Current model considering oxide thickness non-uniformity in a MOS tunnel structure

M.I. Vexler a,*, A.F. Shulekin a, Ch. Dieker b, V. Zaporojtschenko b, H. Zimmermann b,1, W. Jäger b, I.V. Grekhov a, P. Seegebrecht b

a A.F. Ioffe Physicotechnical Institute, Polytechnicheskaya Str. 26, 194021 St. Petersburg, Russian Federation
b Technische Fakultät der Christian-Albrechts-Universität zu Kiel, Kaiserstraße 2, D-24143 Kiel, Germany

Received 28 February 2000; received in revised form 19 June 2000; accepted 28 July 2000

Abstract

A problem of oxide non-uniformity in MOS tunnel structures on n-type Si with a 2–3 nm nominal SiO₂ thickness is investigated. Thickness distribution is described with Gauss law. Concepts of modeling of reverse current–voltage characteristics are formulated for metal-oxide-semiconductor (MOS) tunnel structures, regarding the effect of non-uniformity. The current crowding in relatively thin device sections is shown to result in a noticeable transformation of curves for electron and hole tunnel currents. Introduction of a mean effective thickness instead of the nominal one is possible only for the weak-inversion regime. The experimental part of this work includes the examination of samples by TEM and atomic force microscope methods, and their electrical characterization. Measured characteristics have been satisfactorily fitted by curves simulated for a standard thickness deviation of σ = 0.3 nm. © 2001 Published by Elsevier Science Ltd. All rights reserved.

1. Introduction

Recent advances in scaling conventional MOSFETs (e.g. Ref. [1]) prompted the global intensification of studies of metal-oxide-semiconductor (MOS) structures with an extremely thin – less than 3 nm – oxide layer.

Among the problems existing in this thickness range and waiting for careful consideration is oxide uniformity. Tunnel currents very strongly depend on the insulator thickness, so that even small local thickness deviations may be expected to cause a crowding of current in the thinnest device sections. This is not an exclusive problem of ultra-thin insulators, but for these it is heavily aggravated by generally large values of flowing currents.

Furthermore, for a tunnel charge transport mechanism, the sensitivity of current density on thickness change (evaluated by the logarithmic derivative ) grows with a reduction of nominal thickness dn. It is therefore quite normal to address the problem of gate oxide non-uniformity in connection with very thin MOS structures.

Not long ago, the authors of Ref. [2] have intentionally fabricated MOS structures (dₙ = 4–6 nm) with giant (more than ±0.8 nm) thickness variations and demonstrated that their direct (accumulation) I–V- characteristics do not coincide with the curves predicted by the casual formulas for any effective thickness dₑffective. This work is very interesting as an illustration of consequences of the thickness distribution effect (and it was the work [2] that inspired us to perform this study), but one must not forget that the ordinary values of thickness deviation are less than in Ref. [2]. It would be therefore more practical to analyze the same effects in thinner MOS structures where even smaller (realistic) non-uniformity is expected to be of great importance.
In our present work, we are studying the role of oxide thickness distribution in the formation of the reverse current–voltage characteristics of large-area tunnel MOS structures with a sub 3 nm nominal oxide thickness \( d_n \). Unlike the authors of Ref. [2], we consider the case of reverse bias (inversion), because it is a conventional operation mode for most MOS devices (MOSFETs, tunnel transistors [3,4]). The purpose of this investigation is to clarify how strongly the device characteristics might deviate from those of an ideal structure and also whether the introduction of some effective thickness \( d_{\text{eff}} \) might satisfactorily substitute the aspects of statistics.

2. Samples

For the experimental part of this work, two types of tunnel structures have been fabricated. The first one is simply a two-electrode MOS tunnel diode (Fig. 1a) and the second one is a MOS tunnel emitter transistor (Fig. 1b). The former configuration served for direct characterization of the thickness distribution while the latter was used for electrical measurements. Emitter area \( S \) was \( 1.3 \times 10^{-3} \) cm\(^2\) in diodes and \( 8 \times 10^{-6} \) cm\(^2\) in transistors.

