

Quantum Capacitance in Scaled Down III-V FETs

Donghyun Jin, Daehyun Kim, Taewoo Kim and Jesús A. del Alamo

Massachusetts Institute of Technology, Cambridge, MA 02139; jinnara@mit.edu; +1-617-253-1620; FAX +1-617-258-7393

Abstract

We have built a physical gate capacitance model for III-V FETs that incorporates quantum capacitance and centroid capacitance in the channel. We verified its validity with simulations (Nextnano) and experimental measurements on High Electron Mobility Transistors (HEMTs) with InAs and InGaAs channels down to 30 nm in gate length. Our model confirms that in the operational range of these devices, the quantum capacitance significantly lowers the overall gate capacitance. In addition, the channel centroid capacitance is also found to have a significant impact on gate capacitance. Our model provides a number of suggestions for capacitance scaling in future III-V FETs.

Introduction

As Si CMOS approaches the end of the roadmap, finding a new transistor technology that allows the extension of Moore's law has become a technical problem of great significance. Among the various candidates, III-V-based Field-Effect Transistors represent a very promising technology. In particular, low-effective mass materials with high electron velocities, such as InGaAs and InAs are of great interest [1,2].

A concern with this approach is the relatively small inversion-layer capacitance that is associated with the channel and the limits that this imposes on the gate capacitance that can be attained from barrier thickness scaling [3]. This can seriously limit the current driving ability of scaled down devices. The inversion-layer capacitance has two main contributions: quantum capacitance [4] and centroid capacitance [5]. The first one originates in the penetration of the Fermi level inside the 2D subbands of a quantum well due to the finite density of states. The second one is related to the shape of the charge distribution in the inversion layer. In low effective mass III-V channels, both capacitances can be relatively small.

In order to understand the scaling potential of III-V FETs, we have built a physical gate capacitance model and compared it with experimental measurements on InGaAs- and InAs-channel HEMTs. From this analysis, we conclude that the relatively small quantum capacitance of InAs-rich channels significantly limits the overall gate capacitance in scaled down designs. In addition, our experiments suggest a large increase of the in-plane effective mass in very thin channel designs as a result of non-parabolicity and strain. This should help to achieve a relatively high electron concentration in future scaled down high-k dielectric III-V FETs.

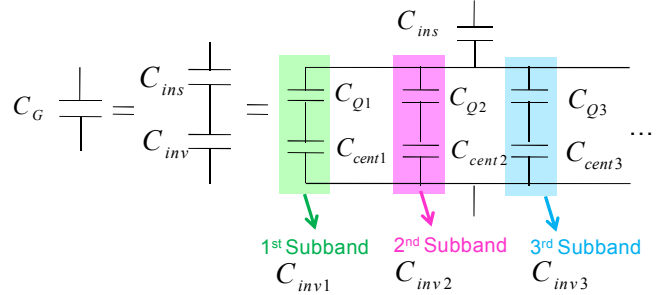


Fig 1. Equivalent circuit diagram of gate capacitance in a III-V FET

• Insulator capacitance

$$C_{ins} = \frac{\epsilon}{t_{ins}}$$

• Centroid capacitance of subband i

$$C_{cent_i} = C_{Q_i} \cdot \frac{\partial(E_F - E_i)}{\partial(E_i - E_C)}$$

• Quantum capacitance of subband i

$$C_{Q_i} = \frac{m_{\parallel}^* q^2}{\pi \hbar^2} \frac{1}{1 + \exp\left(\frac{E_i - E_F}{kT}\right)}$$

• Inversion-layer capacitance

$$C_{inv} = \sum_i \left(\frac{1}{C_{Q_i}} + \frac{1}{C_{cent_i}} \right)^{-1}$$

Fig 2. Physical formulas for capacitance model components of Fig. 1.

