Self-Aligned InGaAs FinFETs with 5-nm Fin-Width and 5-nm Gate-Contact Separation

Alon Vardi[1], Lisa Kong[1], Wenjie Lu[1], Xiaowei Cai[1], Xin Zhao[1], Jesús Grajal[1,2] and Jesús A. del Alamo[1]
[1]Microsystems Technology Laboratories, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A
60 Vassar St., Rm. 39-615, Tel: 617-258-5752, Fax: 617-258-7393, Email: alonva@mit.edu
[2]ETSI Telecomunicación, Universidad Politécnica de Madrid, C/ Ramiro de Maeztu 7, 28040 Madrid, Spain

***Abstract*—**We demonstrate self-aligned InGaAs FinFETs with fin widths down to 5 nm fabricated through a CMOS compatible front-end process. Precision dry etching of the recess cap results in metal contacts that are about 5 nm away from the intrinsic portion of the fin. The new process has allowed us to fabricate devices with undoped fins and compare them with delta-doped fins. We find that in highly scaled transistors, undoped fin devices show better OFF-state and a tighter V_T distribution but similar ON-state characteristics, as compared with δ-doped-fin transistors. 2D Poisson-Schrodinger simulations reveal undoped fins making more effective use of the fin height.

I. Introduction

InGaAs is a promising channel material candidate for CMOS technologies beyond the 7 nm node [1-2]. In this dimensional range, only high aspect-ratio (AR) 3D transistors with a fin or nanowire configuration can deliver the necessary performance. Impressive InGaAs FinFET prototypes have recently been demonstrated [3-6]. However, unlike planar InGaAs MOSFETs [8], their performance is lagging behind Si FinFETs. There are many challenging issues that are unique to InGaAs FinFETs that have yet to be tackled. At the point of insertion of this technology, ~5 nm fin widths with steep sidewalls will be required [9-10]. Such aggressively scaled transistors have never been demonstrated in InGaAs. To mitigate short-channel effects and device variations, it is also necessary to remove all dopants from the fin. However, this requires very tight self-alignment between the ohmic contacts and the edge of the conducting channel.

In this work we demonstrate a new process flow that uses dry recess to etch the cap. This results in a very close contact to gate-edge distance of 5 nm. The high process margin of our fabrication flow has yielded devices with fin widths as narrow as 5 nm and channel aspect ratios of 10. The tight self-aligned nature of the device has enabled the fabrication of high-performance devices with fully undoped channels.

II. Process Technology

Two heterostructures have been studied in this work (**Fig. 1**). The significant difference is the presence or absence of a δ-doped InAlAs layer just below the conducting channel**.** For both samples, the channel is 50 nm thick InGaAs lattice matched to InP and the cap consists of 30 nm thick heavily-doped Si:InGaAs. In the delta-doped sample, there is an InP stopper between the cap and the channel which is used to calibrate the dry etching time for both samples.

Our fabrication process (**Fig. 2**) integrates several features developed in our group in the last few years [11-13]. The process starts with sputtering a low-ρ (R_{sh}=5Ω/□) W/Mo ohmic-contact bilayer. This contact-first approach yields outstanding contact resistance in planar and fin devices [11,12].

E-beam lithography is used to define the gate recess. The SiO_2 hard mask and W/Mo contact stack are etched by anisotropic RIE using CF_4:H_2 and SF_6:O_2 chemistries, respectively [11,13]. After a mesa definition step, the highly conductive cap is removed using Cl:BCl_3 dry etching. On the δ-doped sample, the InP stopper is removed controllably using digital etch (DE). This also smooths the top surface of the channel (**Fig. 3a)** [14]. The undoped sample was processed side by side using the calibration established in the delta-doped sample. Due to the 60^0 slope of the dry recess, the highly doped cap extends ~15 nm on top of the fins. The DE pulls back the cap ~5 nm under the metal (**Fig. 4c)**. This is also the distance between the ohmic contact and the fin as compared with ~20 nm in [15] obtained via a wet recess process.