The tunnel transistor (Fig. 1b) is topologically identical to a MOSFET, but acts as a bipolar device whose inversion layer (channel) is taken as a base, gate electrode as an emitter of electrons and Si bulk as a collector [4]. The p+-contact area provides an access to the inversion layer and the possibility of measuring its potential \( \phi \) in respect to the emitter. Additionally, one can separate the electron and hole tunnel components. The averaged (over the device area) electron current \( j_e \) can be found by measuring the current flowing through the collector electrode (C) while the measurement of the current \( j_{\text{ext}} \) flowing through the base electrode (B) at \( V > \phi \) gives the average value of the hole tunnel component \( j_h \). One should, however, note that the hole current can be found by such a way only for moderate \( V \) and \( \phi < 2.5 \) V, i.e. while no impact/Auger ionization occurs [4,5]. Besides, measured \( j_{\text{ext}} \) should be much larger than typical values of a thermal generation current.

All the samples were fabricated by oxidizing Si(1 0 0) wafers \( (N_d \sim 10^{16} \) cm\(^{-3}\)\) at low \( (700^\circ \text{C}) \) temperature in dry \( O_2 \). This is the most popular method of thin oxide fabrication today. Aluminum was used as gate material. Oxide parameters (nominal thickness \( d_n \), as well as its standard deviation \( \sigma \)) are assumed to be the same for diodes and transistors, on the basis of the identity of thin oxide growth conditions. Additionally, it was checked that the density of free-base collector current had been the same as the current density present in the diode (if the applied biases \( V \) are equal in both cases).

3. Characterization of oxide thickness distribution

Insulator thickness variations may be quantitatively described with a normal Gauss distribution. The values of standard thickness deviation \( \sigma \) reported in the literature for the oxide films grown in the furnace normally lie around 0.15 nm (e.g. Ref. [6]). A new exotic method of oxide formation in O3 [7] can provide even better oxide uniformity. But simultaneously one cannot exclude that \( \sigma \) will be somewhat larger in practice. Taking this in mind, we suggest \( \sigma = 0.1 \) and 0.3 nm to be reasonable values for modeling.

Fig. 1. (a) MOS tunnel diode structure used in this work. (b) MOS tunnel emitter transistor structure studied in our work (E-emitter/gate, B-base/source, C-collector/substrate). (c) Local energy band diagram of the structures shown in figs. (a), (b) with notation used in text. The values of \( V, \chi_m \) and \( \phi \) are the same throughout the entire device area.

Eventually worries about the fictive areas with negative thickness may then be inhibited. Of course, Gauss law formally admits any value of local thickness, but if the ratio of \( \langle d \rangle / \sigma \) exceeds 3.7, only \( 10^{-8} \) of the device area
has a negative thickness. If $\langle d \rangle = d_n > 1.5$ nm and $\sigma$ is less than 0.3 nm, one can evidently apply the commonly accepted Gauss law even without a re-normalization. It is also not necessary to consider the thickness distribution separately within a “native” oxide (if any) and in an oxide layer grown onto it, since for two Gauss distributions, corresponding $\langle d \rangle$ and $\sigma^2$ simply act as additive values.

To control whether our samples meet conventional assumptions regarding $\sigma$, we have undertaken their investigation by TEM, which yielded (Fig. 2a) the mean thickness $\langle d \rangle$ of about 2.6 nm and the dispersion $\sigma$ of about 0.1–0.2 nm. Regrettably, only a restricted number of measured values of local thickness has been obtained, so that an exact value of $\sigma$ cannot be reported. The estimation of $\sigma$ is done on the basis of the suggestion that the probability for a local oxide thickness to exceed $d_{\text{max}}$ in at least one of $n$ points is of the order of $1/n$. Mathematically, this is to be written as $1 - \Phi(\langle d \rangle - \langle d \rangle)/\sigma) \sim 1/n$, where $\Phi$ denotes the probability integral. The values of $d_{\text{max}}$ and $\langle d \rangle$ are known from experiment (they were equal to 2.89 and 2.59 nm, respectively). Such estimation is justified for the case of a small number of measured values [6].

We have therefore checked that the samples used in this work do not present anything extraordinary themselves and are quite regular. It is also important that very similar oxides had been used earlier in our operational devices whose performance was quite satisfactory [4].

