NON-EQUILIBRIUM DIELECTRIC RESPONSE
OF HIGH-κ GATE STACKS IN Si MOSFETS:
APPLICATION TO SO INTERFACE PHONON SCATTERING

J. R. BARKER AND J. R. WATLING
Department of Electronics and Electrical Engineering, University of Glasgow,
Glasgow G12 8LT, Scotland, UK

The problem of coupling between phonon and plasmon modes across the idealised gate stack of a semiconductor MOSFET with one or more high-κ dielectrics has come under close scrutiny in the last few years. High-κ dielectrics are a possible technology for achieving the same gate capacitance as with silicon dioxide but for a much thicker insulating region. The effect is to drastically reduce the gate leakage current which is driven by tunnelling. Unfortunately, the much higher values of the dielectric parameters leads to the emergence of soft optical phonons that give rise to strong electron-phonon scattering at the channel-dielectric interface. Even for perfectly flat interfaces with homogeneous dielectric layers the technical problem of determining the modes and scattering amplitudes for electron-phonon-plasmon interactions is formidable. Here the problem is briefly reviewed and an outline is presented of a computational scheme to describe the modelling of MOSFETs with realistic high-κ gate stacks that are dominated by interface roughness and layer inhomogeneities. The long term objective is to build both ensemble Monte Carlo modelling schemes and non-equilibrium Green function schemes for this problem. Preliminary data based on the simplest version of the computational scheme leads to the conclusion that the inhomogeneities in the gate stack give rise to a inhomogeneous electron-SO phonon scattering amplitude across the channel where the amplitude is modulated by remote and near surface roughness and by the electron concentration and electron temperature across the channel. Polycrystallinity of the dielectric is shown to lead to strong device parameter variations across ensembles of devices. The presence of an interfacial silicon dioxide layer between channel and high-κ dielectric leads to a decrease in SO phonon scattering rates and higher effective mobilities as a function of increasing interfacial layer thickness.

1. Introduction

The present paper draws attention of the Green function modelling community to an interesting and complex many-body problem that has arisen recently in the study of the physics and modelling of ultra-small (< 45 nm) MOSFETS (metal-oxide-semiconductor-field-effect-transistors) constructed with high-κ dielectric gate stacks (see Figure 1) rather than conventional silicon dioxide. It concerns the non-equilibrium dynamic shielding of electron-phonon and electron-electron interactions in the channel by coupling to single-particle and collective excitations in the gate, source and drain; it occurs at small scales and high carrier densities. The phenomena involved are currently studied by methods that build on continuum dielectric models of the gate-stack dielectric function that controls the coupled phonon-plasmon modes. The present paper discusses the challenges and suggests improved models for non-equilibrium screening, damping models and electron-phonon coupling including the inclusion of the processes in both Monte Carlo (MC) and (tentatively) non-equilibrium Green function (NEGF) modelling of silicon devices. Detailed results from our findings will be published elsewhere.

2. The high-κ gate stack MOSFET

According to the International Technology Road Map for Semiconductors (ITRS) the scaling of MOSFETs to the 45nm technology node is expected by 2010 (the physical gate length will
be 18 nm). It will require extremely thin ($t_{ox} \sim 0.7$nm) SiO$_2$ gate oxides. The requirement for scaling to a thin oxide is fundamental as it is necessary to maximize the gate capacitance $C_{ox}$:

$$C_{ox} = \frac{\kappa_{ox} \varepsilon_0}{t_{ox}}$$  \hspace{1cm} (1)

Unfortunately, these thin oxides lead to intolerably high gate leakage arising from tunneling between the channel and the gate. Industry is therefore seeking to introduce high-$\kappa$ dielectrics into the gate insulator of silicon MOSFETs. This move allows a thicker dielectric - $t_{high-\kappa}$ - to reduce the gate leakage (tunneling) current that would otherwise rise dramatically as the silicon dioxide in present devices is scaled downwards.

$$t_{high-\kappa} = \frac{\kappa_{high-\kappa}}{\kappa_{ox}} t_{SiO_2}$$  \hspace{1cm} (2)

The leading high-$\kappa$ contender at present is HfO$_2$. More complex designs have been invoked involving for example a very thin layer of silicon dioxide between the silicon channel and the high-$\kappa$ dielectric.

