a linear region and a negative-gate-voltage region.

(A) High positive gate voltages

In this region the breakdown voltage is independent of gate voltage. Breakdown occurs at the curved junction near the source, which is not covered by the gate electrode. The breakdown voltage is given by the doping of the substrate and the junction radius. This radius is about 3 μm for units with a thermally grown oxide layer of 0·2 μm thickness and about 12 μm for units with an oxide layer of 1 μm thickness. The results of the measurements are in good agreement with the theoretical values calculated by Sze and Gibbons \(^7\) for abrupt junctions.
(B) The linear region

The breakdown in this region occurs at the surface of the silicon substrate and becomes apparent when the related breakdown voltage is lower than the one caused by the junction curvature.

Normally the surface of p-type silicon would be inverted at positive gate voltages. However, in the presence of a reversely biased p-n junction, the inversion disappears and the surface becomes depleted, as pointed out by Grove and Fitzgerald 6). From the measured C–V curves it follows that in the whole linear region of the \( V_{BR} - V_g \) curve the surface is depleted.

In order to obtain more insight the following model was analyzed with the aid of a computer. A two-dimensional plane diode is considered. The oxide layer with thickness \( x_0 \) is covered by a metal electrode (see fig. 6). The heavily doped \( n^+ \) region is taken as an equipotential region with potential \( V_d \). The junction is assumed to be abrupt. The potential of the field-free region in the p-type substrate is taken as zero while the potential of the metal electrode is \( V_g \).

For this structure the two-dimensional Poisson equation is solved numerically with a digital computer using a relaxation method. As a first guess the voltages of a plane junction are prescribed at the mesh points, except for the right boundary. Here the voltages of a depleted surface are prescribed, taking the gate voltage \( V_g \) into account. The different relative dielectric constants of the oxide and the silicon (3.8 and 12 respectively) are taken into account by using a modified difference equation for the interface. Surface-state charges and oxide charge can be neglected in this case 8). This model is not suited for calculating the field strength in the corner of the \( n^+ \) region at the interface, because theoretically this field strength tends to infinity. In reality a thin depletion layer in the \( n^+ \) region keeps the field strength everywhere finite. The model can be used,
however, for calculating the equipotential curves. A few examples of such calculated equipotential curves are sketched in fig. 7. For $V_g < V_d$ it can be seen that the equipotential curves bend to the left in the corner near the interface. This increases the field strength in this area; the highest field strength will be encountered at the point where the interface and the metallurgical junction intersect. Avalanche multiplication will start here and the multiplied carriers are forced to move along the interface. The multiplication factor is determined by the integral of the ionization rate, taken along the interface.

![Diagram of calculated equipotential curves](image)

Fig. 7. Calculated equipotential curves for several drain and gate voltages and with various oxide thicknesses. The substrate doping $N_a = 10^{15} \text{ cm}^{-3}$. 
The ionization rate itself is a function of the x-component (along the interface) of the field strength at the interface. Breakdown occurs when the multiplication factor goes to infinity. Instead of integrating the ionization rate, one can use as an approximation the critical-field concept. That means in this case that breakdown will occur when the x-directed component of the field strength at the corner point \((E_{xx}(0))\) reaches the critical value \(E_{cr}\).
For \( V_g \geq V_d \) the field strength at the interface will decrease and breakdown will not occur at the interface, but somewhere in the bulk, e.g. at the junction curvature in our units.

By examining the plots of the equipotential curves the following became apparent:

1. For \( 0 < V_g < V_d \) the equipotential curve with \( V = V_g \) is almost a straight line, terminating at the gate electrode. In the coordinate system as indicated in fig. 6 this means that the surface potential \( V_s \) in the silicon at the interface is \( V_s = V_g \) when \( x = x_g \). The distance \( w_g \) is known from the plane diode and given by

\[
w_g = W \left\{ 1 - (V_g/V_d)^{1/2} \right\},
\]

where \( W \) = depletion width of the plane diode.

2. In the silicon, far away from the interface, the field strength in the \( y \)-direction \( (E_y) \) is zero. As an approximation we can say that \( E_y \) begins to deviate from zero at a distance \( k \) from the interface (dashed curves in fig. 7). This distance \( k \) is approximately independent of the position \( x \), and of the substrate doping, the oxide thickness and the applied voltages. The value obtained for \( k \) from the computer results is 1.5 to 2 \( \mu \text{m} \) (see fig. 7b, c and d). This leads to the introduction of the following approximations: for \( 0 < V_g < V_d \) the \( y \)-directed component of the field strength in the silicon at the interface \( E_{sy} \) is given by

\[
E_{sy} \approx \frac{V_s - V_g}{(\varepsilon_s/\varepsilon_{ox}) x_0},
\]

where \( \varepsilon_s \) and \( \varepsilon_{ox} \) are the permittivities of the silicon and the oxide respectively.