Gate Capacitance Model

In this work, the gate capacitance of a III-V FET is modeled as the series combination of the insulator capacitance and the inversion-layer capacitance (**Fig. 1**). This one consists of a parallel combination of the contributions of each occupied electron subband in the channel. For each subband i , the inversion-layer capacitance (C_{inv}) consists of the quantum capacitance (C_{Q_i}) and the centroid capacitance (C_{cent_i}) which are connected in series (**Fig. 1**). We derived this from the definition of inversion-layer capacitance:

$$C_{inv} = \frac{\partial(-Q_s)}{\partial\psi_s} = \frac{q\partial(-Q_s)}{\partial(E_F - E_C)} \quad (1)$$

$$Q_s = \sum_i Q_i = \sum_i \int_{E_i}^{\infty} \frac{m_{\parallel}^* q}{\pi \hbar^2} \frac{1}{1 + \exp\left(\frac{E - E_F}{kT}\right)} dE \quad (2)$$

$$\text{and } \left(\frac{q\partial(-Q_i)}{\partial(E_F - E_C)} \right)^{-1} = \left(\frac{q\partial(-Q_i)}{\partial(E_F - E_i)} \right)^{-1} + \left(\frac{q\partial(-Q_i)}{\partial(E_i - E_C)} \right)^{-1} \quad (3)$$

where ψ_s is surface potential, and E_C is conduction band edge at the barrier-channel interface on the channel side. From (1), (2) and (3), inversion-layer capacitance is

expressed as $C_{inv} = \sum_i \left(\frac{1}{C_{Q_i}} + \frac{1}{C_{Cent_i}} \right)^{-1}$ where analytic

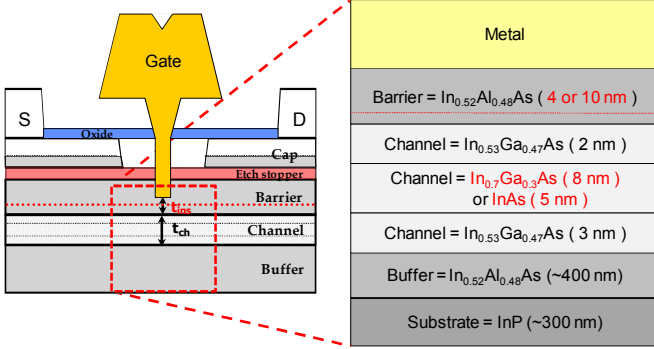


Fig 3. Experimental HEMT cross section and layer details of the three heterostructures explored in this work:

8 nm $\text{In}_{0.7}\text{Ga}_{0.3}\text{As}$ center channel layer, $t_{\text{ch}} = 13$ nm, $t_{\text{ins}} = 4$ nm
 5 nm InAs center channel layer, $t_{\text{ch}} = 10$ nm, $t_{\text{ins}} = 4$ nm
 5 nm InAs center channel layer, $t_{\text{ch}} = 10$ nm, $t_{\text{ins}} = 10$ nm

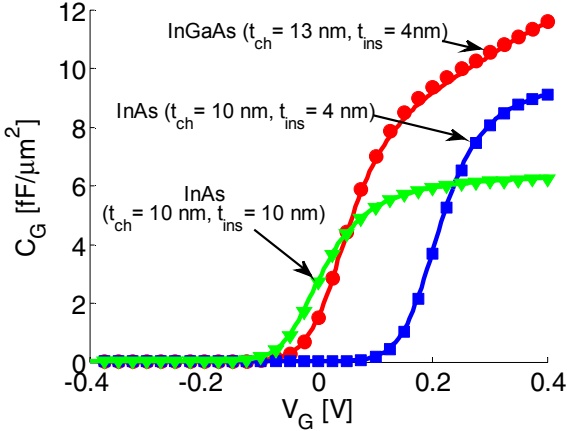
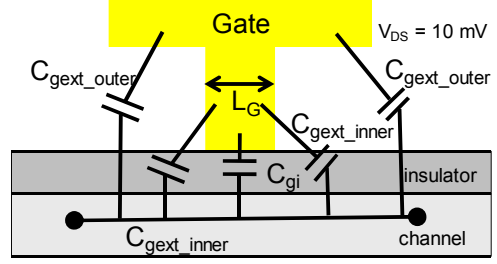


Fig 4. Gate capacitance as a function of gate voltage for three heterostructures. Solid lines are simulation results and symbols are model results.

formulas for C_{Q_i} and C_{Cent_i} are given in Fig. 2. If the location of each subband energy level (E_i) and the Fermi level are known with respect to the conduction band edge, then all capacitance components can be evaluated. Rather than attempt to analytically solve the quantization problem for realistic FET structures, in this work, we used a one-dimensional Poisson-Schrodinger solver (Nextnano) to obtain the subband energy levels as a function of V_G .

