THE HOT-ELECTRON PROBLEM IN SUBMICRON MOSFET

W. Hänsch, M. Orlowski, W. Weber

JOURNAL DE PHYSIQUE
Colloque C4, supplément au n°9, Tome 49, septembre 1988

THE HOT-ELECTRON PROBLEM IN SUBMICRON MOSFET

W. HÄNSCH, M. ORLOWSKI and W. WEBER

Siemens AG, Corporate Research and Development, Microelectronics, Otto-Hahn-Ring 6, D-8000 München, F.R.G.

ABSTRACT - A review of the hot electron problem in MOSFET is given. This includes: Key experimental features, the problem of modelling hot carrier transport in Si and SiO_2 after injection into the oxide, and an evaluation of technological measures to obtain hot carrier resistant structures.

I) INTRODUCTION:

Long-term stability is a key to the successful design of a MOSFET in the submicron regime. An important factor in achieving this aim is a quantitative analysis of hot electron effects. Hot carrier induced damage was known long before the submicron region was entered. Here we deal with the always present high energetic electrons (holes) that can surmount the Si/SiO_2 barrier of 3.2 (4.8) eV and then get trapped in the oxide. These charges now interfere with the controlling gate and will produce threshold shifts, transconduction changes, saturation current modifications etc. It is clear that all this can only be tolerated to a certain degree if the device is to operate properly in its circuit environment.

Secondly, we have the influence of the elevated fields on the carrier transport and the carrier injection into the oxide. It is a well known fact that the conventional drift diffusion approximation is only valid for small current levels, e.g. close to the thermal equilibrium.

The link between the latter and the former is that Poisson's, continuity, and current equations provide carrier distributions and electrical fields which are used to calculate the carrier injection into the oxide.

To prevent the undesired effects of hot-electron damage, a careful analysis of the experimental material is indispensable in an early state of device development. Measurement and simulation, as an analytic tool, have to go hand in hand.

There is quite a variety of effects caused by hot-electron (carrier) injection. Trapped electrons and/or holes in the oxide (Q_{ox}), fast surface states at the Si/SiO_2 interface (N_{it}), series resistance effects in LDD, mobility degradation due to oxide charges, time evolution of damage, etc. To monitor all these phenomena there are only four terminal currents measurable in a MOSFET: Source, drain, bulk, and gate current. Because of the high degree of complexity it is absolutely essential to have sufficient experimental data.

In the next chapter this review we will address the key experimental features for a hot-carrier analysis and their evaluation. In Chapter III a detailed discussion will be presented on modelling hot-electron effects on the basis of the drift-diffusion approximation and its extensions. In Chapter IV we will discuss how simulation is used to prevent hot-electron damage in submicron MOSFET structures.

II) THE HOT-CARRIER EFFECT FROM THE EXPERIMENTAL STANDPOINT

In the real operation environment, devices can change their properties to such a high degree that the functioning of the whole circuit may be affected after a certain time of operation. An example of this is shown in Fig. 1. The current-voltage behaviour and as a consequence the current-time path are changed after a certain time of operation.

In order to compare different technologies with respect to their hot-carrier stability the question of the right parameter arises. One may propose the threshold voltage shift at $V_G = V_D = 5V$ to qualify a technology as was done some time ago, since this is a circuit-relevant parameter. However, technological developments require a fast evaluation of the stability so that sensitive parameters were sought at the expense of circuit relevance, since the aforementioned one is not sufficiently sensitive. Empirically it was found that the characterization of the damage at low drain voltage, in the linear region, can provide the desired fast evaluation of device stability (Later on it turned out that for the inverter operation these new linear region parameters are very relevant and in fact proportional to the frequency shift of an inverter chain /1/).

Now the question arises whether we can, having found sensitive characterization parameters, immediately proceed to heuristic optimizations of technology parameters? Is there a conclusive way to obtain optimal drain profiles and oxide growth conditions to design a device with a high hot-carrier stability without a detailed physical understanding?

The issue has been approached in most labs in a more or less empirical way. However, a physical insight into degradation phenomena can provide guidance for optimization, accelerating it and making it more effective. This approach will be addressed in Chapters III and IV. Furthermore, the danger is too great that without understanding the physical background, false conclusions about stability are drawn for the operating conditions which can never be used in stress test experiments because of the slowing down of the hot-carrier effect at low operating voltages . As a consequence we conclude that a detailed understanding of the hot-carrier effect is necessary.
A second question arises: how far are we in understanding hot-carrier effects? Have we understood the basic features of the hot-carrier effect or are we still right at the beginning? In the following we try to obtain an answer to this question on the basis of a few experimental examples. These examples led to understanding but new open questions arose.

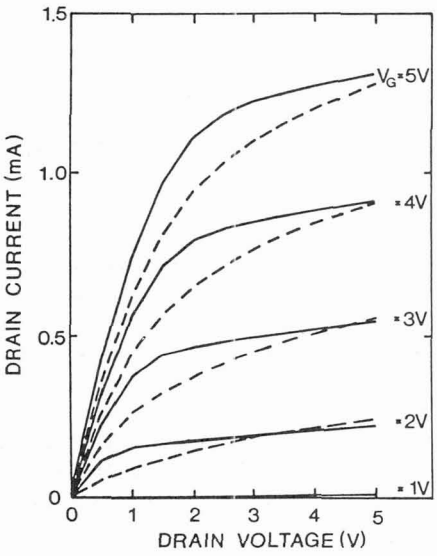

Fig. 1: $I_D(V_G,V_D)$ - characteristics before (full lines) and after (broken lines) a hot-carrier stress. A conventional nitride passivated 2.3um device with t_{ox}=42nm is stressed with t_{stress}=14h, V_D=8V, and V_G=3V.

