Criteria for Plasmon-Enhanced Electron Drag in Si Metal–Oxide–Semiconductor Devices

Ming-Jer Chen, *Senior Member, IEEE*, Shang-Hsun Hsieh, Yu-Chiao Liao, Chuan-Li Chen, and Ming-Fu Tsai

Abstract—Plasmons in the highly doped source and drain regions of silicon field-effect transistors can strongly drag channel electrons via long-range collective Coulomb interactions and cause deteriorations in overall device performances. To examine such interactions, two different methods have been published: 1) sophisticated device simulations and 2) carefully calibrated experiments. To provide a more transparent understanding, we propose a third method in terms of two criteria: 1) one for the occurrence of the plasmon resonance and 2) the other for the strength of the plasmons. The former is determined based on our published dragged mobility data due to the interface plasmons and the bulk plasmons in gate. The latter makes use of our more recently experimentally extracted potential fluctuations due to plasmons in crystalline silicon. The effects of the temperature on the criteria are considered. It is a straightforward task to confirm that for a channel density larger than approximately 5×10^{12} cm^{-2}, the source and drain plasmons act as key limiters in silicon device scaling. Therefore, the underlying device physics, modeling, simulations, experimental analyses, and data interpretation may be inaccurate if the limiting factors are not incorporated.

Index Terms—Drag, field-effect transistors (FETs), fluctuations, long-range Coulomb interactions, plasmon, resonance, scaling, transport.

I. INTRODUCTION

IN THE semiconductor industry, silicon-based field-effect transistors (FETs) have been aggressively miniaturized in the past four decades and will expectedly enter the 5-nm technology node within the next few years [1]. At such ultra-short device regimes, many-body effects emerge in terms of plasmons [2], particularly in the highly doped source and drain regions [3]. Plasmons are quantized collective modes of strongly interacting electrons and arise from excited potential fluctuations resulting from charge density fluctuations in plasma-like environments [2]–[5]. Plasmons and their excited potential fluctuations can penetrate into the channel region and strongly drag channel electrons via long-range collective Coulomb interactions [3], [5]. To examine such interactions and their effects, two different methods have been published: a sophisticated device simulation [3], [5], [6] and a carefully-calibrated experiment [7].
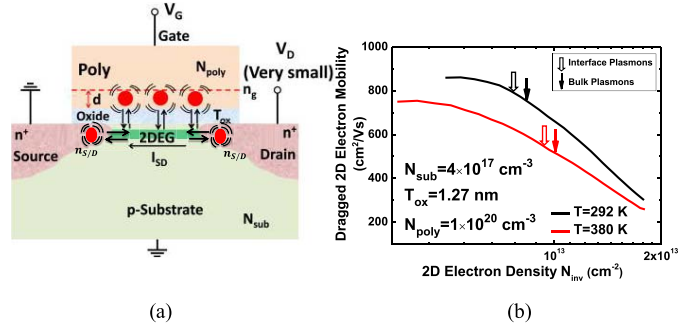
Fig. 1.   (a) Cross section view of silicon FET structure showing the interactions of 2D electron gas with bulk plasmons in source and drain and with bulk and interface plasmons in gate. For a given channel electron density N_{inv}, there may be 3D electrons with a density n_g at a depth d from the poly/oxide interface and those with a density n_{sd} inside the source and drain, each of which can produce a strong drag on the channel electron current through the plasmon resonance (vibrating colored dots with the opposite arrows). Gate plasmons dominate in affecting the long-channel device. (b) Experimental data [8] of dragged channel electron mobility versus N_{inv} for two temperatures. Main process parameters are given. The solid arrows represent the minimum resonance N_{inv} for bulk plasmons and the empty arrows represent one resonance N_{inv} for interface plasmons. Bulk plasmons dominate in the high N_{inv} region, whereas interface plasmons dominate in the low N_{inv} region. Increasing temperature accelerates the drag process, because plasmon-excited potential fluctuations increase with temperature.

To provide a more transparent explanation, we propose a third method in terms of the two criteria by which one can determine how strongly source and drain plasmons will drag channel electrons. These criteria are experimentally determined on a long-channel silicon FET (see Fig. 1(a)). Relative to high-k, metal gate long-channel FETs, plasmons naturally exist in gate and effectively dominate over those of source and drain. In addition, their long-range collective Coulomb interactions with the underlying channel electrons are relatively uniform, thus greatly easing the task of experimentally determining the two criteria and making a close connection to the drag data (see Fig. 1(b)).