We have investigated the gate capacitance in three HEMT structures with different channel and barrier designs (Fig. 3). In essence, we have two channels, one with $\text{In}_{0.7}\text{Ga}_{0.3}\text{As}$ at the center of a 13 nm thick channel [1] and another one with pure InAs at the center of a 10 nm channel [2]. In both cases, the channel cladding is $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ (2+3 nm). For the InAs -channel design, we have two $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ barrier thicknesses, t_{ins} , of 4 and 10 nm. For the InGaAs sample, t_{ins} is 4 nm. Before analyzing the experimental gate capacitance in these devices, we verified the accuracy of our physical gate capacitance model. Fig. 4 shows that our model agrees very well with simulation results obtained directly from Nextnano through $C_G = d(-Q_S)/dV_G$ for the three heterostructures. In the



$$C_G \text{ (in fF/mm)} = C_{g_i} \text{ (in fF/}\mu\text{m}^2) \times L_G + 2 \times C_{g_{\text{ext_inner}}} (V_{GS}) + 2 \times C_{g_{\text{ext_outer}}}$$

Fig 5. Intrinsic and parasitic components of gate capacitance in a HEMT. Two different kinds of parasitic gate capacitance are defined in this model. $C_{g_{\text{ext_outer}}}$ is associated with the top of the T-gate. $C_{g_{\text{ext_inner}}}$ is associated with the sidewall of the gate stem. Our experimental methodology separates all these components.

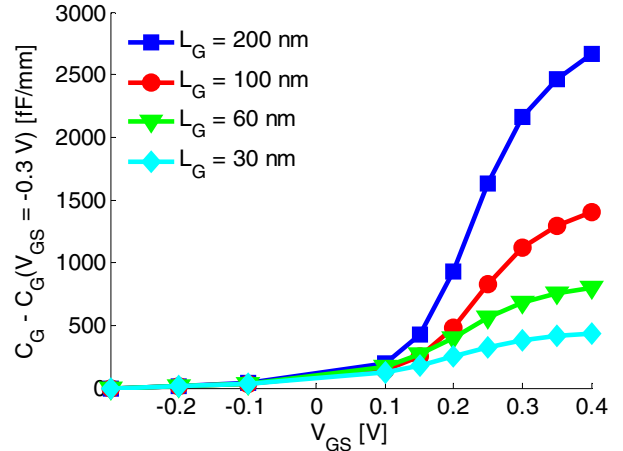


Fig 6. $C_G - C_G(V_{GS} = -0.3 \text{ V})$ vs. V_{GS} in different gate length InAs channel HEMTs with $t_{\text{ins}} = 4$ nm. $C_G(V_{GS} = -0.3 \text{ V})$ corresponds to $C_{g_{\text{ext_outer}}}$ in Fig. 5.

Nextnano simulations, the Schottky barrier height was adjusted to match the experimental threshold voltage.

C_G Measurements on Scaled Down HEMTs

We have compared our model against experimental gate capacitance for these three heterostructures. C_G was obtained from S-parameter measurements [6] of HEMTs with gate lengths (L_G) from 30 to 200 nm in the linear regime ($V_{DS} = 10$ mV) for different values of V_{GS} . After deembedding the pads, we extracted the intrinsic gate capacitance by eliminating all the parasitic components (Fig. 5). We define two different kinds of parasitic gate capacitance. There is an outer component, $C_{g_{\text{ext_outer}}}$, associated with the top of the T-gate. There is also an inner parasitic capacitance, $C_{g_{\text{ext_inner}}}$, that is associated with the sidewall of the gate stem. Unlike $C_{g_{\text{ext_outer}}}$, this parasitic component depends on V_G . In order to eliminate these two parasitic terms, we measured C_G in devices with different L_G as a function of V_{GS} . $2C_{g_{\text{ext_outer}}}$ is taken as $C_G(V_{GS} = -0.3 \text{ V})$ (Fig. 6). $C_{g_{\text{ext_inner}}}(V_G)$ is then obtained from the extrapolation of $C_G - C_G(V_{GS} = -0.3 \text{ V})$ with L_G at each

