Mesoscopic noise in VLSI devices

Iannaccone, Giuseppe


Event: SPIE's First International Symposium on Fluctuations and Noise, 2003, Santa Fe, New Mexico, United States
Keynote Address

Mesoscopic noise in VLSI devices

G. Iannaccone
Dipartimento di Ingegneria dell’Informazione, Università di Pisa
Via Diotisalvi 2, I-56122, Pisa, Italy

ABSTRACT

The rich variety of noise properties that make the field of mesoscopic transport so fascinating is going to be shared with "common" VLSI devices. Typical MOSFETs used of present-day VLSI circuits and systems already have feature sizes smaller than what we usually consider mesoscopic devices. In this talk, we focus on shot noise of the drain and gate currents in nanoscale MOSFETs. The subject is of interest from the point of view of applications, since adequate models of noise in such MOSFETs are required, especially for high-frequency analog and mixed-signal applications, and from the point of view of the understanding of the underlying physics, since effects typical of mesoscopic devices can now be observed at room temperature and in silicon.

Keywords: Noise, Shot Noise, MOSFETs, Nanoelectronics, Mesoscopic devices

1. INTRODUCTION

When considering nanoelectronic devices, one should certainly include the old-fashioned MOSFET. Indeed, the present day 0.13 mm CMOS technology is characterized by devices with effective gate length of 70 nm and gate oxide thickness of 2 nm. Within the next ten years, according to the 2001 International Technology Roadmap for Semiconductors, devices with gate lengths shorter than 20 nm and gate dielectrics with equivalent oxide thickness below 1 nm are expected.

Such dimensions are much smaller than those of typical mesoscopic devices, for which noise properties have been extensively studied in recent years. Therefore, we should expect that the rich and intriguing properties of noise that have been unveiled in mesoscopic devices, should manifest themself also in the more secular MOSFETs, as they enter the ballistic transport regime.

In this talk we will show that this is the indeed the case. We will discuss the aspects of charge transport which are suitable to be described in terms of ballistic, diffusive, or single electron transport, and are therefore within the realm of what is usually defined as “mesoscopic” transport. In particular, we will focus on the suppressed shot noise of the drain current, due to Pauli exclusion and electrostatic force among the ballistic electrons in the channel, and on the suppressed shot noise of the gate current, due to the presence of defects in the oxide as a consequence of electrical stress and aging.

Shot noise in multinode ballistic conductors, a class of devices including MOSFETs, has been recently investigated. Suppression of shot noise has been predicted as a result of Fermi statistics at the source contact, and of electrostatic interaction in the channel. Here, we focus on nanoscale ballistic MOSFETs, using a model already developed for the simulation of their DC characteristics. In such devices, quantum confinement is particularly strong in the direction perpendicular to the silicon-silicon oxide interface, and transport mainly occurs in the lowest-lying two-dimensional subband in the inversion layer.

We show that in the assumption of completely ballistic transport we can obtain an analytical expression for the shot noise of the drain current, that can be easily computed once the values of the parameters involved are obtained from detailed numerical simulations. The analytical expression, however, provides important insights into the mechanisms of shot noise suppression.

Then, we focus on noise properties of the gate current. In that case we have already shown that the power spectral density of shot noise associated to the current through the dielectric is reduced in stressed oxides with respect to fresh oxides. Here, we review the model of noise in the case of Stress Induced Leakage Currents

Further author information: E-mail: g.iannaccone@iet.unipi.it, Telephone: +39 050 2217677
(SILCs), the excess currents through a thin oxide Metal-Oxide-Semiconductor (MOS) capacitor observable at low voltages after the structure has been stressed by a large electric field.

SILCs have been observed more than twenty years ago,\textsuperscript{9,10} and have been extensively studied from the experimental and theoretical point of view.\textsuperscript{11-15} Such a wide interest is due to the fact that SILCs are a major problem for the reliability of MOS structures, and presently constitute the major obstacle to the downscaling of non-volatile memory devices.