II. CRITERIA FOR GATE PLASMONS

In his pioneering simulation task on long-channel silicon FET, Fischetti [5] showed that for a given channel density N_{inv}, despite the many different electron densities and hence, many different plasmon energies or frequencies in gate (including the depletion region), one and only one electron density n_g with its own plasmon energy $E_{plasmon}$ can critically produce the plasmon resonance, as explicitly reflected by a significant drop or "tip" in the channel mobility curve. This links the $E_{plasmon}$ to the Fermi energy E_F (with
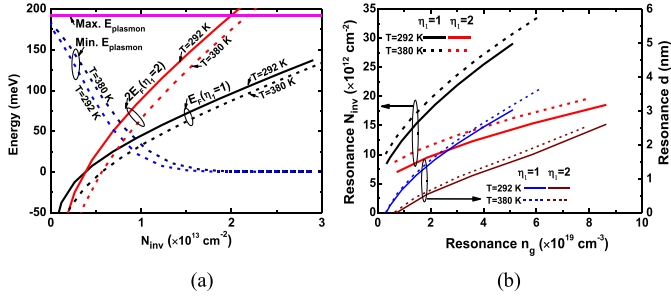
Fig. 2.   (a) Simulated 2D E_F and $2E_F$ and 3D maximum and minimum plasmon energy versus channel electron density for two temperatures. (b) Resulting resonance 3D electron density, plotted versus corresponding 2D electron density and corresponding position d in polysilicon depletion region.

Fig. 3.   (a) Plasmon-excited silicon potential fluctuation data σ [9] and its product with the oxide penetration coefficient $\eta_2 = 0.5$, versus 3D electron density. (b) Scatter plot of experimental dragged mobility in Fig. 1(b) versus the potential fluctuation value obtained by combining Fig. 2(b), Fig. 3(a), and Eq.(2). Four different conditions of T and η_1 were considered, but only two showed considerable overlap. This led to an optimum η_1 of 1.

respect to lowest subband level) of the 2D channel electrons. We determine the E_F via solving the Schrödinger-Poisson equations in a *static* (i.e., no fluctuations due to plasmons) MOS system. The same static MOS system had earlier been adopted in the experimental determination of the dragged mobility [8], as shown in Fig. 1(b) for two temperatures. This leads to the following semi-empirical formula in this work describing the occurrence of the plasmon resonance:

$$E_{Plasmon} = \eta_1 E_F \tag{1}$$

The ratio of $E_{plasmon}$ to E_F, η_1, is the first criterion. To experimentally determine η_1, we start by solving the Schrödinger-Poisson equations with the process parameters (labeled in Fig. 1(b)) as input. The *bulk* plasmon energies for the maximum 3D electron density n_g (equal to polysilicon dopant concentration), the minimum n_g (at the surface), and $\eta_1 E_F$, versus N_{inv} can be created, as demonstrated in Fig. 2(a). The intersection between $\eta_1 E_F$ and the plasmon energy associated with the minimum n_g can determine the minimum resonance N_{inv} and the corresponding resonance n_g at the surface. We adjusted η_1 until the resulting dragged mobility matched the experimental mobility. The optimum η_1 was 1, regardless of the temperature, as will be demonstrated later. For a higher N_{inv}, the corresponding resonance n_g in the poly depletion region can be accordingly obtained (i.e., the $E_{plasmon}$ is made equal to E_F times η_1). The resulting resonance n_g and its depth d from the poly/SiO$_2$ interface both increase with N_{inv}, as depicted in Fig. 2(b).