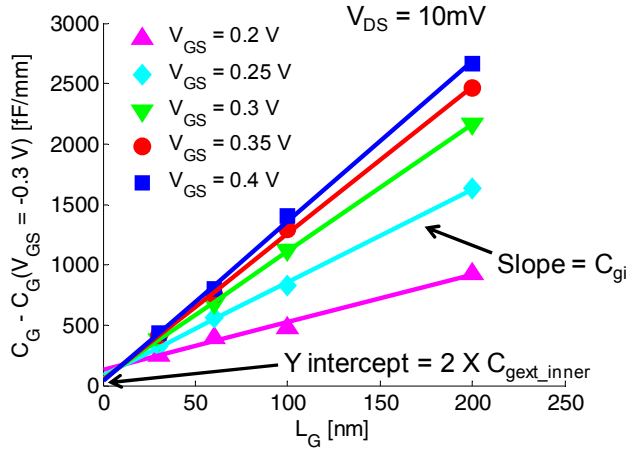


Fig 7. $C_G - C_G(V_{GS} = -0.3V)$ as a function of L_G at different V_{GS} for InAs channel HEMTs with $t_{ins} = 4$ nm. The slope of this line at each bias point corresponds to C_{gi} . The Y intercept is $2C_{gext_inner}$ of Fig. 5.

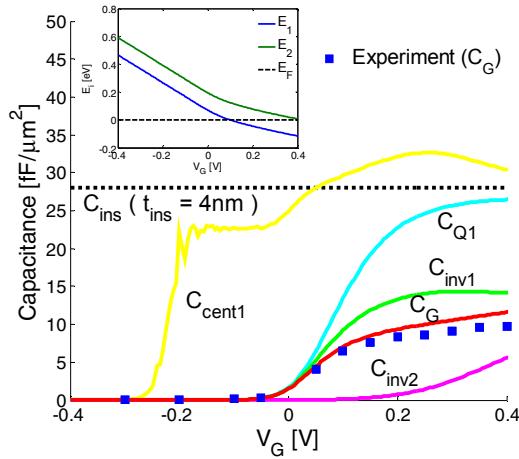


Fig 8. Experimental C_G and modeled C_G components of $In_{0.7}Ga_{0.3}As$ channel structure with $t_{ins} = 4$ nm vs V_G . The inset describes the evolution of the 1st and 2nd subband energy levels with respect to the Fermi level. The 2nd subband is only populated in a minor way.

value of V_{GS} (Fig. 7). The intrinsic gate capacitance per unit area (C_{gi}) corresponds to the slope of these lines (Fig. 7). The intersection is $2C_{gext_inner}$.

Comparison of Model and Experiments

Figs. 8, 9 and 10 show the experimental intrinsic gate capacitance vs V_{GS} for the three heterostructures as well as the gate capacitance components that are derived from our model. The agreement between modeled and experimental capacitance is reasonable although there are some discrepancies that are discussed below. In all three cases, the degradation in overall gate capacitance that comes from the finite inversion-layer capacitance is evident. The measured C_G in strong inversion is between 35% and 62% of C_{ins} .

An additional conclusion is that the 1st subband dominates the overall gate capacitance in the operational range of a scaled down HEMT. This is particularly the case of the InAs channel structures which have significant more channel

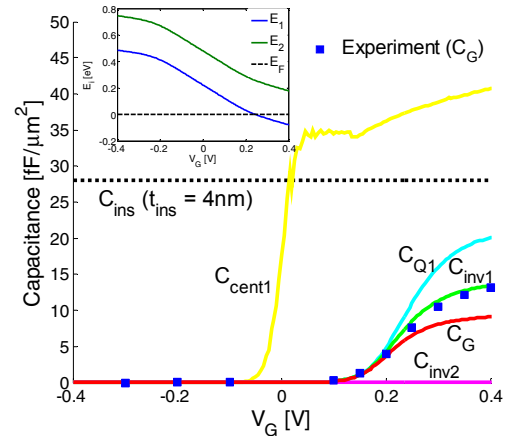


Fig 9. Experimental C_G and modeled C_G components of InAs channel structure with $t_{ins} = 4$ nm vs V_G . The inset describes the evolution of the 1st and 2nd subband energy levels with respect to the Fermi level. The 2nd subband contribution is negligible.