3. Degradation of mobility in high-$\kappa$ gate stack MOSFETs

Unfortunately, there is a serious problem with high-$\kappa$ gate stacks. It has been pointed out$^{1-5}$ that they lead to a severe mobility degradation in the device because of the strong coupling of carriers in the channel to surface soft-optical (SO) phonons in the vicinity of the dielectric interface. Table I compares relevant parameters for silicon dioxide and HfO$_2$. The SO modes are easily excited at room temperature and lead to strong mobility reduction. This interaction is enhanced by coupling to plasmon modes in the gate metal (or poly-silicon) and the channel.

| Parameter | SiO$_2$ | HfO$_2$ |
|-----------|---------|---------|
| $\varepsilon_{ox}^0 (\varepsilon_0)$ | 3.90 | 22.00 |
| $\varepsilon_{ox}^\infty (\varepsilon_0)$ | 2.50 | 5.03 |
| $h\omega_{TO1}$ (meV) | 55.60 | 12.40 |
| $h\omega_{TO2}$ (meV) | 138.10 | 48.35 |
| $h\omega_{LO1}$ (meV) | 62.57 | 22.2 |
| $h\omega_{LO2}$ (meV) | 153.3 | 56.5 |
| $h\omega_{SO1}$ (meV) | 57.14 | 16.79 |
| $h\omega_{SO2}$ (meV) | 140.78 | 50.67 |

Table I.
Comparison of material parameters for computation of soft-optical phonon coupling from silicon dioxide and hafnium oxide gate dielectrics.

We have recently shown$^{4,5}$ that for ideal interfaces and homogeneous dielectrics the mobility degradation is as much as 50% depending on $V_D$, the drain bias and the gate over-drive ($V_G - V_{th}$). However, this degradation may be offset by a mobility enhancement that derives
from using a strained silicon channel. The model adopted assumed a Fröhlich interaction having an unscreened field amplitude at the dielectric interface given by:

$$\phi_{\text{SO}} = \frac{h \omega_{\text{SO}}}{2Q^2} \left( \frac{1}{\varepsilon_{\text{Si}}^\infty + \varepsilon_{\text{ox}}^\infty} - \frac{1}{\varepsilon_{\text{Si}}^0 + \varepsilon_{\text{ox}}^0} \right)^{1/2}. \quad (3)$$

Here, $\varepsilon_{\text{ox}}^\infty, \varepsilon_{\text{ox}}^0$ are the high frequency and static permittivities for the oxide, $\varepsilon_{\text{Si}}^\infty$ is the high frequency permittivity for Si and $\omega_{\text{SO}}$ is the soft-optical (SO) longitudinal phonon frequency, computed from the two dominant transverse-optical (TO) phonon modes using the Lyddane-Sachs-Teller relationship. Key data are presented in Table 1. The scattering amplitude matrix element, $\phi_{\text{SO}}$, decreases exponentially as a function of distance $z-t$ away from the dielectric interface (at $z=t$) into the channel: $\phi_{\text{SO}} \sim \exp[-Q(z-t)]$, where $Q$ is the in-plane wave vector representing the momentum transfer. The Born approximation scattering rate for soft-optical phonon scattering from an initial state $k$ to final state $k'$ at the interface follows an effective Fröhlich interaction:

$$R(k,k') = \frac{e^2}{4\pi \varepsilon_0} \frac{\omega_{\text{SO}}}{2\pi q^2} \left[ N_q + 1 \pm 1 \right] \left( \frac{1}{\varepsilon_{\text{Si}}^\infty + \varepsilon_{\text{ox}}^\infty} - \frac{1}{\varepsilon_{\text{Si}}^0 + \varepsilon_{\text{ox}}^0} \right) \exp[-2Q|t|]D(k') \quad (4)$$

Here, the $+(-)$ refers to phonon absorption (emission) and the Bose-Einstein distribution is

$$N_q = \exp\left[ \frac{\hbar \omega_{\text{SO}}}{k_B T} \right] - 1 \quad (5)$$

and $D(k')$ is the final density of states. For bulk silicon, $D(k) \sim E^{1/2}$, but because of the $q^{-2}$ term the scattering rate decreases with increasing energy.