Among the transport mechanisms proposed to explain SILCs (tunneling enhancement due to hole trapping,\textsuperscript{14} trap-assisted tunneling,\textsuperscript{11} an effective reduction of the oxide thickness due to the growth of a conductive filament\textsuperscript{15}) trap-assisted tunneling is presently considered as the most likely. The nature of traps and of the type of tunneling (elastic or inelastic) is not yet clear.\textsuperscript{15,10}

2. SHOT NOISE OF THE BALLISTIC DRAIN CURRENT

Given a MOSFET bias point, we can assume that the subband profile is obtained from the self-consistent solution of the Poisson-Schrödinger equation in two dimensions.\textsuperscript{6} The assumption of fully ballistic transport simplifies the enforcement of current continuity, that is directly guaranteed by filling up propagating states according to Fermi Dirac statistics with the chemical potential of the originating contact. Fluctuations of the drain current are due to fluctuations of the occupancy of incoming states (Pauli interaction) and to potential fluctuations induced by the mentioned fluctuations of occupation factors (through long-range electrostatic interactions). Both terms are expected to suppress current noise with respect to Poissonian shot noise.

A typical profile of the doping profile of a 25 nm MOSFET and of the first subband as a function of $V_{DS}$ is taken from Ref. 6 and shown in Fig. 1. We can extract the maximum value of the first subband in the channel $E_M$ for the considered bias point.

If we consider that only the first subband is occupied, the density of states in the channel is $N_{2D} = \frac{2m_y}{\pi^2 E_y E_z}$, where $y$ is the direction of propagation, and $N_{2D}$ is only associated to one direction of propagation, from source to drain ($k_y > 0$) or vice versa ($k_y < 0$); the factor 2 takes into account the two degenerate minima in the $k_x$ direction, $m_y$ is silicon transversal mass, $h$ is Planck's constant, $E_y$ ($E_z$) is the kinetic energy in the $y$ ($z$) direction.
The electron density per unit area for \( y \) corresponding to the subband maximum is

\[
n_{2D} = 2 \int_0^\infty dE_y \int_0^\infty dE_z N_{2D} [f_s(E_y + E_z + E_M) - f_D(E_y + E_z + E_M)],
\]

where \( f_s(E) \) \( [f_D(E)] \) is Fermi Dirac occupation factor with Fermi energy of the source \( E_{FS} \) \( [\text{drain} \ E_{FD}] \).

We assume that fluctuations of the propagating states occupation factor affect \( n_{2D} \) directly and through their electrostatic effect on \( E_M \). From (1), we have

\[
\delta n_{2D} = 2 \int_0^\infty dE_y \int_0^\infty dE_z N_{2D} [\delta f_s + \delta f_D] + 2\delta E_M \int_0^\infty dE_y \int_0^\infty dE_z \left( \frac{\partial f_s}{\partial E} + \frac{\partial f_D}{\partial E} \right)
\]

We use a very simple approximation for the electrostatics, synthesizing all electrostatic effects in a unique gate capacitance per unit area \( C_G \), which gives us the relationship between the electron density and the subband potential at the maximum:

\[
\delta E_M = q^2 \delta n_{2D} / C_G,
\]

from which we obtain

\[
\delta E_M = \frac{2q^2 \int_0^\infty dE_y \int_0^\infty dE_z N_{2D} (\delta f_s + \delta f_D)}{C_G + C_{\text{degS}} + C_{\text{degD}}},
\]

where we have defined

\[
C_{\text{degS}} \equiv 2q^2 \int_0^\infty dE_y \int_0^\infty dE_z N_{2D} \frac{\partial f_s}{\partial E_{FS}}, \quad \text{and} \quad C_{\text{degD}} \equiv 2q^2 \int_0^\infty dE_y \int_0^\infty dE_z N_{2D} \frac{\partial f_D}{\partial E_{FD}}.
\]

The ballistic current density \( I \), computed at the subband maximum, is given by

\[
I = 2q \int_0^\infty dE_y \int_0^\infty dE_z N_{2D} v_y \left[ f_s(E_y + E_z + E_M) - f_D(E_y + E_z + E_M) \right],
\]

where \( v_y = \sqrt{2E_y/m_y} \) is the velocity along direction \( y \). Again, fluctuations of \( f_s \) and \( f_D \) act directly on \( J \) and through long range electrostatic interactions (through \( E_M \)).

\[
\delta I = 2q \int_0^\infty dE_y \int_0^\infty dE_z N_{2D} v_y (\delta f_s - \delta f_D) + \delta E_M \int_0^\infty dE_y \int_0^\infty dE_z N_{2D} v_y \left( \frac{\partial f_s}{\partial E} - \frac{\partial f_D}{\partial E} \right).
\]