After the last DE cycle, 1/4 monolayers of Al_2O_3/HfO_2 are deposited to passivate the top facet of the fins in the undoped sample where there is no InP stopper. 2 nm of Si_3N_4 is deposited by CVD on both samples as adhesion layer for the subsequent fin etch mask which consists of 90 nm thick Hydrogen Silsesquioxane (HSQ). This is exposed by e-beam lithography.

The fins are then etched in inductive coupled plasma (ICP) using a BCl_3/$SiCl_4$/Ar chemistry (**Fig. 3b**) [16]. This yields fins as narrow as 15 nm with highly vertical walls in the top ~70 nm. The fin height (H_f in Fig. 2) is 170-200 nm. To further thin down the fins and smooth the sidewalls, we perform several cycles of DE [14] (etch rate ~1 nm/cycle per side). This yielded fins as narrow as 5 nm, as seen in Fig. 4 in a finished device.

Immediately after the last DE cycle, 1 monolayer of Al_2O_3 and 3 nm HfO_2 (EOT~0.8 nm) are deposited by ALD. Here, we increased the EOT w.r.t previous experiments, in order to reduce gate leakage at long gate lengths and enable a detailed study of the mobility. The frontend process is completed by Mo gate metal sputtering and definition using SF_6:O_2 dry etch. The gate covers the fins completely and overlaps with the source and drain regions. The backend comprises of inter-level dielectric (ILD) deposition, via opening and pad formation.

We have fabricated devices with W_f=5 to 25 nm, L_g=30 nm to 2 μm and number of fins N_f=1 to 200. Consistent with the

double-gate nature of our devices, all figures of merit are normalized by twice the channel height.

III. Electrical Characteristics

The electrical characteristics of undoped and δ-doped fin array devices (N_f=34) with W_f=5 nm (AR= H_c/W_f=10) and L_g= 50 nm are shown in **Fig. 5**. Well-behaved characteristics and good sidewall control with S_{lin}=65 mV/dec in the undoped channel are obtained. **Fig. 6** shows electrical characteristics of W_f=25 nm, L_g=50 nm devices with single (top) and multiple fins (bottom). For both doped and undoped fins, OFF-state performance of array and single fin transistors is similar. This suggests interface-state dominated behavior. In contrast, g_m of single-fin devices is larger than in arrays, as expected from R_{sd} considerations.

Fig. 7 summarizes g_m and S_{sat} (both at V_{DS}=0.5 V) of undoped- and doped-fin array devices of different W_f and L_g. Overall, undoped fins exhibit much better OFF-state behavior while doped fins show better ON performance over most of the range (this trend is kept even when removing the effect of R_{sd}). For short L_g and narrow fins, the ON performance of undoped fins exceeds that of doped fins due to better scalability.

V_T rolloff (V_{DS}=50 mV) for different W_f is shown in **Fig. 8**. Excessive rolloff is observed for very narrow fins, a possible consequence of line-edge roughness. Except for the W_f=5 nm samples, the V_T distribution in undoped fin transistors as W_f changes is much tighter than in doped fin devices, a key advantage with relevance to future manufacturing. R_{on} at a fixed $V_{GT}=V_{GS}-V_T$=0.6 V is shown in **Fig. 9**. R_{sd} and R_{fin} defined in **Fig. 9a-b** are shown in **Fig. 9c-d**, respectively. R_{sd} of the doped sample is lower than in the undoped one suggesting more effective fin to S/D link. The fin resistance R_{fin} is similar in both structures at the given overdrive, and scales roughly as $1/W_f$.