---

**Figure 1.** (a) Features of realistic high-$\kappa$ gate stack MOSFETs – adapted from reference 4. (b) Layer inhomogeneities at different cross sections in the stack.
4. Simplified gate stack dielectric response

The derivation of equations (3) and (4) and the tabulated energies in Table 1 may be understood from a simple idealised gate stack model comprising a single layer of dielectric sandwiched between a metal or highly doped poly-silicon gate occupying \( z \leq 0 \) and a p-type silicon channel occupying \( z \geq t \). Assuming the channel is inverted we treat the channel as a two-dimensional electron gas (2DEG). Let the \((x,y)\) plane be parallel to the interfaces. Following earlier studies, the dielectric response of each region (gate, oxide, channel) is described by dielectric functions \( \varepsilon_g(\omega), \varepsilon_{ox}(\omega) \) and \( \varepsilon_s(Q,\omega) \) in the long-wavelength limit where we use the notation \( R, Q \) to denote in-plane position vectors and wave vectors respectively. The electrostatic potential is Fourier analysed at time \( \tau \) as:

\[
\phi(R,z;\tau) = \phi_{Q\omega}(z) \exp[iQ.R + i\omega \tau] \quad (6)
\]

The mathematical problem reduces to solving Laplace’s equation for the potential: it reduces to the following Helmholtz equation because of (6):

\[
\frac{d^2}{dz^2} - Q^2 \phi_{Q\omega}(z) = 0 \quad (7)
\]

subject to the boundary conditions that the in plane electric field components are continuous across the interfaces at \( z=0 \) and \( z=t \); and, the normal component of the displacement field is continuous at the same interfaces. The potential has therefore the form:

\[
\begin{align*}
\phi_{Q\omega}(z) &= a_{Q\omega} \exp[Qz] & (z < 0) \\
\phi_{Q\omega}(z) &= b_{Q\omega} \exp[-Qz] + c_{Q\omega} \exp[Qz] & (0 < z < t) \\
\phi_{Q\omega}(z) &= d_{Q\omega} \exp[Qz] & (z < 0)
\end{align*} \quad (8)
\]

where the four boundary conditions lead to the matrix equation

\[
A.u = 0 \quad (9)
\]

\[
A = \begin{bmatrix} 
1 & -1 & 0 & a \\
\varepsilon_g & \varepsilon_{ox} & -\varepsilon_{ox} & 0 \\
0 & \exp[-Qt] & \exp[Qt] & b \\
0 & -\varepsilon_{ox} \exp[-Qt] & \varepsilon_{ox} \exp[Qt] & 0 \\
\end{bmatrix} ; \quad u = \begin{bmatrix} a \\
b \\
c \\
d \end{bmatrix}
\]

Solutions exist, if the determinant of \( A \) vanishes, leading to the secular equation for the possible modal frequencies \( \omega_j^2 \):
\[ \epsilon_{\text{ox}}(\omega_{\text{Q}j}) + \epsilon_{\text{ox}}(\omega_{\text{Q}j})[\epsilon_{\text{g}}(\omega_{\text{Q}j}) + \epsilon_{\text{s}}(Q,\omega_{\text{Q}j})] \coth[Qt] + \epsilon_{\text{s}}(\omega_{\text{Q}j})[\epsilon_{\text{g}}(Q,\omega_{\text{Q}j})] = 0 \]  

(10)

The dispersion relations \( \omega(Q) = \omega_{\text{Q}j} \) for each branch \( j \) may be obtained by using explicit expressions for the dielectric functions. Let us take the two-pole approximation for the oxide response:

\[ \epsilon_{\text{ox}}(\omega) = \epsilon_{\text{ox}}^{\infty} \left( \frac{\omega_{\text{LO}1}^2 - \omega^2}{\omega_{\text{LO}2}^2 - \omega^2} \right) + \epsilon_{\text{ox}}^{\infty} \left( \frac{\omega_{\text{TO}1}^2 - \omega^2}{\omega_{\text{TO}2}^2 - \omega^2} \right) \]  

(11)

The long-wave limit for the gate response is taken as:

\[ \epsilon_{\text{g}}(\omega) = \epsilon_{\text{Si}}^{\infty} \left( 1 - \frac{\omega_{\text{pg}}^2}{\omega^2} \right) \]  

