Numerical Simulation and Analytical Modeling of Strong-Inversion Gate Capacitance in Ultra-Short (30nm) MOSFETs

C. Sudhama*, Oana Spulber**, Colin McAndrew* and Rainer Thoma*

*Motorola SPS, Maildrop EL-615, 2100 E. Elliot Road, Tempe, AZ 85284, c.sudhama@motorola.com
**De Montfort University, Leicester, United Kingdom

ABSTRACT

Bulk and novel MOSFET structures with gate-lengths in the *30nm* regime are expected to become industry standards in *~2007*. As devices are scaled down to these lengths, overlap- and fringe-capacitance between the gate and source/drain regions gain more importance as a fraction of the total gate-capacitance. Therefore in numerical simulations of these structures it is necessary to carefully include various parasitic capacitance components. Further, finite doping levels in poly-gates lead to appreciable poly-depletion effects. For the *30nm* MOSFET, the inclusion of the actual gate thickness and dielectric overlayer in the simulation increases fringing capacitance by *~6X*, compared to an infinitesimally thin gate and no overlayer. Analytical expressions for these strong inversion capacitance components are useful in device design and circuit simulation. We have derived analytical capacitance formulae for various geometric configurations, which yield values that match numerical simulations quite well.

Keywords: novel, MOSFET, fringe-capacitance, strong-inversion, poly-depletion, TCAD, numerical simulation.

1 STRONG-INVERSION GATE CAPACITANCE

Several researchers have demonstrated experimental prototypes of MOSFETs that have channel lengths (L_{eff}) and gate lengths (L_{gate} or L_{poly}) in the *30nm* range [1-3]. It is expected that bulk and novel MOSFET structures in this length regime will become industry standards in the *~2007* time-frame [4]. With increased interest in their fabrication and electrical characteristics, there is now a need to carefully model these devices using numerical device simulators, and a need to provide accurate models for circuit simulation.

Fig. 1 shows the geometric cross-section of a MOSFET gate, along with the various capacitance components. A thickening of the gate-oxide beneath the edge of the polysilicon is sometimes termed a "smile", and is the result of oxidant penetration beneath the polysilicon during the re-oxidation step. In reality this oxidant penetration results in non-planarity of the silicon substrate as well. It can be shown that attributing all of the oxide-thickening to the polysilicon region is an acceptable approximation with a negligible effect on modeled capacitance [5]. As Fig. 1 shows, C_{smile} and C_{ox} together comprise the "parallel-plate" component of the gate capacitance, termed C_{pp}. These two components of capacitance are associated with nearly *vertical* electric field lines. In reality, the 'smiling' poly has a curved surface, even if it is restricted to the polysilicon; however a linear (or piece-wise linear) approximation is sufficiently accurate for most applications. Assuming vertical electric field lines under the gate, integration yields the following formula for the gate capacitance in strong-inversion:

$$C_{GG} = 2\frac{\epsilon W K_L}{K_T} ln(K_T + 1) + \frac{\epsilon W L_{poly}}{t_{ox}} - 2\epsilon W K_L + 2C_{fringe} \quad (1)$$

where ***e*** is the dielectric permittivity, t_{ox} is the nominal gate oxide thickness, $K_L=L_{smile}/t_{ox}$ (see Fig. 1), $K_T=T_{smile}/t_{ox}$, W is the width of the gate in the third dimension, and C_{fringe} is the fringing capacitance on each side of the gate. The first three terms in Eq. 1 correspond to C_{pp}. The final term, C_{fringe} ($=C_1+C_2+C_3$), arises from *non-vertical* electric field lines, irrespective of which location on the gate these lines arise from.

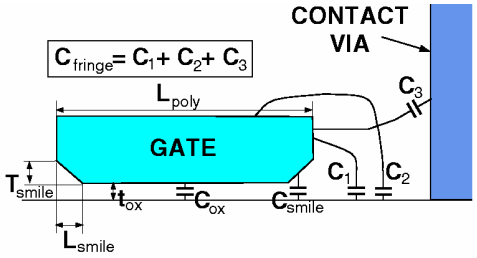

Fig. 1. Cross-section of the gate for a symmetric MOSFET. The "smiling-poly" effect is characterized by $T_{smile}=K_T t_{ox}$ and $L_{smile}=K_L t_{ox}$. C_{fringe} is composed of components C_1, C_2 and C_3, as shown.