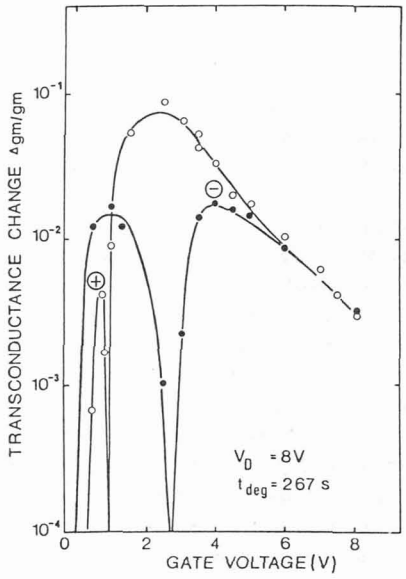

Fig. 2: Degradation vs. stress gate voltage is shown for devices as in Fig. 1 (open symbols) and devices from the same process but without a final nitride passivation (full symbols).

1) A few years ago a number of puzzling degradation results for different conventional n-MOSFETS led to a degradation model /2/ in which electrons and holes play an important part in explaining the resulting degradation. Holes are injected and trapped inside the oxide close to the interface and fixed positive charge is formed (visible by a positive $\Delta g_m/g_m$ in Fig. 2). Injected electrons are trapped with a large cross-section (due to Coulomb interaction) at the trapped holes and interface states are formed. These are evident through their negative charge in the negative $\Delta g_m/g_m$ branch in Fig. 2. A simple theory using rate equations was set up giving a reasonable fit to the experimental data with a proper choice of the two parameters of hole trap density and electron capture cross section. This model not only worked for different conventional MOSFETs (Fig. 2), but also for modern devices with reduced drain field such as LDD and DID.
Recently, however, R. Bellens et al. /3/ performed charge pumping measurements on devices under hole conditions (low V_G) and discovered the direct formation of interface states at the hole-injection condition. They claim that the subsequent electron injection merely causes a charge compensation. In order to decide finally on these findings, more data, especially space and energy resolved charge pumping measurements and 2D-simulations, are needed. Furthermore, essential features of the previous model, like the importance of the hole injections, are not questioned. - Nevertheless we can state that a precise degradation mechanism still remains unsettled.

2) A few years ago a relation ship between the lifetime and the substrate current of a device was found by Takeda and Suzuki /4/ and explained by C. Hu et al /5/. Although derived from an electron injection model, the result is also valid for a hole-injection model /6/. In Fig. 3 typical results are shown for different technologies, both LDD and conventional ones with a slope of 2.9 on the double-logarithmic scale. Agreement with the theory is good.

However, data have recently been measured at LDD devices which show a much steeper slope around 5 which cannot be explained by this theory (Fig. 3). Neither a technological reason is known for this difference nor a satisfactory physical explanation available for this phenomenon.

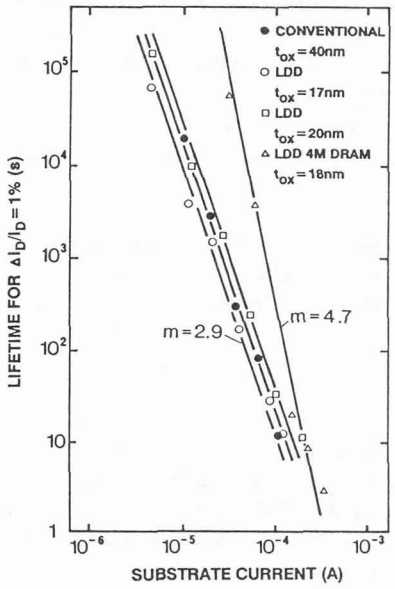

Fig. 3: Lifetime vs. substrate current for different technologies. One technology shows a steeper slope than the others.

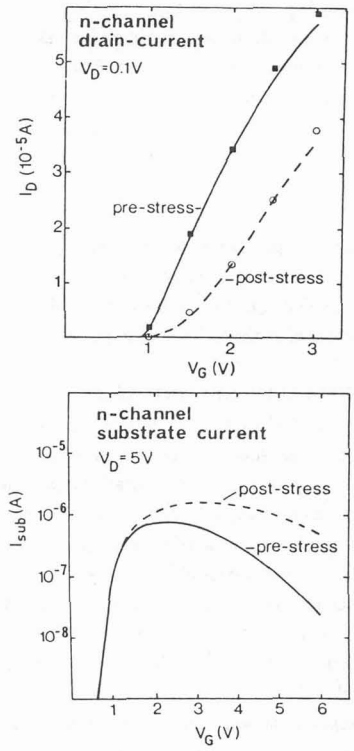

Fig. 4: Drain- and substrate currents of an n-MOSFET of the process in Fig. 1 before and after hot-carrier stress. The stress conditions were V_G=3V, V_D=8V and t_{stress}=5x10^{4}s.

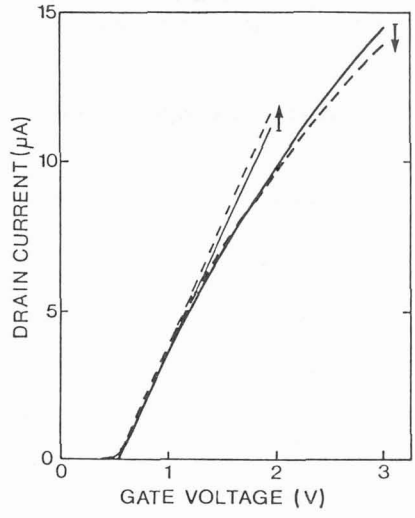

Fig. 5: $I_D(V_G,V_D$=0.1V) for a LDD-n-MOSFET with t_{ox}=25nm, l_{spacer}=0.2um and l_{gate}=1.0um before (full lines) and after (broken lines) a stress with V_G=0.6V, V_D=8V and a stress time of 12h. To clarify the result a straight line is drawn through the steepest parts of the $I_D(V_G)$-curve.