One should, however, say that our TEM measurements have been performed just inside a small (compared to the device area) region. For this reason, the values of $\sigma$ found by TEM can underestimate (which is more probable) or overestimate the standard thickness deviation related to the whole structure.

In this work we are dealing exclusively with large-area devices. It means that their linear sizes $S^{1/2}$ (S-area) substantially exceed typical spatial measure $L$ of thickness deviation along the substrate.

In order to find the value of $L$ experimentally, local current measurements could be performed using a scanning tunneling microscope (STM). However, the exact estimate of $L$ is not essential for our purposes. It is enough to be sure, that there are no too slow changes of thickness, which would introduce a difference between the devices fabricated at different places of the substrate. For this reason, we decided to restrict ourselves to the atomic force microscope (AFM) measurements of the top oxide topography (Fig. 2b). This is a superposition of oxide non-uniformities and silicon surface roughness. But our comparative AFM measurements of an etched

![Fig. 2](image-url)
(i.e. without oxide) surface suggest a minor role of the roughness, as the Si wafer is more flat than the SiO$_2$ surface. Anyway we will never underestimate $L$ analyzing the top topography, since all the oxide thickness variations are involved. Having the profile “height”(x) recorded, 1/L (and thus $L$) may be evaluated, for example, by expanding this profile in a Fourier series.

AFM studies have led us to the conclusions that: (a) thickness changes at the distance of about $d_n$ are negligible; and (b) for the spatial measure $L$, one may adopt $\sim 0.1–1$ $\mu$m. In such a case, one should expect that all the devices will have identical characteristics, which will, however, deviate from those for an uniform oxide of nominal thickness $d_n$.

The small-area structures ($S^{1/2} < L$) remained beyond our consideration. For these, the statistical fluctuations of current from sample to sample could be suggested, which may eventually cause the threshold voltage fluctuations in MOSFETs [8].

4. Modeling of device characteristics in the case of inversion

Under inversion conditions, the behavior of MOS structures and therefore the occurrence of a thickness distribution effect are substantially more complex than in accumulation. The matter is that the insulator voltage in the inversion case is not uniquely determined by the applied bias $V$, but dramatically depends also on the rate of external supply of minority carriers toward the inversion layer (i.e. on the so called external control current $j_{ext}$, or base current [3]).

Further, we will focus on the structures with an n-type Si substrate, where the electrons are majority and holes are minority carriers. This is not only for definiteness, but also because such structures, unlike their counterpart, exhibit a large current gain when being exploited as tunnel transistors [3,4]. (Due to the asymmetry of tunneling probability, electron injection current always dominates.) All the results are, of course, transferable to the p-Si case.

Our analysis has shown that the inversion layer is to be considered equipotential. To clarify the situation, we calculated the distance $L_{inf}$ at which the perturbation of the local potential within the inversion layer would affect the neighbouring areas. This distance was estimated as $L_{inf} \sim (\lambda / g)^{1/2}$ where $\lambda = qN_s \mu_p$ ($\Omega^{-1}$) is the sheet conductivity of an inversion layer ($N_s$ is the surface hole concentration) and $g$ ($\Omega^{-1} \text{cm}^{-2}$) is the local “leakage conductivity” for holes. With realistic assumptions about the hole mobility $\mu_p$ and about tunneling in general, $L_{inf}$ proved to be substantially larger than the spatial measure $L$ of thickness change in our structures. For example, for a 2.0 nm oxide, estimated $L_{inf}$ exceeds 50 $\mu$m for all practical insulator voltages till breakdown, while the parameter $L$, evaluated in the previous section, lies within 0.1–1 $\mu$m.

After this remark, we can proceed with the simulations of tunnel currents. Local device characteristics may be generated, for example, using our one-dimensional model [9]. Such parameters, as tunnel barrier heights and effective masses are also to be borrowed from [9]. The equipotentiality condition dictates the requirement for the potential $\phi$ of the inversion layer in respect to the gate: the integral of the hole tunnel current $j_h$ over the oxide thickness $d_{ox}$ weighted with a normal distribution $f_N$ should be equal to the total intrinsic minority carrier supply plus the control current:

$$\int_0^{\infty} j_h(d_{ox}, \phi)f_N(d_{ox})d_{ox} = \int_0^{\infty} j_{int}(d_{ox}, \phi)f_N(d_{ox})d_{ox} + j_{ext}$$