Gate capacitance (C_{GG}) affects the digital switching delay of the device, while C_{GD} (one of the components of C_{GG}) is critical for analog operation because of Miller multiplication (which is also present, at a reduced level, for digital circuits). Minimizing C_{GG} is thus an important goal of technology development.

In *30nm* MOSFETs it is possible that smiling-poly effects are negligible, depending on the exact fabrication scheme chosen [1]. If polysilicon gates are used, the doping level in the poly affects the strong-inversion value of C_{GG}. As shown in Fig. 2, poly-depletion effects can cause a

reduction in C_{GG} in strong-inversion. While a capacitance reduction is desirable, the accompanying process-variability, increase in gate resistance, and decrease in transconductance are not. Eq. 2 describes the total gate capacitance as a series combination of the poly-depletion capacitance and the oxide capacitance of Eq. 1:

$$C_{GG,poly-depletion}(V_{GS}) = \frac{C_{GG}(V_{GS})C_{poly}(V_{GS})}{C_{GG}(V_{GS}) + C_{poly}(V_{GS})} < C_{GG}(V_{GS}) \quad (2)$$

where $C_{poly}(V_{GS})$ is the V_{GS}-dependent, non-infinite capacitance which results from depletion in the poly, and $C_{GG}(V_{GS})$ is the capacitance described in Eq. 1. Clearly, in strong inversion, $C_{GG,poly-depletion}<C_{GG}$, as demonstrated in Fig. 2. In the simulations, V_{GS} is taken to unreasonably large values for the sake of clarity.

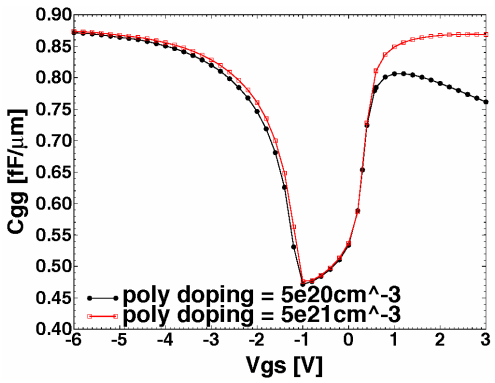

Fig. 2. C_{GG} for two gate poly doping levels for an n-MOSFET with L_{poly}=*30nm* & t_{ox}=*1.5nm*, showing significant poly-depletion effects for poly-doping=$5x10^{20}cm^{-3}$. See Eq. 2 for an explanation. Here V_{DS}=*0* while the V_{GS} range is exaggerated (beyond reasonable V_{DD} values) for clarity.

Note that poly-depletion occurs only in the inversion regime ($V_{GS}>V_{TH}$). In accumulation ($V_{GS}<V_{TH}$) the gate poly is also accumulated. In this simulation using ISE-DESSIS [7], quantum effects in the channel are excluded; accumulation and inversion saturation-capacitance are therefore asymptotically identical for the case where there is no poly depletion. In Fig. 2 it is evident that in accumulation C_{GG} continues to increase slightly at *–6V*. The capacitance does not saturate but can be shown to be inversely proportional to V_{GS}-V_{FB}, where V_{FB} is the flatband voltage.

2 C_{FRINGE}: NUMERICAL SIMULATION

As gate lengths are scaled down, t_{ox} and gate-thickness are not scaled down proportionately. As Fig. 3 shows, this has important implications for gate capacitance: the relative importance of C_{fringe} as a percentage of the total gate capacitance increases significantly, reaching *~25%* for the *30nm* bulk MOSFET. Consequently it is important to understand the origin of C_{fringe}, to generate analytical models for it and to discover ways of minimizing it. Here we examine first the dependence of C_{fringe} on gate-shape, gate-thickness and dielectric properties. Numerical simulations using ISE-DESSIS were performed to obtain physical insight into C_{fringe}. As shown in Fig. 4, fringing field lines were evaluated for two idealized structures with infinitesimally thin gates. In Fig. 4(a) no dielectric overlayer is present in the structure, as a result of which C_{fringe} components C_2 and C_3 (see Fig. 1) are not evaluated. Moreover due to the zero thickness of the gate, C_1 is under-estimated. In the structure of Fig. 4(b), the oxide overlayer permits the simulation of component C_2, such that only C_3 is not evaluated. Here also C_1 is under-estimated due to the infinitesimally thin gate. It is clear from Fig. 4(b) that fringing field lines originating from the top of the gate are numerous and therefore C_2 is significant.