3) It can be seen in Fig. 1 how the current-voltage characteristics change after hot-carrier stress. A discussion has been going on for years on the question of the phenomena causing these changes. Are they due to fixed negative charges inside the oxide, to acceptor-type interface states acting through their potential, or to acceptor- or donor-type interface states acting through their mobility effect? It

is not easy to decide this question on the basis of electrical measurements alone, and surely the results will be different for different technologies. Recently Schwerin et al. /7/ have performed careful comparisons of 2D simulations and electrical data including the effects of stress-generated surface charges. This work revealed the complicated interplay of different types of oxide charges on drain and substrate currents. The essential result is this: without complicated simulations, statements on N_{it} and Q_{ox} can be made just by looking at I_D (V_G, V_D) and I_{sub} (V_G, V_D) characteristics together. An example of this is shown in Fig. 4. An n-channel MOSFET shows a decreasing drain current but an increasing substrate current. This cannot be explained by a mobility effect since this would not lead to an increased substrate current. Fixed negative charges would essentially lead to a parallel shift of the I_D (V_G) characteristics. Only acceptor-type interface states can explain changes of the curves in Fig. 4 satisfactorily.
Meanwhile, the experiment has brought new and puzzling results. Can we explain the behaviour shown in Fig. 5 with the above method? Here the I_D (V_G) curves before and after hot-hole stress intersect, possibly due to a combination of mobility and potential effects.

In the preceeding paragraph, several results and new questions have been discussed, showing that our understanding of the hot-carrier effect is more or less at an early state and fragmentary. These questions can become important for the way the device stability under operating conditions is extrapolated or calculated, especially for the coming generations of devices. There the safety margins will be further decreased as the technological possibilities are most likely exhausted.

III) HOT-ELECTRON MODELLING

The ultimate aim of hot-electron modelling is to reduce the hot carrier damage by choosing an optimal device structure. As a very first step to accomplish this task, models enabling the experimental data to be verified must be developed on the basis of the relevant physics. The key idea is to incorporate simple, but physically appropriate models into a realistic simulation program that allows to consider technologically relevant structures. This requires that source/drain and channel doping profiles to be accurately known. Process simulation, which itself is supported by experimental techniques to verify the doping profiles, can provide this data /8/. Another technological aspect concerns the quality of the oxide to be taken into account, in the gate and spacer regions. It is generally believed that near the gate edge the oxide quality decreases due to mechanical stress and/or implantation damage. The degradation properties certainly depend sensitively on the trap distribution in the oxide.
The problem of hot electron-modelling splits into two parts: (a) Calculation of the internal distributions in the device in the presence of very high fields; and (b) calculation of the parasitic effects such as oxide injection and charge transport in the oxide, that eventually lead to trapped charges which interfere with the controlling gate. The simulation can be used to analyze degraded samples to get information about the amount and location of the trapped charges. Or the degradation process itself, can be monitored.
We will initially discuss a modified description of carrier transport which is appropriate for submicron structures:

A) EXTENDED DRIFT-DIFFUSION APPROXIMATION

In an ideal device the field and carrier distributions are the solutions of the following semiconductor equations with appropriate boundary conditions.

$$\mathrm{div}(\varepsilon \underline{E}) = -q(n-p+C) \qquad \text{(III.1)}$$

$$\mathrm{div}\underline{j}_n - \frac{q}{c}\frac{\partial n}{\partial t} = G-R \qquad \text{(III.2a)}$$

$$\mathrm{div}\underline{j}_p + \frac{q}{c}\frac{\partial p}{\partial t} = R-G \qquad \text{(III.2b)}$$

$$\underline{j}_n = q\mu_n n\underline{E} + qD_n \mathrm{gradn} \qquad \text{(III.3a)}$$

$$\underline{j}_p = q\mu_p p\underline{E} - qD_p \mathrm{gradp} \qquad \text{(III.3b)}$$

Poisson's equation Eq. (III.1) and the continuity equations Eq. (III.2a/b) are justified by classical electrodynamics and are valid as long as fluctuations in the system are negligible. The current equations Eq. (III3a/b) are based on the general assumption valid in macroscopic systems that the particle current has a field driven drift term and a concentration gradient driven diffusion term/8/. A justification of the current equations Eq. (III3a/b) can be given by the Boltzmann equation if only small deviations from thermal equilibrium are considered. In this case the momentum relaxation time τ_m is not field dependent and therefore the mobility $\mu_{n/p}$ is a constant and Einstein s relationship holds/9/.

$$qD_{n/p} = kT\mu_{n/p} \qquad \text{(III.4)}$$

The quasi-Fermi potential $\varphi_{n/p}$ is now a very useful concept for transforming the current equations.

$$\underline{j}_n = -q\mu_n n \,\mathrm{grad}\varphi_n \qquad \text{(III.5a)}$$

$$\underline{j}_p = q\mu_p p \,\mathrm{grad}\varphi_p \qquad \text{(III.5b)}$$

It was, however, realized long ago that the linear response regime is not sufficient to obtain satisfactory results for MOSFET. Velocity saturation had to be included to explain the experimental data. This was done by replacing the mobility in Eq. (III.5) by a field dependent function $\mu_{n/p}(F_{n/p})$. The choice of the saturation force $F_{n/p}$ is not unique and has been disputed in the literature/10/. Of several possibilities $F_{n/p} = |\mathrm{grad}\varphi_{n/p}|$ is the most often used. Replacing μ by a field-dependent mobility already takes a situation in which the electrons (holes) are no longer in equilibrium with the lattice partly into account . The physical reason for this is the onset of optical phonon scattering of carriers with elevated energies. Optical phonons carry away some of the energy that the carriers gain in the electrical field. For very high fields this scattering process is no longer sufficient to maintain thermal equilibrium. As a consequence the carriers will have an elevated average energy compared with that of the phonons and their average velocity will saturate because optical phonons randomize the motion of the carriers very effectively. Therefore if phonon scattering in high fields is properly taken into account an increased carrier temperature has to be included.