IV. Mobility analysis

In order to study the scaling behavior of our FinFETs in the ON regime, we have carried out an extraction of the mobility. In an effort to mitigate the effect of interface states, we performed simultaneous measurements of I_D-V_{GS} characteristics and S-parameters at 1 GHz for V_{DS}=10 mV for different L_g on relatively long-channel devices. From these measurements, we extracted the total gate capacitance, $C_g=C_{gs}+C_{gd}$, as a function of L_g. Results obtained for δ-doped- and undoped-fin devices with W_f=9 nm are shown in **Fig. 10**. From the slope of C_g-L_g (see inset) we extract the intrinsic fin capacitance per unit length, C_{fin} (**Fig 11**). The combination of I_d and C_{fin} yields the mobility (**Fig 12)** which is corrected for series resistance. Results of this analysis are summarized in **Fig. 13** at two different values of electron concentration in the channel. First, there is a strong fin width dependence for W_f<20 nm, presumably due to sidewall roughness scattering. For a linear carrier concentration in the fin $n_l=10^7$ cm^{-1}, δ-doped fins show somehow higher mobility at all W_f. For $n_l=3\times10^7$ cm^{-1} equivalent results are obtained in both structures.

V. 2D Simulations

To explore the impact of doping on device characteristics, we performed 2D P-S simulations in the fin cross-section. The charge distribution in OFF and ON state are shown in **Fig. 14**. In the OFF state, the charge distribution is sharply concentrated against the bottom interface for doped fins but towards the top of the fin in undoped fins. The shielding effect of the δ-doped layer in the doped fins weakens charge control from the gate and hurts the subthreshold behavior. In the ON state, the charge distribution in the doped fins remains highly crowded against the bottom interface where interface roughness scattering should be at a minimum. In the undoped fins, the electrons pile up along the sidewalls with a higher concentration towards the top where interface roughness scattering should be worse. This might explain our mobility and transconductance results. For W_f=9 nm, there is complete volume inversion in the undoped sample mitigating sidewall roughness scattering.

C_{fin} as obtained from simulations is shown on **Fig. 11**. The simulated capacitance is in reasonable agreement with the measured data. Undoped and δ-doped fins have similar capacitance, however, the δ-doped structure is stretched out around threshold and, as a result, S degrades. This degradation is partially mitigated with W_f scaling.

To further explore the reason for the significantly lower mobility in undoped wide fins at low n_l, we evaluated the weighted average $\langle n_s E_s \rangle_y$ (**Fig. 14**) at the fin sidewall surface. E_s represents the normal electric field at the fin sidewall surface. The results are shown in **Fig. 15** as a function of n_l. These data were also used to plot μ vs. $\langle n_s E_s \rangle$ in **Fig. 12**. In wide fins and for the δ-doped structure, there is significant reduction in the electric field at low n_l with respect to the undoped fin. This is consistent with a higher mobility. For larger n_l, the difference in $\langle n_s E_s \rangle$ disappears and the mobility behavior is strongly affected by sidewall roughness scattering. In narrow fins, the impact of doping is small throughout the full n_l range.

VI. Benchmark

Fig. 16 benchmarks g_m, normalized by conducting gate periphery, and g_m/W_f for InGaAs FinFETs from this work and the literature (V_{DD}=0.5 V) as well as state-of-art Si FinFETs (V_{DD}=0.8 V) as a function of fin width. g_m/W_f is an FOM that highlights the desire to obtain high current on a minimum transistor footprint. When viewed this way, this work shows record performance that approaches Si in spite of the rather different operating voltage. In addition, this a first-time demonstration of working transistors at the target W_f of 5 nm.

VII. Conclusions

We demonstrate InGaAs FinFETs with W_f=5 nm and gate-contact separation of 5 nm through an advanced self-aligned process. Undoped-fin devices exhibit better OFF-state characteristics and tighter V_T than doped fin devices. In the most scaled devices, g_m/W_f approaches state-of-art S_i FinFETs.

Acknowledgment

Device fabrication was carried out at the Microsystems Technology Laboratories and the EBL Facility at MIT. This research was funded by DTRA (HDTRA1-14-1-0057), Lam Research, NSF (E3S STC grant #0939514), and Korea Institute of Science and Technology. The authors thank Dr. Winston Chern for many useful discussions.