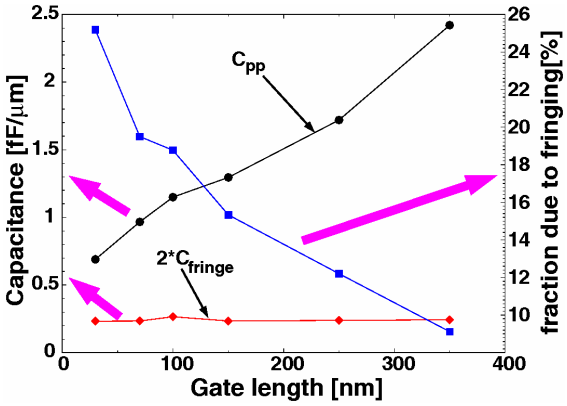

Fig. 3. Parallel-plate and fringe components of the strong-accumulation gate capacitance (C_{GG}). Over technology generations, L_{poly}, t_{ox} and t_{poly} do not scale in proportion to each other. With decreasing L_{poly}, C_{fringe} increases as a fraction of the total capacitance, and hence should be carefully studied.

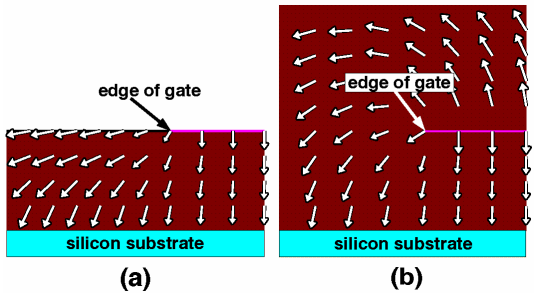

Fig. 4. Electric field lines from numerical simulations of a MOSFET with **(a)** no dielectric layer (C_2 and C_3 not evaluated) and **(b)** with a dielectric overlayer (C_3 not evaluated). An infinitesimally thin gate is used in each case, implying that C_1 is under-estimated. See Fig. 6 and Table 1 for a comparison of C_{GG} values. A sufficiently fine grid has been used for these simulations.

The complexity of the numerically simulated structure can be increased further, as in Fig. 5, to accommodate more

of the fringing capacitance. Fig. 5(a) shows the field lines for a structure with a realistic gate thickness, but without a dielectric overlayer, such that C_1 is accurately simulated but C_2 and C_3 are neglected. The structure in Fig. 5(b) includes the overlayer, thus accurately capturing both C_1 and C_2. The structure in Fig. 5(c) has contact vias placed close to the gate, and thus includes component C_3 as well. However, it must be noted that typical processing constraints require vias to be placed far enough away from the gate stack such that $C_3 << C_1+C_2$. In light of this, the structure of interest for two-dimensional numerical simulations is that of Fig. 5(b), which includes the fringing components of practical importance. It must be emphasized that in addition to including all the relevant layers and regions, a sufficiently fine grid is necessary to allow accurate numerical simulation of fringing field lines.

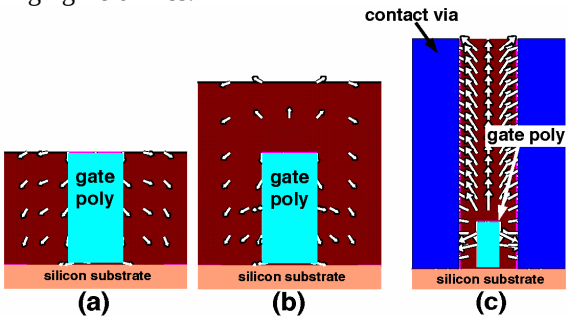

Fig. 5. Electric field lines from numerical simulations of a MOSFET with **(a)** a poly gate but no dielectric overlayer (C_1 included), **(b)** a poly gate and a sufficiently thick dielectric overlayer (C_1, C_2 included), and **(c)** contact vias placed close to the poly gate (C_1, C_2, C_3 included). In practice the vias are placed far from the gate such that $C_3 << C_1+C_2$. The structure in 5(b) is thus of greatest interest.

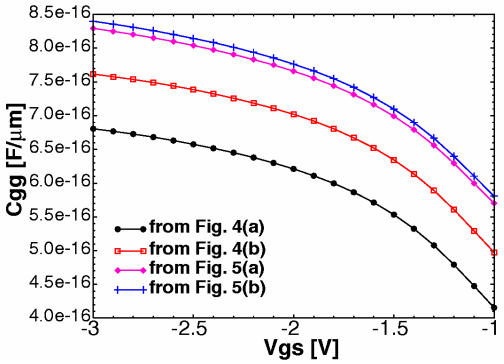

Fig. 6. C_{GG} for the structures in Figs. 4 & 5 in the accumulation regime. The differences are due to C_{fringe}, which is included to an increasing extent when going from Fig. 4(a) to Fig. 5(b). C_{fringe} is independent of V_{GS} and poly-depletion effects.