(12)

and the response of the inversion layer is similarly

\[ \epsilon_{\text{s}}(Q,\omega) = \epsilon_{\text{Si}}^{\infty} \left( 1 - \frac{\omega_{\text{ps}}^2(Q)}{\omega^2} \right) \]  

(13)

In these expressions, \( \omega_{\text{pg}}, \omega_{\text{ps}} \) are the plasma frequencies for the gate and substrate electrons. With these approximations, the secular equation is a polynomial equation of order six in \( \omega^2 \) giving six positive branches \( \omega_{\text{Q}j}, j=1\ldots6 \).

| Branch | Property | \( \omega_{\text{Q}j} \) |
|--------|----------|----------------|
| 1      | Coupled interface plasmon-phonon mode | \( \omega_{\text{Q}1} \geq \omega_{\text{Q}2} \) |
| 2      | Coupled interface plasmon-phonon mode | \( \omega_{\text{Q}2} \geq \omega_{\text{Q}3} \) |
| 3      | Coupled interface plasmon-phonon mode | \( \omega_{\text{Q}3} \geq \omega_{\text{Q}4} \) |
| 4      | Coupled interface plasmon-phonon mode | \( \omega_{\text{Q}4} \) |
| 5      | Localised near gate interface, weak scattering field |
| 6      | Localised near gate interface, weak scattering field |

Table 2: Properties of the six modes

Detailed analysis of these modes for the case of a large oxide thickness shows that modes 5 and 6 may be completely neglected. Indeed, for \( Qt \to \infty \), we may expand the secular equation up to leading terms in \( \exp[-Qt] \) to obtain two effective SO modes. If we further neglect screening from the channel and gate we obtain:
\[ \omega_{\text{SO}} = \omega_{10} \frac{\varepsilon_{\text{ox}}^0 + \varepsilon_{s}^{\infty}}{\varepsilon_{\text{ox}}^0 + \varepsilon_{s}^{\infty}}^{1/2} \]  

(14)

Under this approximation we obtain expressions (3) and (4) for the SO scattering, with energies given in Table I. The output of this highly-simplified model corresponds to the unscreened “remote phonon” scattering first described by Wang and Mahan\(^{14}\) and deployed for Si-SiO\(_2\) interfaces by Hess and Vogel\(^{15}\) and Moore and Ferry\(^{16}\).

In Figure 2 we show the simplified SO phonon scattering rate in the \(X\)-valley for phonon mode 1 (absorption) computed as a function of energy and the distance from interface. Here the equivalent silicon dioxide thickness using HfO\(_2\) is EOT=2.2nm and we assume a bulk Si MOSFET with planar interfaces.

Figure 2: SO phonon scattering rate in the \(X\)-valley for phonon absorption as a function of energy and the distance from interface for a bulk Si MOSFET.
5. “Realistic” gate stacks

A first improvement on the simple model (2), (3) and (14) is obtained by re-instating screening from the channel by introducing channel plasmons. The dynamically screened scattering amplitude becomes

\[
\phi_{s0} = \frac{\hbar \omega_{SO}(Q)}{2Q} \frac{1}{\varepsilon_1(Q,\omega_{SO}(Q)) + \varepsilon_{\alpha x}^m} - \frac{1}{\varepsilon_2(Q,\omega_{SO}(Q)) + \varepsilon_{\alpha x}^O} \frac{1}{\varepsilon_{\alpha x}^O} \quad 1/2
\]

(15)

where the dispersion of the phonons is given by (14) with \( \varepsilon_{Si}^\infty \) replaced by \( \varepsilon_{s}(Q,\omega_{SO}(Q)) \) using (13). Screening and anti-screening is picked up by this approximation. However, it does not describe Landau damping which requires a more sophisticated expression for \( \varepsilon_{s}(Q,\omega_{SO}(Q)) \) than expression (13). This problem is re-visited in section 8. The gate plasmons also have a key role.