Next, the strength of the plasmons associated with a resonance n_g is considered as a second criterion. This can be effectively dealt with by means of the excited potential fluctuations. Experimental values of the potential fluctuation σ due to plasmons in *crystalline* silicon [9] can be empirically expressed as a function of the 3D electron density n and the temperature T:

$$\sigma = 62.3 \text{ meV} \left(\frac{n}{2 \times 10^{19} \text{ cm}^{-3}} \right)^{0.18} \left(\frac{T}{295 \text{ K}} \right)^{0.48} \tag{2}$$

While applying to the polysilicon gate, the collision damping in a fluctuating system, in combination with the grain-size effect [10], should be taken into account, as had been demonstrated by Sano's group [6] in simulation of crystalline silicon. We noticed that the experimental sheet resistance
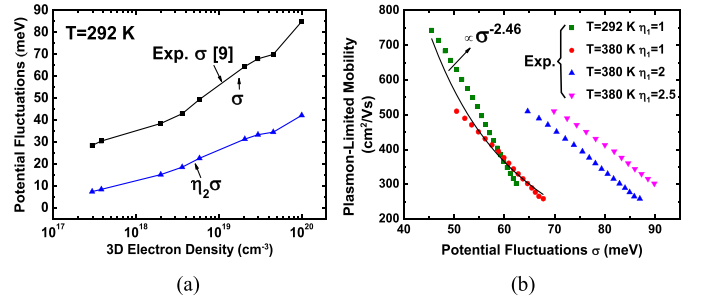
was 15.4 Ω/□ on a polysilicon layer 120-nm thick in the same industrial manufacturing process [8]. The corresponding polysilicon electron mobility was approximately 80 cm^2/Vs, a value quite close to that of the crystalline silicon (typically 100 cm^2/Vs at 10^{20} cm^{-3} doping). Thus, we can argue that polysilicon crystallization occurred in the device under study. The TEM picture [8] revealed that there was a large area of crystalline-like patterns in the polysilicon region. Specifically, we observed that the grain size spans from 10 nm to 100 nm, which is much larger than the Landau damping wavelength of 0.5 nm at 10^{20} cm^{-3} doping. Therefore, the grain-size issue [10] is not significant in this work. The experimentally determined silicon σ [9] implicitly contains the collision damping (2) and can be applied to the polysilicon side of the presented device.

We show in Fig. 3(a) the experimental silicon σ [9] versus the 3D electron density at 292K. These plasmons penetrate through the gate oxide to the 2D channel. Fischetti and Laux [3] simulated potential fluctuations in the penetration path (see [3, Fig. 9]) and showed that the penetrated potential fluctuations through the oxide dropped by a factor of approximately 0.5. This value remained until a few nanometers deep [3], which covered a distance approximately 1 nm from the surface to the 2D gas centroid. To account for this, we introduce the penetration coefficient $\eta_2 = 0.5$. Multiplying σ by η_2, as shown in Fig. 3(a), represents the effective potential fluctuation seen by the 2D electrons.

Next, we combine the two criteria. This is performed by combining Figs. 1(b), 2(b), and 3(a) with Eq. (2), leading to a scatter plot showing a power-law relation of the experimental dragged mobility versus the corresponding potential fluctuation. The power-law exponent of approximately -2.46 is quite close to that (-2) claimed elsewhere [9]. Remarkably, for $\eta_1 = 1$, the same power-law relation applies, regardless of the temperature. This strongly supports the validity of the proposed criteria.

The bulk plasmons are accountable in the high N_{inv} region but cannot explain the experimental data (see Fig. 1(b)) in the low N_{inv} region. We attribute this to the interface plasmons in gate. We re-calculated Eq.(1) with the $E_{plasmon}$ divided by $\sqrt{2}$ (the quantity of the interface modes [4], close to that

in the zero oxide thickness limit and in the long wavelength limit [5]). We found that the calculated resonance N_{inv} due to the interface plasmons is less than the bulk plasmons, as depicted in Fig. 1(b). In addition, *one must keep in mind that the lower N_{inv} near the resonance N_{inv} does not necessarily indicate the absence of the actual plasmon resonance. The reason is that the aforementioned "tip" due to the interface plasmons spans a considerable range [5] including those of the N_{inv} below that of the calculated resonance.* Furthermore, we can argue that *(i) in the low N_{inv} region where the potential fluctuations due to the interface plasmons are very large, 2D channel electrons tend to emit 2D plasmons, some of which enter the interface plasmons in gate; and (ii) in the high N_{inv} region featured by both large potential fluctuations and bulk plasmon resonance, the 2D electrons are likely to emit 3D plasmons to gate.*

III. APPLICATION TO SOURCE AND DRAIN PLASMONS

The aforementioned experimentally-determined criteria can be readily applied to the plasmons in the source and drain, for $\eta_2 = 1$ due to the absence of the gate oxide. Consequently, the drag rate due to the source and drain plasmons is significantly faster than that of the gate plasmons, as seen from the larger potential fluctuations in Fig. 3(a). Such enhanced drag by the source and drain plasmons took place for certain critical operating conditions in terms of the N_{inv} above its onset of approximately 5×10^{12} cm^{-2}, regardless of temperature, as seen from Fig. 2(a). This onset was experimentally observed in our previous work [7].