If we substitute (4) into (7), after straightforward derivation, we obtain

\[
\delta I = 2q \int_0^\infty dE_y \int_0^\infty dE_z N_{2D} v_y \left( \frac{\partial f_s}{\partial E_{FS}} - \frac{\partial f_D}{\partial E_{FD}} \right) \left[ \left( 1 - \frac{\tilde{v}_s C_{\text{degS}} - \tilde{v}_D C_{\text{degD}}}{C_G + C_{\text{degS}} + C_{\text{degD}} v_y} \right) \delta f_s - \left( 1 - \frac{\tilde{v}_D C_{\text{degD}} - \tilde{v}_s C_{\text{degS}}}{C_G + C_{\text{degS}} + C_{\text{degD}} v_y} \right) \delta f_D \right],
\]

where \( \tilde{v}_s \) and \( \tilde{v}_D \) are weighted averages of the velocity \( v_y \):

\[
\tilde{v}_s \equiv \frac{\int_0^\infty dE_y \int_0^\infty dE_z N_{2D} \frac{\partial f_s}{\partial E_{FS}} v_y}{\int_0^\infty dE_y \int_0^\infty dE_z N_{2D} \frac{\partial f_s}{\partial E_{FS}}}, \quad \tilde{v}_D \equiv \frac{\int_0^\infty dE_y \int_0^\infty dE_z N_{2D} \frac{\partial f_D}{\partial E_{FD}} v_y}{\int_0^\infty dE_y \int_0^\infty dE_z N_{2D} \frac{\partial f_D}{\partial E_{FD}}},
\]

Considering that the current is a sum of pulses each carrying charge \( q \), that \( \int f_s = f_s(1 - f_s) \), and that the occupation factors of different modes are uncorrelated, we can write the power spectral density of the shot noise

Proc. of SPIE Vol. 5115 33
as

\[ S = 4q \int_0^\infty dE_y \int_0^\infty dE_z N_{2D} v_y \times \left[ \left( 1 - \frac{\tilde{v}_S C_{\text{deg}S} - \tilde{v}_D C_{\text{deg}D}}{C_G + C_{\text{deg}S} + C_{\text{deg}D} v_y} \right)^2 f_S(1 - f_S) + \left( 1 - \frac{\tilde{v}_D C_{\text{deg}D} - \tilde{v}_S C_{\text{deg}S}}{C_G + C_{\text{deg}S} + C_{\text{deg}D} v_y} \right)^2 f_D(1 - f_D) \right]. \] (10)

Let us consider a far from equilibrium condition, with an applied \( V_{DS} \) of a few \( k_B T/q \), so that \( f_D \) is much smaller than \( f_S \), and therefore can be removed from any expression. The shot noise suppression factor \( \gamma \), defined as \( \gamma = \frac{S}{S_D} \), in that case, becomes, from (6) and (10):

\[ \gamma = \frac{2 \int_0^\infty dE_y \int_0^\infty dE_z N_{2D} v_y \left( 1 - \frac{\tilde{v}_S C_{\text{deg}S}}{C_G + C_{\text{deg}S} v_y} \right)^2 f_S(1 - f_S)}{2 \int_0^\infty dE_y \int_0^\infty dE_z N_{2D} v_y f_S}. \] (11)

If we define \( \gamma_C \) as

\[ \gamma_C(E_y) = \left( 1 - \frac{\tilde{v}_S C_{\text{deg}S}}{C_G + C_{\text{deg}S} v_y} \right)^2, \] (12)

we can see from Eq. (11) that the noise suppression factor is a weighted average of the term

\[ \gamma_C(1 - f_S). \] (13)

where the weight is the contribution of the state defined by \( E_y \) and \( E_z \) to the drain current.

Clearly, \( \gamma \) is always smaller than one, and suppression occurs through two well defined terms: a term \((1 - f_S)\), that is the effect of Pauli principle, and \( \gamma_C \), that is the effect of electrostatic interaction.