Assuming a drifted Maxwellian as an Ansatz for the microscopic distribution function f_n in $\underline{x}$- and $\underline{k}$-space,

$$f(\underline{x},\underline{k}) = \exp\left(-\frac{\varepsilon(\underline{k})+\underline{k}\cdot\underline{v}_d(\underline{x})}{kT(\underline{x})}\right) \qquad \text{(III.6)}$$

in the Boltzmann equation Blotekjar/11/ obtained an alternative set of transport equations by using the moment method. Blotekjars equations quoted for one carrier type, the stationary state, and an isotropic system are

$$\underline{j} = q\mu n\underline{E} + \frac{2}{3}\mu \,\mathrm{grad}\left(\epsilon - n\frac{1}{2}mv_d^2\right) \qquad \text{(III.7)}$$

$$q\underline{v}_\epsilon = -\frac{5}{3}\underline{j}\frac{\epsilon - nmv_d^2/5}{n} \qquad \text{(III.8)}$$

$$\frac{\epsilon-\epsilon_o}{\tau_\epsilon} = \underline{j}\cdot\underline{E} - \mathrm{div}\underline{v}_\epsilon \qquad \text{(III.9)}$$

Blotekjar does not consider the determination of the momentum (τ_m) and energy (τ_ϵ) relaxation times. These have to be suitably chosen and added to complete the formulation. This and the need to add the missing heat current in Eq. (III.8), are the major drawbacks of this formulation. By now there is clear evidence that a drifted Maxwellian is a poor approximation of the distribution function/12/. Before we discuss the physical contents of Eqs. (III.7 - III.9) we will present the results obtained by one of the authors. Because the drifted Maxwellian is a poor approximation, Hänsch and Miura-Mattausch/13/ proceeded differently. Utilizing the moment method for the Boltzmann equation as well their approach focuses on the self-consistency in the four lowest moments of the distribution functions, which are: $n, \underline{j}, \epsilon, \underline{v}_\epsilon$. This is achieved by expressing the distribution function as:

$$f(\underline{x},\underline{k}) = \alpha(\underline{k})n(\underline{x}) + \underline{\beta}(\underline{k})\underline{j}(\underline{x}) + \gamma(\underline{k})\epsilon(\underline{x}) + \underline{\delta}(\underline{k})\underline{v}_\epsilon(\underline{x}) \qquad \text{(III.10)}$$

They derived an alternative set of equations to Eqs. (III.7 - III.9) that reads:

$$\underline{j} = q\mu n\underline{E} + \frac{2}{3}\mu \,\mathrm{grad}(\epsilon) \qquad \text{(III.11)}$$

$$q\underline{v}_\epsilon = -\frac{5}{3}\underline{j}\frac{\epsilon}{n} - \frac{10}{9}\mu\epsilon \,\mathrm{grad}\left(\frac{\epsilon}{n}\right) \qquad \text{(III.12)}$$

$$\frac{\epsilon-\epsilon_o}{\tau_\epsilon} = \underline{j}\cdot\underline{E} - \mathrm{div}\underline{v}_\epsilon \qquad \text{(III.13)}$$

In addition they can also provide the mobility μ and energy relaxation time τ_ϵ in a self-consistent way:

$$\frac{1}{\tau_m} = A + B\frac{\underline{j}\cdot\underline{v}_\epsilon}{\underline{j}^2} \qquad \text{(III.14)}$$

$$\frac{1}{\tau_\epsilon} = \frac{1}{\tau_{\epsilon^o}} = \text{const.} \qquad \text{(III.15)}$$

Here A, B and $\tau_\epsilon{}^0$ no longer depend on the active variables n, $\underline{j}$, ϵ and $\underline{v}_\epsilon$. The heat current, the second term on the right side of Eq. (III.12), does not vanish. Although its actual form depends on the truncation scheme of the moment method, it is constructed such that it is exact for small disturbances from equilibrium.

Eqs. (III.7 - III.9) and Eqs. (III.11 - III.15) constitute possible extensions of the classical drift-diffusion Eqs. (III.3a/b). Both contain a finite energy relaxation time τ_ϵ , which is the characteristic time for the energy transfer from the carrier system to the lattice. In the classical drift- diffusion approximation the carriers transfer their excess energy instantaneously to the lattice, which means $\tau_\epsilon = 0$. Therefore the carrier system remains in thermal equilibrium with the lattice. However, in real systems momentum changes occur more frequently than energy losses, the system will heat up. Comparing the current Eq. (III.3) with Eq. (III.7) or Eq. (III.11) we notice that the diffusion current is now driven by the energy gradient. The current Eq. (III.7) agrees with Eq. (III.11) if the kinetic energy $n\frac{1}{2}m v_d^2$ is small compared to the total energy ϵ. This is the case if the variation of the electric field over a distance $v_d \tau_\epsilon < 100$ A is moderate, which is anyway a limit of the classical field concept. Adding in Eq. (III.8) a suitable heat current and neglecting $n\frac{1}{2}m v_d^2$ again Eq. (III.8) coincides with Eq. (III.12) for the energy current v_ϵ . More essential differences in their solution are attributed to the models utilized for τ_m or μ and τ_ϵ . Whereas in Eqs. (III.11 - III.15) they are included self-consistently, Eqs. (III.7 - III.9) demand an input from other sources.