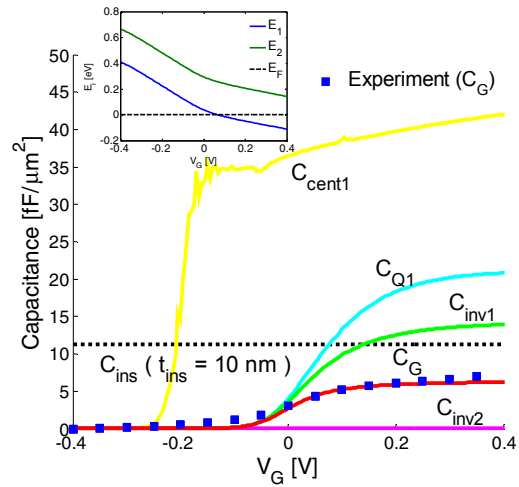


Fig 10. Experimental C_G and modeled C_G components of InAs channel structure with $t_{ins} = 10$ nm vs V_G . The inset describes the evolution of the 1st and 2nd subband energy levels with respect to the Fermi level. C_{ins} is much lower in comparison with Figs. 8 and 9.

quantization due to the thinner channel and the lower effective mass. A third conclusion is that C_{cent1} is also highly relevant to determining C_G . Comparing the two devices with $t_{ins} = 4$ nm in Figs. 8 and 9, we see that C_{cent1} is significantly larger in the thin channel device (Fig. 9) vs. the thicker channel device (Fig. 8). This compensates for the lower quantum capacitance of the InAs device which ends up with a higher overall value of C_G . This suggests that scaling down the channel thickness is an effective way to enhance gate capacitance.

Discussion

The agreement between model and experiments is worst for the $t_{ins} = 4$ nm InAs channel device (Fig. 9). This is the structure in which quantum capacitance is most relevant. This discrepancy cannot come alone from experimental uncertainties in t_{ins} , which is measured by TEM [7]. The

uncertainty in t_{ins} is estimated to be about ± 0.5 nm which translates in error bars in the calculated C_G such as ± 0.6 fF/ μm^2 at $V_G = 0.4$ V as shown in Fig. 11.

In our model, we are limited to the use of spherical bands with a single well defined effective mass for each material which we have set to be equal to the bulk effective mass. Biaxial strain between InAs and InP lattices, and the combination of strong quantization and non-parabolicity are expected to increase the in-plane effective mass (m_{\parallel}^*) [8,9]. We explored the impact of this by artificially selecting a higher value of m_{\parallel}^* for the InAs layer. We have not done the same for the InGaAs cladding layers because we estimate the electron concentration there to be less than 25% of the total for the entire V_G range.

Fig. 11 shows that the agreement between experiments and the model for the InAs-channel devices improves when we increase m_{\parallel}^* of InAs from bulk value ($0.026m_e$) to the value around $0.05m_e$ with $\pm 0.005m_e$ variation. This seems like a large increase but it is actually expected by theoretical and experimental studies of these effects [8-11]. With the combination of thickness uncertainty and effective mass increase, the discrepancy between model and measurements is significantly reduced.

Our model allows us to estimate C_G in future scaled down high-k dielectric III-V FETs. In future devices, the adoption of a high-k gate dielectric and the use of a very thin quantized channel with a low-effective mass material will establish the quantum capacitance of the first subband as the dominant term in C_G . Using our model, we can examine the implications of this. Fig. 12 shows the sheet carrier concentration as a function of gate overdrive for a future 10 nm gate length prototype device with $t_{\text{ch}} = 3$ nm and $t_{\text{ins}} = 2.6$ nm ($\epsilon = 25\epsilon_0$). The dotted lines show the expected N_S for different values of the in-plane effective mass in the channel. Due to the dominance of C_{Q1} , the effective mass in the channel greatly affects N_S . As m_{\parallel}^* increases, so does C_{Q1} and N_S at a given overdrive. It is clear that in future III-V FETs, an enhancement of channel effective mass is essential to attain N_S in the high 10^{12} cm^{-2} range. However this seems eminently feasible through non-parabolicity, the strong quantum confinement expected from a very thin channel [8,10] and by proper engineering of in-grown biaxial strain [9].