In particular cases only one term, i.e., one type of interaction, has a role in shot noise suppression. If, for example \( f_S \ll 1 \), we are in the case in which the Fermi-Dirac distribution can be approximated with a Maxwell distribution. In that case, the term \((1 - f_S) \approx 1\), and the suppression factor is a weighted average of \( \gamma_C \). On the other hand, if the gate capacitance is extremely effective in screening the electrostatic interaction \((C_G \rightarrow \infty)\), we have \( \gamma_C = 1\): the suppression factor is a weighted average of \((1 - f_S)\).

Let us consider for a moment only the effect of electrostatic interaction \((f_S \ll 1)\): if we assume that the velocities in \( \gamma_C \) are of the same order, the suppression factor due to electrostatic interaction is \( \gamma_C \approx |C_G/(C_G + C_{\text{deg}S})|^2 \). This expression allows us to evaluate the order of the suppression of shot noise for different MOSFET structures.

For a typical device structure, \( C_{\text{deg}S} \) is close to so called “quantum capacitance” at zero temperature \( q^2 m_i/2\pi^2\hbar^2 = 0.13 \, \text{F/m}^2 \). In strong inversion \( C_G \) may be well approximated with the oxide capacitance. For an equivalent oxide thickness of 2 nm (including quantum correction and polysilicon depletion), we have, \( C_G = 172 \, \text{F/cm}^2 \), providing a suppression term \( \gamma_C = 0.0135 \).

A more precise numerical evaluation of noise can be obtained from Eqs. (6) and (10), once subband profiles are computed with a 2D Poisson-Schrödinger solver. From the application point of view, particular care has to be devoted to the resistance of source and drain extensions, since the associated thermal noise is expected to be the largely dominant term in channel noise.

3. SHOT NOISE OF THE TUNNEL GATE CURRENT

The tunneling current \( I_{\text{tresh}} \) through a perfect oxide layer consists of several events per unit time, each corresponding to the transfer of an electron charge \(-q\) from the cathode to the anode as a consequence of a tunneling event. In this case the electrons behave as independent particles, and the tunneling probability is only a function of the electron energy, barrier thickness and shape, and the density of states. The tunneling current is therefore the result of a Poissonian process, and the power spectral density of the current noise \( S \) is therefore given by the known expression for “full” shot noise \( S = 2q I_{\text{tresh}} \).
Figure 2. Band profile of a MOS structure with possible transitions from the contacts to a trap in the oxide of energy $E_\alpha$ and vice versa. Region 1 is the substrate, region 2 is the polysilicon gate.

After electric field stress, the well known phenomenon of SILCs can be observed: the current, especially at low fields, increases by orders of magnitude, meaning that some additional transport mechanisms becomes dominant. Over the years, several microscopic mechanisms have been proposed. Most of them would still allow electrons to behave independently, and therefore would correspond to a “full” shot noise spectrum. Indeed, the localized thinning of the oxide due to the presence of a conductive filament, the localized lowering of the oxide barrier, and the local alteration of the oxide barrier due to hole trapping, would only increase the tunneling probability for electrons in the neighborhood of the defect, but electrons could still be treated as independent particles not interacting with each other, so that the tunneling current would again be governed by Poisson statistics.

On the other hand, the noise properties of SILCs are significantly altered if we ascribe such currents to trap-assisted-tunneling, a two-step process in which electrons first tunnel from the cathode to a trap in the oxide, then from the trap to the anode. Because of Coulomb repulsion and Pauli exclusion principle, only one electron at a time can occupy the same trap; this means that the probability for an electron to undergo a two-step tunneling process through a given trap depends on whether that trap is currently occupied or not. This specific aspect introduces correlation between electrons using the same trap for tunneling, therefore the current is no more a sum of independent tunneling events, and the process is non Poissonian (in fact, super-Poissonian, i.e., with a reduced variance).

In the rest of the section, we present a model for DC and noise properties of SILCs based on trap-assisted-tunneling. For the sake of generality, let us consider the semiconductor-insulator-semiconductor structure whose conduction and valence bands are sketched in Fig. 2. In the case of metal contacts the situation is simpler, since only one band per electrode can be considered. In addition, let us consider a trap in the oxide, consisting in a localized electron state at position $x'$ in the oxide ($0 < x' < d$) and at energy $E_\alpha$. We will assume that the trap has a single level with two possible states (spin up and down), but Coulomb repulsion prevents two electrons from occupying the same trap.