Before we discuss a possible influence on the device performance we would like to point out a possible simplification of Eqs. (III.11 - III.15) which turns out to be very convenient for practical purposes. At first we generalize the concept of the quasi-Fermi-potential, which no longer constitutes the driving potential of the current. Instead we introduce a driving force $\underline{F}_n$ for the current $\underline{j}$:

$$\underline{F}_n = \underline{E} + \frac{1}{n}\text{grad}(u_{T,n}n) \qquad \text{(III.16)}$$

Here we use $\epsilon = q\frac{3}{2}u_{T,n}n$

$$\underline{j}_n = q\mu_n n\underline{F}_n \qquad \text{(III.17)}$$

Since the classical drift-diffusion approximation is justified as an expansion in φ_n we will now expand Eqs. (III.13) and (III.14) in $\underline{F}_n$.

$$u_{T,n} = u_o + \frac{2}{3}\tau_{\epsilon,n}\mu_n(F_n)F^2(1+\theta) \qquad \text{(III.18)}$$

$$\mu_n = \mu_{o,n}\{1-\eta[u_o - u_{T,n}(F_n)(1-\Omega)]\} \qquad \text{(III.19)}$$

As shown in /15/ θ and Ω are negligible small contributions almost everywhere in the limit F_n -----> oo. Therefore Eqs. (III.18) and (III.19) together with Eq. (III.16) constitute a singular perturbed system which is solved to a very good approximation by the reduced system. This is obtained by neglecting θ and Ω . Rewriting the reduced system we finally obtain:

$$\underline{j}_n = q\mu_n n\underline{F}_n \qquad \text{(III.20)}$$

$$u_{T,n} = u_o + \frac{2}{3}\tau_\epsilon v_{sat,n}^2\left(\frac{1}{\mu_n} - \frac{1}{\mu_{o,n}}\right) \qquad \text{(III.21)}$$

$$\mu_n = \frac{2\mu_{o,n}}{1+(1+(2\mu_{o,n}F_n/v_{sat,n})^2)^{1/2}} \qquad \text{(III.22)}$$

Eqs. (III.20 - III.22) represent a simplified version of Eqs. (III.11 - III.15).

All three alternatives for calculating the effect of an elevated carrier temperature on the carrier distribution and the electric field have been realized in a numerical code. We thus find Eqs. (III.7 - III.9) with variations realized in /16,17/, Eqs. (III.11 - III.13) together with a simplified mobility model (which corresponds to the approximation which leads to Eqs. (III.20 - III.22)) in /18/, and finally Eqs. (III.20 - III.22) are used by Hänsch and Selberherr/19/.
All these approaches have an enhanced diffusion constant (coefficient of grad n) in common. This leads to an increased diffusion current which causes the carriers to be pushed further away from the surface and therefore reduces the surface concentration: whenever $u_{T,n} << u_o$. On the other hand, the sharp drop of the carrier density in the pinch-off region is, including energy balance, less pronounced. However, for non-exotic MOSFET structures hot-electron effects will only little affect the peak electric field because it is located near pinch-off where the free carrier density is low compared to the doping. Therefore the redistribution of carriers does not affect the solution of Poisson's equation. This modified carrier distribution will, however, influence the generation rate due to impact ionisation and also the oxide injection of carriers. This is because both depend on the relative position of the maximum current with respect to the maximum electric field (see Chapter IV). The influence on the I-V characteristic is small for a channel length L > 0.5 /um, but depends on the mobility model, especially on the modelling of the gate field reduced mobility.
These remarks will close the overview on possible and feasible extensions of the classical drift-diffusion approach.

B) THE PARASITIC EFFECTS OF OXIDE INJECTION AND CHARGE TRAPPING:

We are now dealing with the carriers in the high energy tail and not with an average of the distribution function. If the energy range considered is far above the average energy it can be modelled independently of the lower moments of the distribution function discussed in the previous section. The problem is to estimate the number of carriers that can overcome the Si/SiO_2 barrier and to calculate the distribution of electrically active trapped charge. This charge can either be present as fixed negative or positive charge which is located away from the interface and only has a very weak coupling to the Si (slow states Q_{ox}) or is located in the vicinity of the interface and has a strong coupling to the Si (fast interface states N_{it}). The charge on the N_{it} is related to the position of the quasi-Fermi-level $\varphi_{n/p}$ at the interface. The origin of these fast interface states is still heavily disputed. There is some evidence that a hole trapping with a successive electron capture is a possible mechanism/2/. However, a model consistent with the experimental findings is required if the degradation process is to be investigated/3/.

The second major problem is the calculation of the number of high energetic carriers in Si. The energetic range we are interested in is 1.2 eV (~50 kT) and higher. The average energy of the carriers that we can calculate with one of the approaches outlined above is of the order of 10 kT, at room temperature. This permits an independent modelling of the high energy tail of the distribution function. These carriers are experimentally accessible only indirectly by substrate and gate current measurements. Both currents test a different energy range. The threshold energy for impact ionisation is between 1.2 eV...1.8 eV. For oxide injection an energy of 3.2 eV (4.8 eV) for electrons (holes) is required. In contrast to the previous problem an explicit form of the distribution function in terms of space and energy has to be known.
The lucky electron model is based on the Shockley model/20/ in which only those electrons are considered that follow the electric field lines until they reach the required energy without experiencing a collision. This ballistic approach implies that scattering is weak. This is not confirmed by Monte Carlo calculations/21/. In contrast, it is shown that once the particles exceed the energy of an optical phonon ω_{opt}~ 60 meV scattering is very effective. Particles with $\epsilon >> \omega_{opt}$ are distributed very isotropically. The physical reason for this is that strong optical phonon scattering randomizes the motion of the carriers and does not allow a path to be followed along the field line. An analytical solution of the Boltzmann equation in the limit $\epsilon >> \omega_{opt}$ in a constant electrical field is available/22/ and can serve as a foundation for further modelling efforts/23/. Ridley/24/ established the lucky-drift-mode model which serves as a bridge between the pure ballistic approach and the statistical approach by Keldysh. The principal ideas of how to model the high energetic carrier distribution are developed for constant electric fields. The MOSFET provides a very inhomogeneous field profile and therefore models have to be developed that take account of the fact that the field varies over the ionization length $d_{ion} = \epsilon_i / q\,E$ /18,23,25/.
A suitable model for the high energetic carrier distribution should be able to explain substrate and gate current. The substrate current is easily calculated by introducing a generation rate due to impact ionisation on the right hand side of the continuity equation. The situation is not so simple for the gate current. Once the carriers are injected into the SiO_2 they will not necessarily end up at the gate electrode to be detected as gate current. If for instance $V_D >> V_G > V_{th}$ there is a strong electron injection after the pinch-off point in an n-channel device. However, the vast majority of these electrons will not reach the gate electrode because the electric field opposes their motion. Nevertheless they can still be very effective in filling possible traps in the oxide near the interface. A rigorous approach to calculating the gate current can only be to consider the transport of the injected carriers (electrons and/or holes) in the gate oxide. This can still be done by using a modified drift-diffusion approach because a stationary state of carriers and field exists for