If all the approximations are relaxed the scattering amplitude for the jth mode is:

\[
\phi_{s0}^j = \frac{\hbar \omega_{j}^O}{2Q} \frac{1}{\varepsilon_1(Q,\omega_{j}^O)} \frac{1}{\varepsilon_2(Q,\omega_{j}^O) - \varepsilon_{g}(Q,\omega_{j}^O)} \exp[-Q(z-t)] \quad 1/2
\]

(16)

where the total effective dielectric function is obtained (following reference 3) after considerable algebra as:

\[
\varepsilon_1(Q,\omega) = \varepsilon_1(Q,\omega) + \varepsilon_{gate}(\omega) \left[ \frac{\varepsilon_{\alpha x}(\omega) + \varepsilon_{s}(Q,\omega)}{2\varepsilon_{\alpha x}(\omega)} \right]^2 \exp[2Qt] + \varepsilon_{ins}(\omega) \left[ \frac{\varepsilon_{\alpha x}(\omega) + \varepsilon_{s}(Q,\omega)}{2\varepsilon_{\alpha x}(\omega)} \right]^2 \left( \exp[2Qt] - 1 \right)
\]

\[
+ \varepsilon_{subs}(\omega) \left[ \frac{\varepsilon_{\alpha x}(\omega) + \varepsilon_{s}(Q,\omega)}{2\varepsilon_{\alpha x}(\omega)} \right]^2 \left( 1 - \exp[2Qt] \right)
\]

(17)

where the functions \( \varepsilon_{gate}(\omega), \varepsilon_{ins}(\omega), \varepsilon_{subs}(Q,\omega) \) control the interface polarisation charges 3.

Evidently, the electron-phonon-plasmon scattering in gate stacks is a complex process that involves the entire gate stack through the coupling of plasmons in the gate and substrate to the phonon modes. We observe that the problem is exacerbated by surface roughness at the various interfaces. Since the electron-phonon interactions are controlled by the various dielectric layer thicknesses we surmise that the surface roughness modulates the electron-SO-phonon interaction. Indeed, it is expected that remote surface roughness scattering from far interfaces also plays a significant role. To complete the scene it is evident that modulated Coulomb drag due to remote interaction with the gate is also possible. A comprehensive schematic showing known processes in high-k gate stacks is shown in Fig. 1. A particularly difficult feature is the inhomogeneity of the dielectric that arises from phase separation of amorphous and crystalline regions having different permittivities.
To date, the dispersion relations for the coupled phonon-plasmon modes (16 modes for the metal-high k-interfacial oxide-channel system) and the dynamically screened electron-SO-phonon interaction in the channel have been computed by highly simplified treatments of the electronic dielectric response following the outline given in section 4 and at the beginning of this section. The understanding of the full coupling problem requires an evaluation of the effects of the variation of the carrier velocity dispersion (basically electron temperature) between source and drain that acts to modulate the Landau damping of the coupled system.

For metal gates there is a predicted long-range Coulomb drag induced by coupling to the channel electrons. Interestingly, experimental measurement\(^5\) of Coulomb drag in devices with poly-silicon gates shows only a small effect that is attributed to strong collisional damping in the gate.

6. Cellular decomposition method

It is clear from Figure 2 that the realistic gate stacks are inhomogeneous. We have extended the formalism developed by Fischetti et al.\(^1-3\) to include an inhomogeneous dielectric, inhomogeneous interfacial layer and surface roughness at the different interfaces. To obtain a computational scheme we note that from the experimentally reported structures it is a reasonable approximation to consider the gate stack as comprising a set of vertical columns with rectangular cross sections where each column has a finite set of different parallel material layers. The solution domain is partitioned into a set of such columns located by discrete coordinates (X, Y) in the stack plane (see Fig. 3). Each column comprises blocks \((m = 1 \ldots N(X, Y))\) in which the material parameters \(\tilde{\lambda}_m(X, Y)\) are approximated as constant. These constants include the dielectric parameters in each block and the averaged self-consistent electron density and electron temperature in the channel at X, Y. Because of interface roughness and separated phases the blocks are disjoint. Each vertical column labelled by (X, Y) is then treated as a slice through a stack of uniform layers from gate down to the channel. The screened SO phonon scattering rates for each column \(R(k, k'; X, Y)\) are then determined by an extension of the formalism in section 4 assuming a bare channel plasmon energy \(\hbar \omega_{\text{ps}}(X, Y)\) and local electron temperature \(T_e(X, Y)\) derived respectively from the averages of the electron density \(n(x, y)\) and temperature \(T_e(x, y)\) over the channel block at (X, Y). The transfer matrix method developed for III-V semiconductor quantum well structures is utilised here. Landau damping has been neglected or treated very simply in earlier work; here it is computed from a hot electron Lindhard model.