(1)

The intrinsic carrier supply $j_{int}$ results from the local currents of thermal generation $j_{th}$, diffusion $j_d$ [3], impact ionization $j_i$ and Auger ionization $j_{Auger}$ [4]. After the position of quasi Fermi level for holes, i.e. the value of $\phi$, is defined by such a way, one can post-calculate the total gate current:

$$\langle j \rangle = \langle j_e \rangle + \langle j_h \rangle = \int_0^{\infty} j_e(d_{ox}, \phi)f_N(d_{ox})d_{ox} + \int_0^{\infty} j_h(d_{ox}, \phi)f_N(d_{ox})d_{ox}$$

(2)

Our method of treatment is seen to be in contrast to that for the case of accumulation [2] where the populations of both electrons and holes are characterized by an unique Fermi level throughout the whole semiconductor substrate. A very important novel point here is the necessity to find the position of the quasi-Fermi level for minority carriers by balancing their supplies and losses in the whole structure, i.e. not locally, that is expressed by Eq. (1).

5. Results of simulations

We have calculated the tunnel components: $\langle j_h \rangle(\phi)$, $\langle j_e \rangle(\phi)$ for a fixed applied bias $V = 5$ $V$ for $\sigma = 0.1$ and 0.3 nm (Figs. 3 and 4). As a reference, the dependencies for $\sigma = 0$ were generated. The characteristics were simulated for a nominal oxide thickness $d_{ox}$ from 1.6 to 3.2 nm with a step of 0.4 nm. Some unsmooth segments appear because of a limited accuracy of calculations.

The increase of $\sigma$ for a fixed $d_{ox}$ is seen to lead to the increase of all the currents at the same applied voltages. This is a consequence of dramatic current crowding in
the areas with small local oxide thickness $d_{ox}$ (Fig. 5). Indeed, as shown in Fig. 5, relatively thin device sections provide an unproportionally large contribution to the total current flowing through the MOS tunnel structure. Similar results may be obtained for any $V$ and $\phi$.

The deviation of the tunneling characteristics from those for $\sigma = 0$ will naturally cause the transformation of output and input characteristics of a transistor (Fig. 1b) both in the bipolar and in the field-effect mode.

Formal substitution of the oxide with parameters $(d_n, \sigma)$ through the ideal film with an effective thickness of $d_{eff}$ and $\sigma = 0$ may be done only for the case of weak inversion, i.e. of relatively low insulator voltage, while the tunneling probability roughly equals $\exp(-2h^{-1} (2m^*\chi)^{1/2} d_{ox})$ with $\chi = \chi_m$ for electrons and $\chi = \chi_m + E_g$ for holes (see Fig. 1c for notation). The interrelation between the effective and nominal thickness can be obtained analytically as

$$d_{eff} = d_n - h^{-1} (2m^*\chi_m)^{1/2} \sigma^2 = d_n - 0.5\sigma^2$$

assuming that $d_n$ and $\sigma$ are expressed in Angstroms. Here, $m^*(=0.3m_0)$ is the electron mass in SiO$_2$. How well this simple expression works can be seen in the numerically calculated curves in the range of small voltages between the gate and the inversion layer (Fig. 3). The expression explains the shift of characteristics from those of an ideal device ($\sigma = 0$). For our structures with $d_n = 26 \, \text{Å}$, $\sigma \sim 3 \, \text{Å}$, we have $d_{eff} = 21.5 \, \text{Å}$, and one may deduce from Fig 3 that the “ideal” electron tunnel current ($\sigma = 0$, $d_n = 21.5 \, \text{Å}$) would lie close to the experimental results.

The effective thickness $d_{eff}$ may not, however, be introduced for satisfactorily matching the characteristics throughout all practical conditions. Indeed, the curves generated for $(d_n, \sigma)$ are seen not to coincide with any curve generated for ideal devices except in the range of small voltages, as discussed above. The matter is that the slopes of curves for $\sigma = 0$ and $\sigma \neq 0$ substantially differ in the range of $\phi > 1.5 \, V$ (Fig. 3). This is not surprising since the tunnel current components strongly and non-trivially depend on many parameters. The absence of $d_{eff}$, that could enable to completely avoid the regard for
a statistics, was also shown in the paper [2] devoted to the accumulation case. For a very small \( \sigma \), however, one can ignore the discrepancy and always use the above-described simple approximation.