REFERENCES: [1] del Alamo, IEDM2013, [2] Riel, MRS Bul. 39(8), 668 (2014), [3] Radosavljevic, IEDM2011, [3] Oxland, EDL 37(3), 261 (2016), [4] Djara, VLSI2015, [5] Zota, IEDM2016, [6] Sun, VLSI2017, [7] Waldron, VLSI2014 [8] Lin, EDL, 37(4), 381 (2016), [9] Yakimets, TED, 62(5), 1439 (2015), [10] del Alamo, EDS 4(5), 205 (2016), [11] Lin, IEDM2012, [12] Vardi, EDL, 36(2), 126 (2014), Lin, IEDM2013, [13] Lin, IEDM2014, [14] Lin, EDL, 35(4), 440 (2014), [15] Vardi, EDL 37(9), 1104 (2016), [16] Zhao, EDL 35(5), 521 (2014)

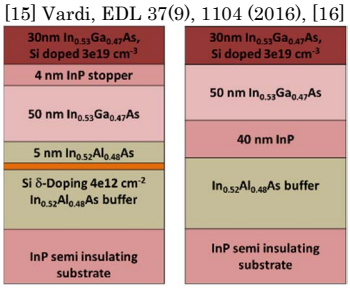

Figure 1: Starting heterostructures.

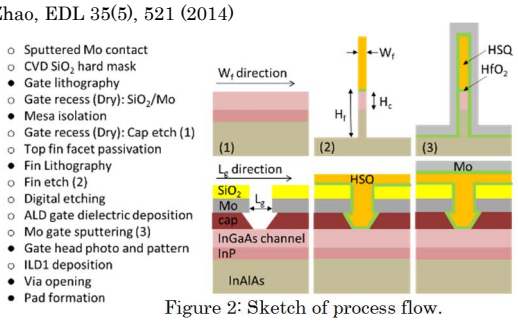

Figure 2: Sketch of process flow.

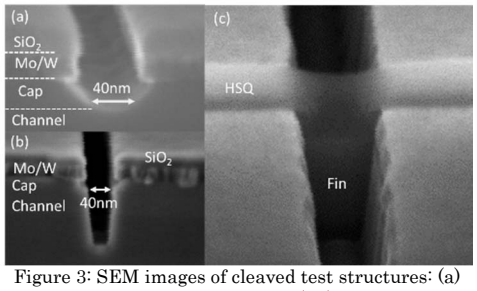

Figure 3: SEM images of cleaved test structures: (a) after dry recess and digital etch, (b,c) after fin etch.

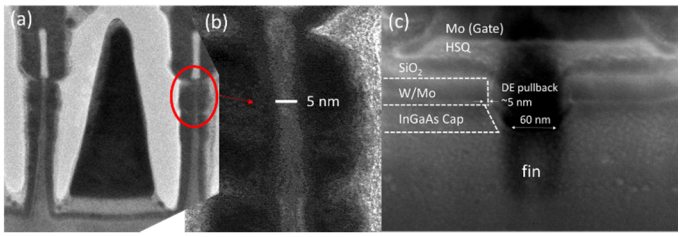

Fig. 4: (a and b) Cross-section TEM images of finished device with W_f=5 nm fin. (c) FIB image along the fin of finished device with L_g=60 nm.

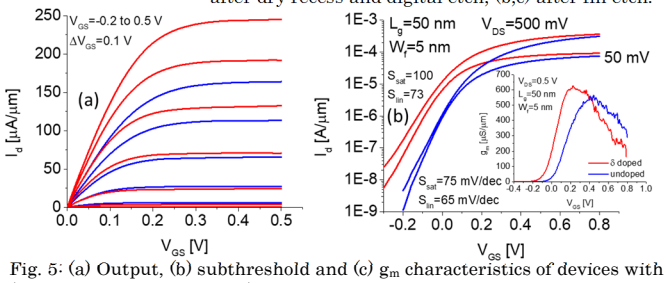

Fig. 5: (a) Output, (b) subthreshold and (c) g_m characteristics of devices with (W_f=5 nm, L_g=50 nm, N_f=34). Blue/red traces for undoped/δ-doped channels.