the oxide thickness used in submicron MOSFET (t_{ox} > 8 nm). It has to be taken into account, however, that the injected carriers are far away from a stationary state with the gate field. They need a distance of 3 nm to relax to the oxide field/26/.
Once the transport problem in the oxide is solved the distribution of carriers in the oxide region is known. This is the precondition for studying the development damage caused by different mechanisms of trapping and detrapping of charge in the oxide. At this point, enter certain assumptions of the trap distribution in the oxide (oxide quality, material properties) and models how this traps are active.
A very important feature here is the time development of a static stress experiment. Several attempts to study this can be found in the literature/2,27/. However, none of these takes into account the feedback of the trapped charges into the electric field in Si, which in turn will modify the number of injected carriers/26/. This is especially important if high-voltage, short-time stress is compared with low-voltage, long-time stress.
Another feature that can be studied if the correct distribution of carriers in the oxide is known, is the drift of carriers into regions of low injection but high trap density. An example is given by trapped charges above the n^--region in a LDD device. These are responsible for a series resistance degradation.

IV) OPTIMIZATION OF SOURCE/DRAIN STRUCTURES

After having considered the experimental degradation data and theoretical aspects of hot electron modelling we turn in this section to the question of how an efficient design of a refined source/drain structure can improve the device lifetime. Most concepts aimed at improving the device reliability have focused first on the reduction of the lateral electric field strength by making the p-channel to n^+-drain transition less abrupt. These concepts are represented by phosphorus, and arsenic-phosphorus source/drain junctions /28/, and most prominently and successfully by the lightly doped drain (LDD) structure /29/. The LDD region is introduced between the p-channel and n^+-drain to spread the high electric field at the drain-pinch-off region. This measure diminishes the impact ionization and consequently reduces both the substrate current and hot carrier injection into the gate and sidewall spacer oxide.
However, a number of puzzling effects have recently been identified in LDD structures and explained by Orlowski et al /30,31/ which render device optimization more difficult. The new effects can be attributed to additional, LDD-specific, peaks of the lateral field in addition to the well-known conventional peak at the lateral drain junction.

The LDD peaks are caused by a sudden variation in vertical band bending, at the silicon surface due to the rapid decrease of the gate field strength beyond the gate edges. On the drain side the situation is aggravated by the fact that for small gate-drain overlaps (depending on LDD concentration) the conventional and LDD field peak are no longer separated but overlap, and as result of this enhance one another considerably. This enhancement of both peaks is detrimental to the device reliability for two reasons: i) the enhanced lateral fields are located just at the surface where they heat up the carriers, thus facilitating injection into the oxide; ii) the enhanced field peak is located at least partly under the spacer oxide, which is probably of lower quality than the gate oxide and thus displays an increased cross section to capture and trap the injected carriers. The oxide damage is locally distributed above the n^--region. The relevant surface states and/or fixed charges deplete the surface of the n^--region and cause early degradation of the transconductance and the drain current. Note that, assuming the same oxide damage distribution the degradation mentioned above will be larger for a lower LDD dose than for a large one. The reason for this is that a n^--region with lower concentration can be more easily depleted than a high concentration region.

We like to point out that the aforementioned overlap and enhancement of the field peaks can already occur at such high LDD doses as 4E13 cm^{-2} and standard drive-ins at 900 ^oC for 40 - 150 minutes. This is due to high surface channel doping $N_{chan} \sim 10^{17} cm^{-3}$ for T_{ox} = 15 - 20 nm and V_{Th} = 0.7 - 0.9 V leading to strong compensation effects with the source/drain structures and partly due to the additional sidewall spacer for the LDD implantation, a necessary modification of the conventional LDD geometry (see discussion below).
For a conventional LDD structure the only way out of this predicament is to increase the LDD dose. This action enhances the overall field distribution, except at the gate edge where the field peaks are separated and mutual enhancement avoided. The simulation predicts higher substrate currents for the higher LDD dose, but smaller electric fields at the surface beneath the gate edge and therefore a longer device lifetime - in perfect agreement with the experiment.
With a modified LDD structure in which the gate electrode overlaps the entire LDD region (known as inside LDD realized in structures such as ITLDD/32/ and GOLD /33/) an entirely opposite procedure would be adopted: instead of increasing the LDD dose, the LDD dose would be lowered. The reason is simple: since the gate overlaps the entire n^--region, there is no gate edge effect and consequently no LDD field peaks and therefore no enhancement of the conventional peak. On the other hand the conventional field peak will decrease with lower LDD dose, leading to a smaller substrate current and smaller degradation at the same time. This has been also confirmed experimentally /32,33/.
Another important issue within the context of conventional LDD structures (but equally important for refined LDD structures) is the sidewall spacer for LDD implantation, or LDD spacer for short. In contrast to the conventional sidewall spacer for n^+-implantation,