6. Implementations: Monte Carlo

The most accurate device modelling deploys either the ensemble Monte Carlo method (MC) or more recently the non-equilibrium Green function (NEGF) formalism. For MC modelling, the implementation is relatively straightforward although compute-intensive. In the MC method, the screened electron-phonon/plasmon scattering is manageable in the second order Born approximation (taking the cue from Boltzmann-Bloch transport theory).
The scattering rate for scattering from state $k$ to $k+q$ is:

$$ R(k, k + q; X, Y) =$$

$$ \frac{2\pi}{\hbar} |V(q)|^2 \Im\left\{ -\frac{1}{\varepsilon(q, \omega, X, Y)} \delta(\varepsilon_{k+q} - \varepsilon_k + \hbar \omega) \right\} $$

where $n_{BE}$ is the Bose-Einstein distribution, $f(k)$ is the electron distribution and the final delta-function represents the conservation of energy for absorption/emission of a quanton of energy $\hbar \omega$. In the approximate model of section 4 the effective total dielectric function $\varepsilon(q, \omega, X, Y)$ for each column yields a form:

$$ \Im\left\{ -\frac{1}{\varepsilon(q, \omega, X, Y)} \right\} = $$

$$ \sum_{i=1}^{N} c_i(q, \lambda(X, Y)) \frac{\pi}{2\varepsilon_0} \omega_q^j(\lambda(X, Y)) \{ \delta(\omega - \omega_q^j(\lambda(X, Y))) - \delta(\omega + \omega_q^j(\lambda(X, Y))) \} $$

where the quanton branches have energies $\hbar \omega_q^j$ and relative strengths $c_i$ that depend on the material parameters for each column in the cellular decomposition.
The full Monte Carlo scheme proposed here is compute intensive. However, simplifications may be made for treating Landau damping of the plasmon modes (section 8). Overall, coarse-graining the column widths will significantly reduce the complexity. This is likely to be a good approximation for the mobility-degrading SO phonon scattering but it will fail to capture the elastic boundary scattering off the rough silicon channel to insulator interface. Boundary scattering is therefore treated by standard methods. The lowest order simulation uses a single column with averaged interfacial layer dimensions and parameters to form the stack blocks.

7. Implementations: Green Function

Here the computational problem is considerable. To date, most NEGF modelling for semiconductor devices incorporates the self-consistent Born approximation (SCBA) into the self-energies to satisfy the Ward identities. For electron-phonon scattering, only non-polar scattering (corresponding to bulk silicon) has been reported. The SO modes involve a polar-interaction because of the coupling to phonons in the dielectric. As we have noted these require dynamical screening. The self-consistency required by the SCBA makes the NEGF computation exceedingly compute intensive. Nevertheless, we have begun to develop an extension of the NASA Ames-Glasgow NEGF codes for this purpose.

8. Landau damping

Landau damping is neglected in the simple model presented so far. But it plays a crucial role in suppressing the plasmons in both gate and silicon channel under certain conditions. Landau damping requires a non-local treatment of the electronic response. The finite temperature Lindhard formula\(^9\) from the random-phase-approximation (RPA) is a good starting point:

\[
\varepsilon_s(q, \omega) = \varepsilon_s^\infty - \frac{\varepsilon^2}{q^2} \frac{f(E_{k+q}) - f(E_k)}{E_{k+q} - E_k - \omega - i\Gamma}
\]

\[
\equiv \varepsilon_s^\infty (1 + \Lambda^2(q, \omega)/q^2)
\]