The derivative of \( E_{sy} \) in the \( y \)-direction is

\[
\frac{\partial E_{sy}}{\partial y} = \frac{V_s - V_g}{(\varepsilon_s/\varepsilon_{ox}) x_0 k},
\]

We suppose this to be also true at the corner \((x = 0)\). With this approximation the Poisson equation for the depletion region along the interface can be written as

\[
\frac{\partial E_x}{\partial x} = -\frac{d^2 V_s}{dx^2} = -\left( \frac{q N_a}{\varepsilon_s} + \frac{V_s - V_g}{(\varepsilon_s/\varepsilon_{ox}) x_0 k} \right),
\]

where \( q \) is the electronic charge and \( N_a \) is the substrate doping. The solutions of (4) give us the potential \( V_s \) at the interface as a function of \( x \). When \( x \) approaches the value of the depletion width \( W \), eq. (4) loses its significance, because assumption (3) is no longer valid.
From the solutions of eq. (4) it can be deduced (see appendix), that the \( x \)-directed field strength at the corner \( (x = 0) \) is given by

\[
E_{xc} = \frac{V_d - V_o}{\left\{ (\varepsilon_s/\varepsilon_{ox}) x_0 k \right\}^{1/2}},
\]

provided that the substrate doping is not too high and the oxide is not too thick \( (N_a < 10^{16} x_0^{-1/2}, \text{see appendix}) \). Under these conditions the breakdown voltage in the linear region is thus given by

\[
V_{BR} = V_o + E_{cr} \left( \frac{\varepsilon_s}{\varepsilon_{ox}} x_0 k \right)^{1/2}.
\]

This is the linear relationship between \( V_{BR} \) and \( V_o \), with slope unity, as known from the measurements. Moreover, at \( V_o = 0 \), the breakdown voltage is proportional to the square root of the oxide thickness. This result is also in good agreement with the experiments (see fig. 5). The encircled points in fig. 5 are the experimental results obtained by Grove et al. \(^5\)). Taking \( k = 2 \, \mu m \), one finds for \( E_{cr} = 6 \cdot 25.10^5 \, V/cm \), which is higher than normally encountered in plane junctions, but quite acceptable when field crowding occurs (see ref. 7).

\( \text{(C) Negative gate voltages} \)

At negative gate voltages the surface of the \( p \)-type silicon becomes accumulated and positive charges are drawn to the interface. They shield the electric field from the gate and make it less effective. This causes a reduction in the slope of the \( V_{BR}-V_o \) curves. At very large negative gate voltages it is even possible that the \( n^+ \) region becomes inverted at the interface \(^9\)). Then the depletion region no longer changes and \( V_{BR} \) becomes constant.

\( \text{4. Discussion} \)

In the foregoing the surface breakdown of silicon \( p-n \) junctions was studied. A quantitative model was developed which shows that the breakdown voltage is given by

\[
V_{BR} = V_o + E_{cr} \left\{ (\varepsilon_s/\varepsilon_{ox}) x_0 k \right\}^{1/2}
\]

within wide limits of substrate dopings and oxide thickness \( (N_a < 10^{16} x_0^{-1/2} \, \text{cm}^{-3}, x_0 \, \text{in } \mu \text{m}) \). The slope of the \( V_{BR}-V_o \) curve equals unity, while at a given \( V_o \) the breakdown voltage is proportional to the square root of the oxide thickness. These results are different from the results obtained by Grove et al. \(^5\)). This arises from the fact that they do not only measure a surface breakdown, but also a gate-voltage-dependent bulk breakdown. In their structure the gate overlaps the junction everywhere, so the gate voltage also modulates the junction curvature in the bulk.
Acknowledgement

The author wishes to express his gratitude to Mr. R. J. M. Verbeek, who took a very active part in programming the computer.

Appendix

The differential equation (4) must be solved with the following boundary conditions:

$$V_s = V_d \quad \text{at} \quad x = 0$$

and

$$V_s = V_g \quad \text{at} \quad x = w_g.$$  

This leads to

$$V_s = V_g + A \exp (x X_0^{-1/2}) + B \exp (-x X_0^{-1/2}) - X_0 \frac{q N_a}{\varepsilon_s},$$

with $X_0 = (\varepsilon_d/\varepsilon_{ox}) x_0 k$ and $A$ and $B$ given by the boundary conditions. The $x$ component of the field strength at the interface is given by

$$E_{sx} = - \frac{dV_s}{dx} = -X_0^{-1/2} \{A \exp (x X_0^{-1/2}) - B \exp (-x X_0^{-1/2})\}.$$  

At the corner ($x = 0$) we find:

$$E_{sx}(0) = E_{xc} = (B - A) X_0^{-1/2}.$$  

Inserting the values for $A$ and $B$ gives

$$E_{xc} = \left\{ \left( V_d - V_g + X_0 \frac{q N_a}{\varepsilon_s} \right) \coth (w_g X_0^{-1/2}) - \frac{X_0 q N_a}{\varepsilon_s \sinh (w_g X_0^{-1/2})} \right\} X_0^{-1/2}.$$  

In all cases encountered in our experiments $w_g \geq 2X_0^{1/2}$, so $\coth (w_g X_0^{-1/2}) = 1$. This reduces $E_{xc}$ to

$$E_{xc} \approx \left\{ V_d - V_g + X_0 \frac{q N_a}{\varepsilon_s} \left( 1 - \frac{1}{\sinh (w_g X_0^{-1/2})} \right) \right\} X_0^{-1/2}.$$  

If the second term between the brackets can be neglected in comparison to the first, the equation simplifies to

$$E_{xc} \approx (V_d - V_g) \left( \frac{\varepsilon_s}{\varepsilon_{ox} k} \right)^{-1/2}.$$
Omitting this second term is allowed when

\[ X_0 \frac{q N_a}{\varepsilon_s} \ll V_d - V_g \approx E_{cr} X_0^{1/2} \]

or when

\[ N_A \ll \frac{1.76}{X_0^{1/2}} \cdot 10^{16} \text{ cm}^{-3} \]

\((x_0)\) is the oxide thickness in microns).

In our experiments \(w_g \geq 2X_0^{1/2}\) for all cases. When \(w_g < 2X_0^{1/2}\) under some circumstances, \(\coth (w_g X_0^{-1/2})\) becomes dependent on \(V_g\) and the slope of the \(V_{BR}-V_g\) curve will be less than unity.

Eindhoven, April 1968

REFERENCES