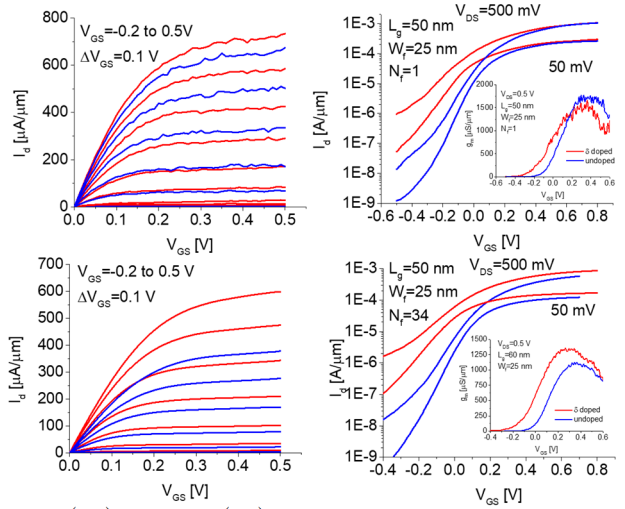

Fig. 6: (a,a') Output and (b,b') subthreshold characteristics of devices with W_f=25 nm, L_g=50 nm and N_f =1 (top) and 34 (bottom). Insets show saturated g_m characteristics. Blue/red traces for undoped/δ-doped channels.

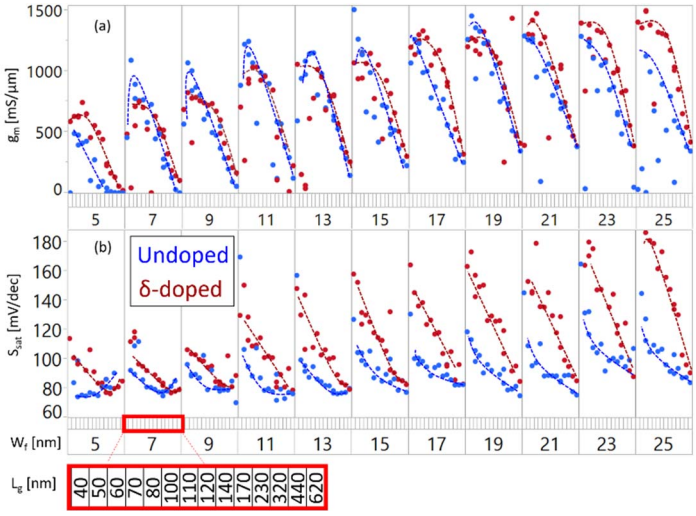

Figure 7: Comparison between g_m and S_{sat} (both at V_{DS}=0.5 V) of undoped (blue) and δ-doped (red) structures at different W_f and L_g.

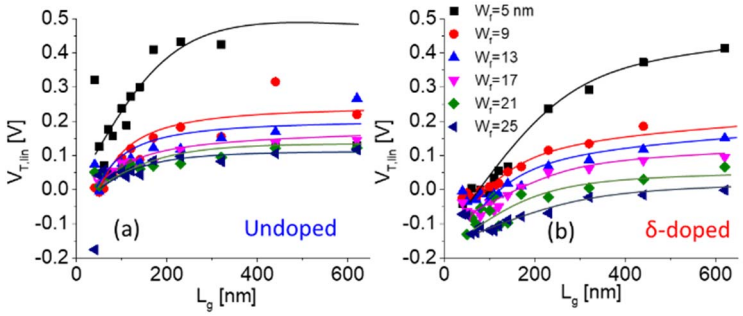

Figure 8: V_T (from max g_m extrapolation at V_{DS}=50 mV) as a function of L_g for different W_f in undoped (a) and δ-doped (b) samples.