Fig. 6 shows accumulation regime gate-capacitance as a function of V_{GS} for the structures in Fig.s 4(a), 4(b), 5(a) and 5(b), in order of increasing complexity. (Accumulation-rather than inversion-regime capacitance is plotted in order to remove the confounding effects of poly-depletion). The differences in the four curves are due to C_{fringe}.

3 C_{FRINGE}: ANALYTICAL MODELING

In this section analytical models are derived for the various structures used in numerical simulation, and a comparison is made to values from numerical simulations. Analytical modeling of fringing capacitance, important for device design and circuit simulation, involves the well-known Schwarz-Christoffel (S-C) transformation [6] of complex analysis. In strong inversion, a layer of mobile electrons forms under the gate in an N-MOSFET, which results in a highly conductive layer stretching from the source contact to the drain contact. In accumulation, a layer of holes forms under the gate. In either situation, from the point of view of C_{fringe} evaluation, the silicon substrate (for $V_{DS}=0$) can be thought of as an infinitesimally thin and infinitely conducting plate. Similarly, in strong accumulation the gate can be thought of as a region of infinite conductivity, whether it is infinitesimally thin or has finite thickness. The analytical problem therefore simplifies to the evaluation of capacitance between perfect conductors in the presence of dielectric layers.

In order to derive an analytical formula for the structure in Fig. 4(a), a formula already available [8] for a related structure is altered suitably. The resulting formula:

$$C_{GG} = \epsilon \frac{L_{poly}}{t_{ox}} + \frac{16\epsilon}{\pi} \sum_{n=1,3,5,\ldots} \frac{1}{n^3} \frac{sinh\left(\frac{n\pi(L-L_{poly})}{2t_{ox}}\right)}{sinh^2\left(\frac{n\pi(L-L_{poly})}{4t_{ox}}\right)} \quad (3a)$$

is easily evaluated for the structure in Fig. 4(a) and yields a value of *0.728fF/mm* for the strong-accumulation gate capacitance. In the equation L is the length of the active area (typically $3xL_{poly}$). The first term is C_{pp} while the second term is $2C_{fringe}$ (we ignore the smiling poly effect here). Note that all the formulae provided in this section are for capacitance *per unit width* (in the third dimension).

C_{GG} for the structure in Fig. 4(b) is derived through S-C analysis, and the resulting approximation is:

$$C_{GG} = \epsilon \frac{L_{poly}}{t_{ox}} + \frac{\epsilon}{\pi}\left\{1 + ln\left(\pi \frac{L_{poly}}{t_{ox}}\right)\right\}, \quad (3b)$$

where once again the first term is C_{pp} and the second term is the total fringing capacitance, $2C_{fringe}$.

The S-C analysis for the structure in Fig. 5(a) is quite involved. However, approximate closed-form solutions are possible. A structure similar to that of Fig. 5(a) has been analyzed by other researchers, and it is possible to use Chang's result for a rectangular conductor flanked by two parallel ground planes [9]. By recognizing that the fringing capacitance in the structure in Fig. 5(a) is exactly half that in the structure considered in [9], a relevant formula can be derived:

$$C_{GG} = \epsilon \frac{L_{poly}}{t_{ox}} + \frac{2\epsilon}{\pi}\left[4\alpha\, tanh^{-1}\left(\frac{1}{\sqrt{q}}\right) + ln\left(\frac{q^2-1}{4q}\right) - 2tanh^{-1}\left(\frac{1}{q}\right)\right],$$

$$q = 2\alpha^2 - 1 + \sqrt{\left(2\alpha^2-1\right)^2 - 1}\,, \qquad (3c)$$

where $a=1+t_{poly}/t_{ox}$, and t_{poly} is the thickness of the gate. Note again that in the expression (3c) for C_{GG}, the first term is C_{pp} while the second term is $2C_{fringe}$.