which is a CVD oxide, the LDD offset spacer is due to the reoxidation of the gate. The purpose of the LDD-spacer is to remove the LDD implantation damage in the oxide from the gate edges. It seems that the implantation damage leads to a greater degradation. MOSFET transistors have been fabricated with and without the LDD spacer under otherwise identical conditions. The stress experiment shows that the variant without the LDD-spacer degrades more strongly than the variant with the LDD spacer, in conflict with the theoretical prediction /30/. On the other hand, all differences in device lifetime for different LDD structures among transistors with and without LDD- spacer were in perfect agreement with theoretical considerations. It was therefore concluded that the implantation generates additional traps in the oxide, thus enhancing the overall capture probability of the injected carriers. Thus, although in the case of an LDD structure without LDD spacer, fewer hot carriers are injected into the oxide, there is more oxide damage and consequently stronger degradation due to a significantly higher capture cross section in the damaged oxide than in the case of an LDD structure with LDD spacer.

Thus the use of an LDD-spacer is an essential precondition for a degradation resistant MOSFET. It should also be noted that an LDD-spacer alleviates the short channel effects too. At the present state of conventional LDD structures, an LDD-spacer (by reoxidation) of about 50 nm with a sufficient gate-drain overlap, 100 - 150 nm, offers optimum reliability.

To summarize this part devoted to conventional LDD structures including the ITLDD-/32/, GOLD-/33/, and PLDD /34/ versions, it can be said that the theory /30/ and simulation give useful and practical guidelines for prevention of hot electron damage in submicron MOSFETs. However, if the aim is to further improve the transistor reliability, or to maintain it with a reduced transistor geometry, conventional LDD structures must be abandoned in favor of refined LDD structures such as profiled LDDs /35/ or buried LDDs /36/ or graded-buried LDDs /36/ or consider a buried channel structure /37/.

Before embarking on a discussion of these modifications it should be pointed out, that although these structures clearly offer more parameters to construct more sophisticated LDD structures and possibly to improve the long-term stability of the device, they also possess many more pitfalls which render the devices worse than their conventional cousins. This plight is aggravated by the circumstance that in contrast to the case for conventional LDD devices, the theory and simulation up to now provide only vague guidance for purposefully optimizing the refined structures.

The profiled LDD structure (PLDD) /34/ results from the insight provided by the simulation that the LDD n^--region is divided into two parts. One is a graded profile region near the n^--channel junction, which must reduce the electric field sufficiently and suppress hot carrier generation, so that the lateral impurity profile of the junction is not changed. The other one is a flat n^--surface profile region. Most of the generated surface states are distributed above the latter region, and give rise to an increase of the parasitic resistance. Therefore, in the latter region, a sufficiently high concentration is required. This is achieved by an additional shallow arsenic implantation to form a double diffused highly doped drain. Although more reliable protection from hot-carrier injection has been reported than for conventional LDD-devices, it might be doubted whether this structure is really better than an optimized conventional LDD structure. This positive judgement on PLDDs is based on a comparison with conventional LDD transistors without the LDD spacer, which entails - as we know - considerably enhanced degradation. Furthermore, the additional As implantation tends to increase the electric fields at the surface dramatically unless the subdiffusion of the phosphorus part of the n^--region is large enough. For this reason this design should be discarded for practical purposes.

A more subtle and more promising approach is represented by the buried LDD (BLDD) structure /35/. In contrast to PLDD the BLDD employs a deep arsenic n^--implant instead of a shallow one. The arsenic dose is chosen such that a peak in the n^--impurity profiles occurs appreciably (100 nm) below the Si-SiO_2 interface forming a "buried" n^--layer whereas in the conventional LDD structures the n^--doping peaks at the surface. Since the maximum lateral field occurs where the n^--doping is heaviest, the maximum field is removed from the Si-SiO_2 interface as verified by the simulation. Thus the impact ionization occurs further away from the surface and the hot carriers near the surface experience a reduced electric field and are thus less likely than surface-heated electrons to overcome the oxide energy barrier and to damage the interface and the oxide. In addition one should attempt to separate the current path and the maximum electric field, such that the maximum current density detours the field peak resulting in a decrease of the bulk current and giving further relief to the oxide damage.

The buried channel structure /37/ exploits a similar idea of shifting the carriers away from the Si-SiO_2 interface. The experiments suggest that buried channel devices are more resistant to hot carrier effects than surface channel devices. This is attributed to the deeper and broader current path in the buried channel structure using p^+-polysilicon gate technology /38/. The probability that hot carriers with an energy greater than the barrier height can reach the Si-SiO_2 interface is supposed to be smaller in the buried channel than in a surface channel device. At the same time, buried channel devices have a higher effective mobility than surface channel devices.

In our opinion, a buried channel in conjunction with a buried LDD structure may be a promising candidate for submicron MOSFETS.

We now turn to the question of how to optimize the refined LDD structures. The experimental data and simulation show that when comparing refined LDD structures with optimized conventional LDD structures it is no longer sufficient to study the magnitude and lateral location of the relevant field peaks alone. The magnitude of the field peaks may be very similar but the degradation may nonetheless be significantly different. There are two main reasons for this circumstance: i) the degradation is extremely sensitive to

even small differences in the field peaks above $3 \cdot 10^5$V/cm, ii) even assuming the magnitude of the field peaks to be identical, the degradation will still depend considerably not only on the lateral but also on the transverse location of the field peaks, or more generally, it will depend on many subtle differences in the entire distribution of the lateral and transverse electric fields between the pinch-off point and the drain contact. It will, for instance, depend on the extent and location of the overlap between the current path and the high field distribution, see Chapter III. There are unfortunately, no clear criteria for weighting such differences with respect to degradation.