The last line of (20) introduces the screening wave-vector function \(\Lambda(q, \omega)\) and we allow for a three-dimensional channel. Here, \(\Gamma\) is a collision damping parameter, that for collisional effects is taken as limiting to zero from above to satisfy causality. The electron energy distribution is taken to be either Maxwell-Boltzmann or Fermi-Dirac with an effective electron temperature \(T_e\). Inserting the Lindhard expression into the secular equations for the eigen-modes (for example in equation 10) will lead to complex roots. The pure imaginary parts determine the damping rates. The problem is generally very difficult to compute. However, when the electronic dielectric function is reducible to a rational fraction of polynomials in frequency \(\omega\), the resulting algebraic equations are amenable to root finding. This polynomial form may be achieved by using a Padé approximant for the function \(\Lambda(q, \omega)\) as demonstrated by one of us elsewhere\(^{10}\). For non-degenerate carriers the calculation involves the plasma dispersion function...
\[
Z(W) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{U^2}{W}\right] dU.
\]

such that

\[
\epsilon_s(q, \omega)/\epsilon_s^\infty - 1 = \Lambda^2(q, \omega)/q^2
\]

\[
\sim C[Z(A^+)-Z(A^-)]
\]

\[
A^\pm = \left(\frac{\hbar^2}{2m^*k_BT_e}\right)^{1/2}\left[\frac{m^*\omega h}{2} \pm \frac{q}{2} + \Gamma\right]
\]

A useful Padé approximant for \( Z \) is given by

\[
Z(W) \approx \frac{p_0 + p_1W}{1 + p_2W + p_3W^2}
\]

It is very likely that the Mermin\(^{11}\) variation of Lindhard’s theory is more applicable but the procedure would be very similar. The key control parameters that determine the importance of Landau damping are the wave vector \( q \), the electron density \( n \) and the electron temperature \( T_e \) (and the appropriate form for the electron distribution: Fermi-Dirac or Maxwell-Boltzmann).

The strong variation in electron concentration and electron heating at various locations in the device severely alters the influence of Landau damping on the plasmon modes. As larger wave-vectors are accessed the plasmon modes decay into single particle modes. The variation of electron temperature along the channel thus provides an additional spatial modulation of the scattering rates and quanta. Some idea of the strong variation in electron temperature may be gleaned from Fig. 4 which shows the Monte Carlo computed electron temperature for a degenerate channel in a 67nm effective channel length Si n-MOSFET. Electron heating at various locations in the device (from ~ 300K near the source to greater than 3000 K near the drain) severely alters the influence of Landau damping on the plasmon modes.

Figure 4: Electron temperature variation along the channel of a MOSFET
Figure 4 shows the variation in electron temperature along the channel for degenerate electrons in a Si n-MOSFET. \( V_G = V_D = 1V \). Vertical dashed lines indicate the start and end of the metallurgical gate.

9. Fractal material models for polycrystallinity

Our preliminary Monte Carlo studies show that strong parameter fluctuations in ultra-scale MOSFETs are introduced by non-uniformity of the dielectric properties of the high-\( \kappa \) material due to random grain orientation (polycrystalline HfO$_2$ with different orientations of grains resulting in varying dielectric properties) and phase separation (HfO$_2$ crystallised in ‘islands’ within an amorphous SiO$_2$ matrix). Figure 5a shows an experimental TEM image in plan view for the structure of a HfSiO film, with phase separation. A 2D Fourier-Fractal model is shown in Fig. 5b.

![Plan-view TEM image of a HfSiO film illustrating phase separation of the Hf and Si oxides.](image)

![Stochastic dielectric pattern from 2-D Fourier-Fractal synthesis](image)

Figure 5: Polycrystallinity in a HfSiO film

The structural fluctuations in the dielectric of the oxide lead to corresponding fluctuations in the surface potential as illustrated in Fig. 6a. The regions of high dielectric yield higher surface potentials. This leads to fluctuations in the threshold voltage of ultra-small devices as can be seen from Fig. 6b, where we show the dispersal of the \( I_D - V_G \) characteristics for ten 50×50 nm MOSFETs modelled using Drift Diffusion methodology.

![Electrostatic potential in a 50×50 nm MOSFET showing the fluctuations in surface potential due to the spatial variations in gate dielectric, shown in the plane above.](image)

![\( I_D - V_G \) characteristics for 10 devices with random dielectric pattern, and the limiting cases.](image)

Figure 6: Effects of random polycrystallinity on a 50X50 nm MOSFET.
Figure 6 also shows the current-voltage characteristics for a pure SiO$_2$ gate oxide device and a pure high-$\kappa$ (above an interfacial layer) device the limiting cases for the fluctuations. In all cases the physical gate oxide is 4 nm thick and degradation in subthreshold slope due to the reduced gate oxide capacitance when SiO$_2$ is employed is evident.