Evidently, the source and drain act as large potential fluctuating reservoirs and not simply as conventional thermal equilibrium reservoirs. Thus, the channel electrons near the source or drain encounter the strongest plasmons, whereas collision damping events are encountered by the remaining channel electrons far away from the source and drain. In other words, the shorter the channel, the more pronounced the effect of the long-range collective Coulomb interactions. We want to stress that the positioning effect of the resonance n_{sd} in the source/drain extension doping profile is not significant. This is because the 3D region under study can be considered to be a quasi-isolated system, where many self-sustainable plasmons with different frequencies behave like electromagnetic waves, regardless of the position of the responsible n_{sd}. This is also the case for n_g in gate with its distance d from the interface. Support for these observations can also be found from the simulated potential fluctuations reported by Fischetti and Laux [3], which revealed that such a distance

factor was not significant. Therefore, the source and drain plasmons, as key limiters in the silicon device scaling toward a 5-nm node [1], are explained. *There have been many articles devoted to device physics, modeling, simulations, experimental analyses and data interpretation where long-range Coulomb interactions and fluctuating environments were not addressed.* Reported results with no considerations for such interactions and fluctuations may be incorrect or misleading.

IV. CONCLUSION

The proposed criteria have exhibited the ability to provide a more transparent understanding of plasmon-enhanced drag. These criteria have been experimentally determined in a long-channel device with plasmons in gate. The source and drain plasmons, as key factors that limit the scaling of silicon devices, have been confirmed. Additionally, critical operating conditions have been introduced, which should be incorporated into device physics, modeling, simulations, experimental analyses and data interpretation, and should be considered for the effects of the long-range Coulomb interactions and potential fluctuations.

REFERENCES

[1] (2013). *International Technology Roadmap for Semiconductors (ITRS)*. [Online]. Available: http://www.itrs.net

[2] D. Pines and D. Bohm, "A collective description of electron interactions: II. Collective vs individual particle aspects of the interactions," *Phys. Rev.*, vol. 85, no. 2, pp. 338–353, Jan. 1952.

[3] M. V. Fischetti and S. E. Laux, "Long-range Coulomb interactions in small Si devices. Part I: Performance and reliability," *J. Appl. Phys.*, vol. 89, no. 2, pp. 1205–1231, Jan. 2001.

[4] N. W. Ashcroft and N. David Mermin, *Solid State Physics*. Belmont, CA, USA: Brooks/Cole, 1976.

[5] M. V. Fischetti, "Long-range Coulomb interactions in small Si devices. Part II. Effective electron mobility in thin-oxide structures," *J. Appl. Phys.*, vol. 89, no. 2, pp. 1232–1250, Jan. 2001.

[6] K. Nakanishi, T. Uechi, and N. Sano, "Self-consistent Monte Carlo device simulations under nano-scale device structures: Role of Coulomb interaction, degeneracy, and boundary condition," in *Proc. IEEE IEDM*, Dec. 2009, pp. 1–4.

[7] M.-J. Chen *et al.*, "Probing long-range Coulomb interactions in nanoscale MOSFETs," *IEEE Electron Device Lett.*, vol. 34, no. 12, pp. 1563–1565, Dec. 2013.

[8] M.-J. Chen *et al.*, "Temperature-oriented mobility measurement and simulation to assess surface roughness in ultrathin-gate-oxide (∼1 nm) nMOSFETs and its TEM evidence," *IEEE Trans. Electron Devices*, vol. 59, no. 4, pp. 949–955, Apr. 2012.

[9] M.-J. Chen *et al.*, "Plasmons-enhanced minority-carrier injection as a measure of potential fluctuations in heavily doped silicon," *IEEE Electron Device Lett.*, vol. 35, no. 7, pp. 708–710, Jul. 2014.

[10] B. Laikhtman and P. M. Solomon, "Electron-drag effect in Si metal-oxide-semiconductor devices with thin oxide layers," *Phys. Rev. B*, vol. 72, p. 125338, Sep. 2005.