This situation calls for a refined model and theoretical insight to guide the optimum design for modified LDD structures. The model given above for conventional LDD structures /30/ cannot discriminate the aforementioned subtle differences in the field distribution, which are vital for the degradation properties of the device. A model is needed which provides clear criteria for process parameters to achieve an optimum design for hot carrier resistant source/drain structures.

Though we still lack such a model, it seems to be clear in which direction the research effort must proceed. Firstly the distribution of injected hot electrons and hot holes into the oxide must be consistently calculated. Secondly the capture, trapping probabilities (depending on interface, gate oxide and spacer oxide quality) and interaction between holes and electrons /2/ have to be reliably described. Thirdly the transient build-up of the oxide damage at different operating conditions has to be taken into account. Fourthly the influence of the oxide damage on the transistor performance must be known. The latter point was already included in the simulation /7,26/.

Finally, by reversing this chain of the analysis, it should be possible to determine the optimum process parameters and the most suitable geometry for the device. The theoretical work outlined in Chapter III shows that significant developments towards such a comprehensive model have already been implemented or are in progress.

REFERENCES

/1/ L. Risch and W. Weber, ESSDERC 1988, 6B3

/2/ K. R. Hofmann, C. Werner, W. Weber and G. Dorda, IEEE ED-32, 691, 1985

/3/ R. Bellens, P. Heremans, G. Groeseneken and H. E. Maes, IEEE EDL-9, 232, 1988

/4/ E. Takeda and N. Suzuki, IEEE, EDL-4, 111, 1983

/5/ C. Hu, S. C. Tam, F.-C. Hsu, P.-K. Ko, T.Y. Chan and K. W. Terril, IEEE, ED-32, 375, 1985

/6/ W. Weber, C. Werner and A. v. Schwerin, IEDM Digest 1986, pp.390

/7/ A. v. Schwerin, W. Hänsch and W. Weber, IEEE ED-34, 2493, 1987

/8/ W. Hänsch and F. Lau, Lecture Notes of an Inter. Short Course - NASECODEV, Dublin 1987,ed. J. J. H. Miller (BOOLE PRESS Ireland 1987) pp. 50

/9/ O. Madelung, Festkörpertheorie II, (Springer Berlin 1972), chp. VIII B

/10/ S. Selberherr, Analysis and Simulation of Semiconductor Devices, (Springer Wien 1984), chp. IV

/11/ K. Blotekjar, IEEE ED-17 38, 1970

/12/ C. Jaccoboni and L. Reggiani, Rev. Mod. Phys.55, 645 1983

/13/ W. Hänsch and M. Miura-Mattausch, J. Appl. Phys. 60 650, 1986

/14/ G. Baccarani and M. R. Wordemann, Solid State Electr.28, 407 1985

/15/ W. Hänsch and Ch. Schmeiser, Zeitschrift für angewandte Mathematik und Physik, submitted for publication

/16/ M. Fukuma and R. H. Uebing, IEDM Digest 1984, pp. 621

/17/ A. Forghieri, R. Guerrieri, P. Camolini, A. Gnudi, M. Rudan and G. Baccarani, NUPAD Symposium in Santa Clara 1986

/18/ B. Meinerzhagen and W. L. Engl, IEEE ED-35 689, 1988

/19/ W. Hänsch and S. Selberherr, IEEE ED-34 1074, 1987

/20/ W. Shockley, Solid State Electr. 2 1537, 1958

/21/ N. I. Tisnek and A. P. Dimitrev, Sov. Phys. Semicond. 31 93, 1987

/22/ L. V. Keldysh, Sov. Phys. JETP 21 1135, 1965

/23/ W. Hänsch, 1988 International Electronic Devices and Materials Symposium, Kaohsing Taiwan

/24/ B. Ridley, Journ. of Phys. C16 3373, 1983

/25/ R. Kuhnert, C. Werner and A. Schütz, IEEE ED-32 1057, 1985

/26/ A. v. Schwerin and W. Hänsch, ESSDERC 1988, Montpellier, 6B1

/27/ S. Baba, A. Kita and J. Veda, IEDM Digest 1986, pp. 734

/28/ H. Kume, E. Takeda, T. Toyabe and S. Asai, Japan I. Appl. Phys., vol. 21, suppl. 21-1, 67, 1982

/29/ P. J. Tsang, S. Ogura, W. W. Walker, J. F. Shepard and D. L. Critchlow, IEEE ED-27, 1540, 1980

/30/ M. Orlowski, Ch. Werner and J. Klink, "Model for the Electric fields in LDD-MOSFETs", Part I and II, submitted for publication in IEEE ED

/31/ M. Orlowski, Ch. Werner, W. Weber, H.-P. Mühlhoff and P. Murkin, VLSI-Symp. 1987, p. 51

/32/ T. Huang, W. W. Yao, R. A. Martin, A. G. Lewis, M. Koyanagi and J. Y. Chen, IEDM Digest 1986, pp. 742

/33/ R. Izawa, T. Kure, S. Iijima and F. Takeda, IEDM Digest 1987, pp. 38

/34/ Y. Toyoshima, N. Nihira and K. Kanzaki, ULSI-Symp. 1985, pp. 118

/35/ H. R. Grinolds, M. Kinugawa and M. Kakumu, IEDM Digest 1985, pp. 246

/36/ C.-Y. Wei, J. M. Pimbley and Y. Nissan-Cohen, IEEE ED-7, 380, 1986

/37/ M.Nakahara, Y.Hiruta, M.Yoshida, K.Maeguchi and K.Kanzaki, IEDM Digest 1985, pp.238

/38/ C. Mazure, U. Schwalke and F. Neppl, ESSDERC 1988, Montpellier 4B4