Conclusions

We have developed a simple quantitative model for C_G in III-V FETs that includes the quantum capacitance. We validate this model through simulations. The model provides reasonable agreement with experiments on InGaAs and InAs FETs with different designs. Residual discrepancies are likely due to the non-parabolic nature of the bands and biaxial strain. Our model suggests that quantum capacitance will dominate in future scaled III-V FETs. Furthermore, the expected increase of effective mass in thin channel designs will provide the required sheet charge density for high-performance operation.

References

- [1] D.-H. Kim *et al*, IEDM, p. 837, 2006.
- [2] D.-H. Kim *et al*, IEDM, p. 722, 2008.
- [3] S. Takagi *et al*, IEEE TED, p. 2125, 1995.
- [4] S. Luryi *et al*, APL, p. 501, 1988.
- [5] H. Pal *et al*, IEEE TED, p. 904, 2008.
- [6] D. Lovelace *et al*, MTT, p. 865, 1994.
- [7] D.-H. Kim *et al*, EDL, p. 830, 2008.
- [8] B. R. Nag *et al*, APL, p. 2416, 1993.
- [9] C. Kopf *et al*, SSE, p. 1139, 1997
- [10] U. Wiesner *et al*, APL, p. 2520, 1994
- [11] M. Luisier *et al*, IEDM, p. 887, 2008

Acknowledgements

This work was sponsored by Intel Corporation and FCRP-MSD Center. Epitaxial heterostructures were supplied by MBE Technology.

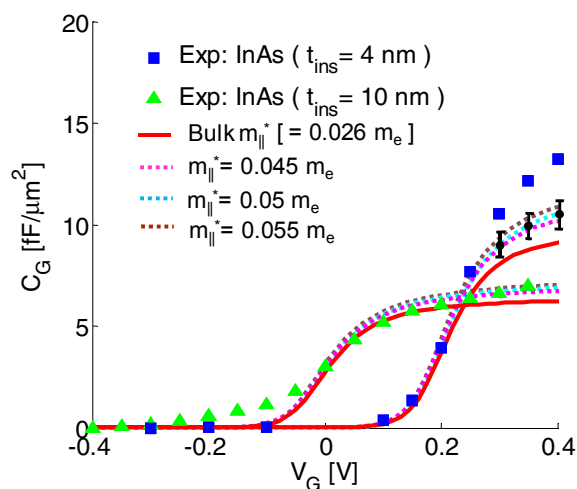


Fig 11. Gate capacitance as a function of gate voltage for the InAs-channel heterostructures. The dotted lines show calculations using heavier values of effective mass in the InAs layer. The error bars indicate the uncertainty in the calculated C_G that comes from an uncertainty of ± 0.5 nm in t_{ins} .

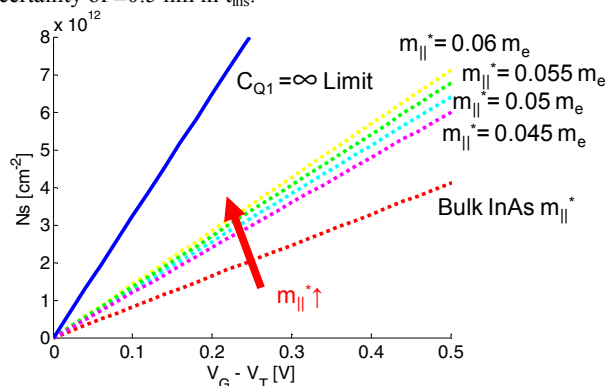


Fig 12. Carrier concentration (N_S) as a function of $V_G - V_T$ for a high-k dielectric 10 nm gate length prototype device with $t_{\text{ch}} = 3$ nm and $t_{\text{ins}} = 2.6$ nm ($\epsilon = 25\epsilon_0$). V_T is defined as V_G at $N_S = 5 \times 10^{10}$ cm^{-2} . Maximum limit when C_{Q1} becomes infinity is shown as a solid line. m_{\parallel}^* increments from strong quantization and biaxial strain can provide high inversion N_S .