We follow the notation used in the case of generation-recombination processes; we call “generation” rate the transition rate from an electrode to the unoccupied trap, and “recombination” rate the transition rate from the occupied trap to one electrode. As can be seen in Fig. 1, we consider four different generation rates, on the basis of the location of the initial state: generation rate from the conduction band of electrode 1 ($g_{1C}$), from the valence band of electrode 1 ($g_{1V}$), from the conduction band of electrode 2 ($g_{2C}$) and from the valence band of electrode 2 ($g_{2V}$). Analogously, we define the four recombination rates, on the basis of the location of the final
state (the same subscript notation is used). Let us call \( |\alpha| \) the electron state in the trap, and let us consider a state \( |\beta| \) in the conduction band of region 1. According to the Fermi "golden rule" the transition rate from \( |\beta| \) to \( |\alpha| \) would be

\[

\nu_{\beta \rightarrow \alpha} = \frac{2\pi}{\hbar} |M(\alpha, \beta)|^2 \Gamma (E_\alpha - E_\beta)
\]

(14)

where \( \hbar \) is the reduced Planck's constant, \( M(\alpha, \beta) \) is the transition matrix element between state \( |\alpha| \) and \( |\beta| \), \( E_\alpha \) and \( E_\beta \) are the energies of states \( |\alpha| \) and \( |\beta| \), respectively. The function \( \Gamma \) is a lorentzian curve of halfwidth \( \Gamma \),

\[

\Gamma (E_\alpha - E_\beta) = \frac{\Gamma}{(E_\alpha - E_\beta)^2 + \Gamma^2}.
\]

(15)

and represents the simplest way to account for inelastic transitions. As can be noticed, \( \hbar \) tends to a delta function as \( \Gamma \) approaches 0, i.e., when only elastic transitions are considered. The larger \( \Gamma \), the larger degree of inelastic transitions is allowed.

The transition rate can also be related to the probability current density \( J(\beta, x') \) of state \( |\beta| \) on the plane \( x' \) where the trap is located through the so-called capture cross section \( \sigma_{\alpha, \beta} \)

\[

\nu_{\beta \rightarrow \alpha} = \sigma_{\alpha, \beta} J(\beta, x') = \sigma_{\alpha, \beta} T_1(E_t) \nu_1(E_t)
\]

(16)

where \( E_t \) is the energy in the \( x \) direction of state \( |\beta| \), \( T_1(E_t) \) is the transmission probability of the one-dimensional barrier from \( x' \) to \( d \), and \( \nu_1 \) is the so-called attempt frequency of the state of longitudinal energy \( E_t \). The trap cross section can depend of course on the trap state and on the state \( |\beta| \) in a non trivial way. However, given our lack of knowledge on the nature of traps, we make the simplest assumption that is consistent with Eq. (14): \( \sigma_{\alpha, \beta} = k \Gamma (E_\alpha - E_\beta) \), where \( k \) is a constant.

The state \( |\beta| \) is defined by its longitudinal energy \( E_t \), its energy in the transverse plane \( E_T \) (\( E_\beta = E_t + E_T \)) and its spin. The generation rate \( g_{1e} \) is obtained by integrating (16) over all occupied states in the conduction band of electrode 1:

\[

g_{1e} = 2 \int_{E_{1e}}^\infty dE_t \int_0^\infty dE_T k \Gamma (E_t + E_T - E_\alpha) \times T_1(E_t) \nu_1(E_t) T_1(E_t + E_T) \rho_1(E_t) \rho_T
\]

(17)

The factor 2 takes into account spin conservation, \( \rho_1 \) and \( \rho_T \) are the densities of states in the longitudinal direction and in the transversal plane, respectively. \( f_1(E_t + E_T) \) and \( E_{1e} \) are the Fermi-Dirac occupation factor and the conduction band edge in the first electrode, respectively.