6. Comparison to experiment

Figs. 3 and 4 are completed with the experimental results for two devices. Data for each are shown with the same symbols on both figures.

One can immediately see that the electron and hole currents substantially exceed the values expected for an ideal structure with a nominal thickness of \( d_0 = 2.6 \) nm. At the same time, there is a satisfactory coincidence with the predictions of calculations for \( d_0 = 2.6 \) nm and \( \sigma = 0.3 \) nm, especially for the sample marked with circles. The experimental data for \( j_n(\phi) \) for another sample (squares) strongly deviate from the simulated curves at low \( \phi \). This may be due to technological defects enabling the parasitic non-tunnel hole transport. With this exception, the experimental dependencies can be fitted by simulated curves for \( \sigma = 0.3 \) nm fairly well. The majority of samples behaved as shown with circles. We have therefore to admit that the standard thickness deviation in our samples is, in fact, larger than its estimation on the basis of TEM measurements.

It is worth to mention here that all our previous experimental data obtained for a MOS tunnel emitter transistor with oxide layers formed within the same technological cycle as in this work, are the evidence for an important role of the thickness distribution. The most general result is that the measured currents are always larger than the theoretical values generated for the ellipsometrical insulator thickness (e.g. Ref. [4]).

To correctly compare the simulated and measured data, one should therefore know not only the value of a nominal oxide thickness \( d_0 \), but also its standard deviation \( \sigma \). Possible differences in the properties of SiO\(_2\), grown in different laboratories are just regarded by the value of \( \sigma \). Tunnel barrier heights and effective masses are thickness-independent. Eventual discrepancy between the experimental results and theory may be attributed to the defects causing the non-tunnel carrier transport, and non-Gaussian thickness distribution (e.g. local pinholes). Model restrictions can come into effect for a small (<1.5 nm) \( d_0 \) and/or a large \( \sigma \), when the equipotentiality condition for an inversion layer may not hold.

Simultaneously, one must note that the existence of thickness non-uniformity does not imply that the sample is out of order. The devices studied in this work demonstrated all the qualitative features which should be exhibited by a homogeneous MOS structure (e.g. bistability), like in Ref. [5].

7. Conclusion

In this work, the behavior of MOS tunnel structures with a 2–3 nm nominal oxide thickness under reverse bias has been analyzed with emphasis on the oxide thickness distribution effect. The following results were obtained:

(a) MOS tunnel diodes and tunnel transistors on moderately doped n-Si wafers have been fabricated. The standard thickness deviation \( \sigma \) (0.1–0.2 nm) found for our samples by TEM measurements, lies within ordinary limits. The spatial measure \( L \) (~0.1–1 \( \mu \)m) of the oxide non-uniformity was less than the device area.

(b) A procedure to model the reverse current-voltage characteristics of tunnel MOS structures considering the thickness distribution effect, has been put forward. On the basis of estimations, the inversion layer is suggested to be equipotential. A key point for simulations is the necessity to find the position of the quasi Fermi level for holes by balancing the flows of minority carriers.

(c) The impossibility of introducing the effective mean thickness \( d_{\text{eff}} \) (instead of considering the statistics) is demonstrated. A simple expression for \( d_{\text{eff}} \) was shown to be only applicable for the case of weak inversion.

(d) A strong current crowding in the device sections with much smaller than nominal oxide thickness has been predicted for \( \sigma = 0.2–0.3 \) nm.

(e) Measured curves for electron and hole tunnel currents were quite well fitted by the simulated ones for \( \sigma = 0.3 \) nm, so that we suggest our TEM studies underestimate, by some extent, the true thickness deviation.

Acknowledgements

Technical assistance of Ms. Asli is greatly acknowledged. This work was supported by the Grant 99-02-18079-a of the Russian Foundation for Fundamental Research and by the program “Nanostructures”. M.I.V. is grateful to the Alexander von Humboldt Foundation and to the GRACENAS center (St.-Petersburg) for support.

References