Finally, the derivation in [9] for a single conductor over a ground plane can be simplified greatly to yield a formula for the capacitance in the structure of Fig. 5(b):

$$C_{GG} = \epsilon \frac{L_{poly}}{t_{ox}} + \frac{2\epsilon}{\pi}\left[1 + 2\alpha\, tanh^{-1}\left(\frac{1}{\sqrt{q}}\right) + ln\left(\frac{q-1}{4q}\right)\right] + \frac{2\epsilon}{\pi} ln\{2\eta + (q+1)ln[max(q,\eta)]\}\,, \qquad (3d)$$

where

$$\eta = \sqrt{q}\left[\frac{\pi L_{poly}}{2t_{ox}} + \alpha + \alpha ln\left(\frac{4}{q-1}\right) - 2tanh^{-1}\left(\frac{1}{\sqrt{q}}\right)\right],$$

and a and q are the same as in Eq. 3(c). Note that in these MOSFETs since $L_{poly}/t_{ox} >> 1$, an iterative solution is not necessary and Eq. 3(d) can be applied directly.

Table 1. Strong-accumulation capacitance obtained from numerical simulations and analytical formulae for the structures in Figs. 4(a), 4(b), 5(a) and 5(b). Here $t_{poly}=60nm$.

Strong accumulation gate-capacitance (all values in fF/μm)			
structure	DESSIS Vgs=-6V	S-C analysis (Eq. 3)	error
MOS capacitor	0.684	ε*Lpoly/tox*=0.691	1.0%
Fig. 4(a)	0.719	0.728 (Eq. 3(a))	1.2%
Fig. 4(b)	0.805	0.804 (Eq. 3(b))	-0.1%
Fig. 5(a)	0.879	0.898 (Eq. 3(c))	2.1%
Fig. 5(b)	0.890	0.923 (Eq. 3(d))	3.6%

Now that fringing capacitance formulae are available for the various structures, a comparison of numerical and analytical results is possible, as in Table 1. In each case the fit is better than 4%, indicating the strength of the approximations made. For the structures in Fig.s 5(a)-(b) with polysilicon gates, it is evident (from Fig. 2) that C_{GG} does not saturate even at V_{GS}=-6V. This implies that the error is smaller than listed in Table 1.

4 CONCLUSIONS

In this work we have highlighted the growing importance of C_{fringe} with shrinking MOSFET dimensions. Important considerations in the numerical simulations of this parasitic capacitance have been discussed, and analytical formulae have been derived for C_{fringe}. Poly-depletion and smiling poly effects have been explored semi-quantitatively. Since C_{fringe} depends on geometry and dielectric properties only and does not depend on voltages, it can be included in the evaluation of parasitic capacitance through the use of a single additive parameter. The formulae presented in this work can easily be incorporated into a circuit simulator, resulting in a boost in the accuracy of the structure-dependent parasitic capacitance. Other, more complex gate geometries can also be analyzed using Schwarz-Christoffel analysis. Mathematical engines such as MATLAB™ may be used for cases where analytical approximations are difficult to obtain. The C_{fringe} formulae derived in this work apply to pseudo-2D structures. If three-dimensional effects are important, for a minimum width MOSFET, for example, new formulae will have to be generated. However even with shrinking gate lengths, it is necessary to maintain certain current drives, which implies that gate width will not scale as rapidly. Eq. 3 should therefore be usable for many generations.

ACKNOWLEDGMENTS

We would like to thank Papu Maniar and his team for providing a context for this work and Len Borucki, Jim Andrews and Prof. Hao Ling for valuable discussions.

This paper is dedicated to the memory of Philip Thomas, whose tireless support made our work so much more productive and enjoyable.

REFERENCES

[1] S. Deleonibus *et al*, IEEE Elec. Dev. Lett., 21 (4), April 2000, p. 173-175.

[2] Yuan Taur *et al*, IEEE IEDM Tech. Dig., 1998, p. 789-792.

[3] D. A. Antoniadis *et al*, Aug. 1999, http://www-mtl.mit.edu/Well/

[4] ITRS Roadmap, Semiconductor Industry Association, http://semichips.org/index2.htm

[5] M. Stecher, "Eine Strategie zur Kopplung von Prozeβ und Bauelemente Simulation," dissertation, Tech. Univ. of Aachen, Shaker Publications, 1995. ISBN: 3-8265-0605-7

[6] Zeev Nehari, "Conformal Mapping," Dover, 1975, p. 194.

[7] Integrated Systems Engineering, Zurich, Switzerland, http://www.ise.com

[8] L. N. Dworsky, "Modern Transmission Line Theory and Applications," Wiley, 1979, p. 183.

[9] W. H. Chang, IEEE Trans. on Microwave Theory and Tech., Sep. 1976, p. 608-611 and Aug. 1977, p. 712.