10. Effect of an interfacial oxide layer

We find that the rough interface boundaries gives rise to a random variation of the effective dielectric thickness that carries through into a modulation of the electron-phonon-plasmon coupling and local phonon energies. This is illustrated by a simple single column Monte Carlo simulation for a gate stack consisting of a HfO$_2$, high-$\kappa$ and a SiO$_2$ interfacial layer, similar to that illustrated in Fig 1. Fig. 7. shows the variation in computed mobility versus oxide interfacial layer thickness with a constant inversion layer field of 1MV/cm.

![Figure 7: Effective mobility within the channel as a function of the interfacial (SiO$_2$) layer thickness.](image)

11. Conclusions

A methodology has been outlined for tackling the difficult problem of modelling MOSFET devices with realistic high-$\kappa$ gate stacks. Preliminary results show the importance of SO phonon scattering and rough boundaries on device parameters. The problem is compute-intensive even for semi-classical Monte Carlo methodology. The incorporation of these concepts in NEGF modelling is important for ultra-small devices where quantum transport is significant. It is likely that SO phonon processes will destroy the coherent flow in regimes that are quasi-ballistic in conventional MOSFETs. However, the overall formalism requires a major simplification if a fully self-consistent treatment of the coupled many body problem is to be manageable by NEGF methods. Some of these results were reported briefly elsewhere.

Acknowledgments

This work is supported by SEMATECH under their Advanced Gate Stack Program and EPSRC grant GR/S80097/01 “Meeting the material challenges of nano-CMOS electronics”. One of the authors (JRW) also acknowledges the support of EPSRC through an Advanced Research Fellowship.
References

1. M. V. Fischetti and S.E. Laux, J. Appl. Phys. 89 1205 (2000).
2. M. V. Fischetti, J. Appl. Phys. 89 1232 (2000).
3. M.V. Fischetti, D. A. Neumayer and E. A. Cartier, J. App. Phys. 90, 4587-4607 (2001).
4. S. Saito et al, IEDM 2003 Tech. Dig, pp. 796-799, (2003).
5. L. Yang, J. R. Watling, F. Adamu-Lema, A. Asenov and J. R. Barker, Scaling study of Si and strained Si n-MOSFETs with different high-k gate stacks, 2004 International Electron Device Meeting (IEDM), San Francisco, December (2004).
6. J.R. Watling, L. Yang, A. Asenov, J.R. Barker and S.Roy, IEEE Trans. Device and Materials Reliability 5, 103-108 (2005).
7. P. M. Solomon and M. Yang, Relevance of remote phonon scattering in gate to channel mobility of thin oxide CMOS devices 2004 International Electron Device Meeting (IEDM), San Francisco, (2004).
8. S. Yu, K.W. Kim, M.A. Strocio, G.J. Iafrate, J-P. Sun and G.I. Haddad, J.Appl.Phys. 82, 3363 (1997).
9. J. Lindhard, Kgl.danske Videnskab. Mat. Fys.Medd. 28, 8 (1954).
10. D. Lowe and J. R. Barker, J. Phys. C 18 2507 (1985).
11. N.D. Mermin, Phys. Rev. B1, 2362 (1970)
12. J.R. Barker, J.R. Watling, A. Brown, S. Roy, P.Zetzoff, G. Bersuker and A. Asenov, International Conference on Hot Carriers in Semiconductors 14, Chicago, (2005)
13. J. R. Watling, L. Yang, M. Borići, R. C. W. Wilkins, A. Asenov, J. R. Barker and S.Roy, Solid State Electronics, 48, pp. 1337-1346, (2004).
14. S.Q. Wang and G. D. Mahan, Phys.Rev B6 4517 (1972)
15. K. Hess and P. Vogl, Solid St. Commun. 30 807 (1979)
16. B. T. Moore and D.K. Ferry, J.Appl.Phys. 51 2603 (1980)