The recombination rate \( r_{1e} \) has an expression very similar to (17), with the difference that the integral has to be performed over unoccupied states in the conduction band of electrode 1 with the same spin of the trapped electron:

\[

r_{1e} = \int_{E_{1e}}^\infty dE_t \int_0^\infty dE_T k \Gamma (E_t + E_T - E_\alpha) \times T_1(E_t) \nu_1(E_t) [1 - f_1(E_t + E_T)] \rho_1(E_t) \rho_T
\]

(18)

At this point, the expressions of the other transition rates can be derived straightforwardly, and we will not write them in detail. We can group transition rates as follows:

\[

g_1 \equiv g_{1e} + g_{1v}; \quad r_1 \equiv r_{1e} + r_{1v} \quad g_2 \equiv g_{2e} + g_{2v}; \quad r_2 \equiv r_{2e} + r_{2v}.
\]

(19)

The occupation factor \( f' \) of the trap in the steady state regime can be readily obtained by imposing the detailed balance of generation and recombination:

\[

f' = \frac{g_1 + g_2}{g_1 + g_2 + r_1 + r_2}.
\]

(20)
The average current $I'$ through the trap can therefore be written as

$$I' = q g_1 (1 - f') - q r_1 f' = q r_2 f',$$

(21)

while the noise spectral density of the noise current at zero frequency is readily obtained with a procedure very close to that used for obtaining generation-recombination noise and noise in resonant tunneling structures\(^{18}\) that will be discussed in detail elsewhere\(^{18}\):

$$S' = 2 q I' \left[ 1 - \frac{g_1 r_2}{(g_1 + r_1 + g_2 + r_2)^2} \right] = 2 q \gamma' I'.

(22)

The shot noise suppression factor $\gamma'$, or Fano factor, is defined as $\gamma' = S'/2 q I'$: as can be seen from (22), is between 0.5 and 1\(^{18}\).

Let us assume that traps are distributed with a density $\eta$ per unit volume per unit energy. The total trap-assisted current density $J_{TA}$ and the associated noise spectral density $S_{TA}$ can be obtained by integrating $I'$ and $S'$ over $E_a$ in the insulator gap, and $x'$, in the longitudinal direction from 0 to $d$, i.e.

$$J_{TA} = \int \int I' \eta(E_a, x') dE_a dx',

S_{TA} = \int \int S' \eta(E_a, x') dE_a dx'.

(23)

$J_{TA}$ is proportional to the product of the capture cross section and the trap density, while the Fano factor

$$\gamma_{TA} \equiv \frac{S_{TA}}{2 q J_{TA}} = \frac{\int \int \gamma' \eta(E_a, x') dx' dE_a}{\int \int \eta(E_a, x') dx' dE_a},

(24)

is again between 0.5 and 1, and is independent from any constant factor in (17) and (23).

The information on $\eta$ is typically very poor, making it difficult to validate the model against experimental results. Here, we consider a MOS capacitor with a 6 nm oxide that has been completely characterized, from the point of view of DC transport and noise, before and after stress, in Ref. 8. We extract $\eta$ by fitting the numerical results on the IV characteristic after stress with the experiments. Results of the experimental curve, from Ref. 8, and of the theoretical curve are shown in Fig. 3. Best fitting is obtained when $\eta$ has a Gaussian profile centered at $E_0 = 0.1$ eV above silicon conduction band at flat band, a standard deviation of 91 meV, and is homogeneous in the oxide volume.

With the same trap concentration, we have computed the noise suppression factor of the SLC component as a function of the gate voltage. Results are shown in Fig. 4 and compared with experiments in Ref. 8. In order to evaluate the sensitivity of the results to the density of states of traps, we show results with slightly different values of the central energy of trap distribution.

As can be seen, the model allows us to reproduce in a rather accurate way both DC and noise properties of tunnel currents through thin oxides.

4. CONCLUSION

In this talk we have shown that typical approaches to the investigation of noise in mesoscopic devices, based for example on Landauer-Buttiker scattering formalism, or on single electron transport formalisms, can provide useful insights into the noise properties of much more common devices for VLSI circuits and devices, such as MOSFETs. As MOSFETs enter the ballistic transport regime, we expect to find experimentally, and even at room temperature, the rich variety of noise properties that has made the field of mesoscopic noise so fascinating in the last two decades.

Acknowledgments

This work has been supported by the Fondazione Cassa di Risparmio di Pisa, and by the EU through the project NANOTCAD (IST-1999-10828).
Figure 3. Experimental I-V characteristic of a 6 nm oxide after stress, from Ref. 8 and theoretical I-V curve after fitting: only SILC component (thin line) and total current (thick line).

Figure 4. Experimental noise suppression factor of a 6 nm oxide after stress, from Ref. 8 and theoretical results obtained for different values of the central energy $E_0$ of trap distribution.
REFERENCES