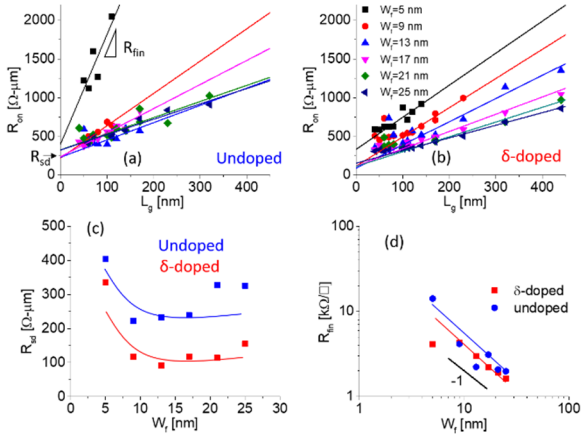

Figure 9: R_{on} (at V_{GT}=0.6 V) vs. L_g for different W_f in undoped (a) and δ-doped (b) samples. (c) R_{sd} and (d) R_{fin} obtained from linear fits of the data in (a-b); blue: undoped fins, red: doped fins.

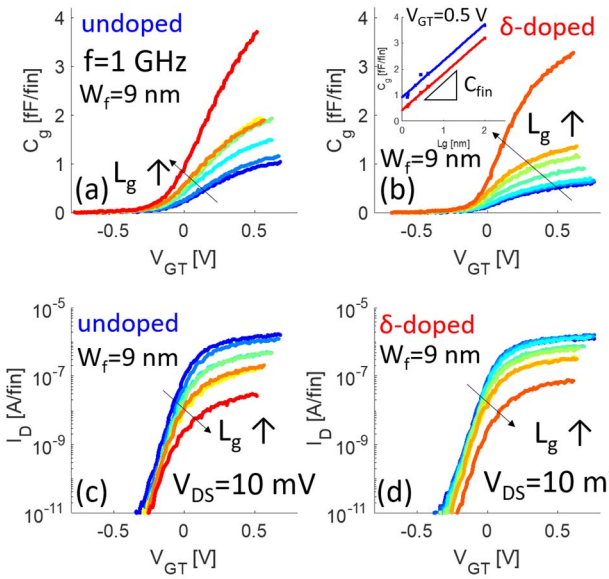

Figure 10: (a,b) Capacitance extracted from S parameters measurements at 1 GHz at V_{DS}=10 mV for (undoped,δ-doped) fins with W_f=9 nm. (c,d) I_D-V_{GT} curves collected during S parameter measurements. (Inset) C_g vs. L_g at V_{GT}=0.5 V, used for C_{fin} extraction shown in

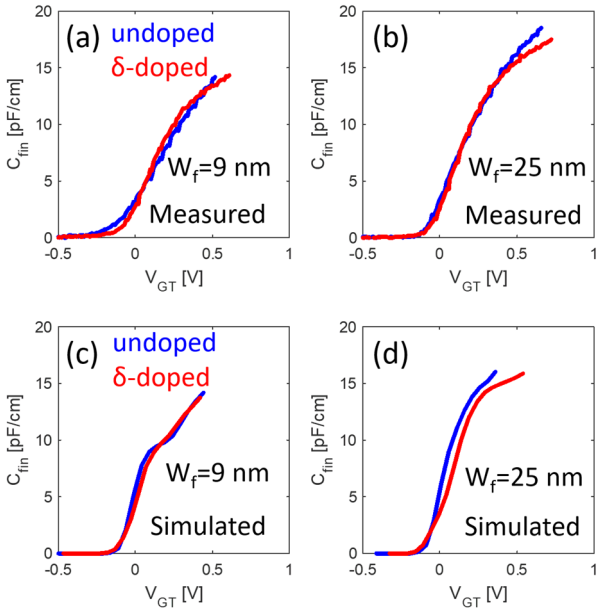

Figure 11: (a,b) Intrinsic fin capacitance extracted from S parameters measurements. (c,d) Fin capacitance evaluated from P-S simulations.

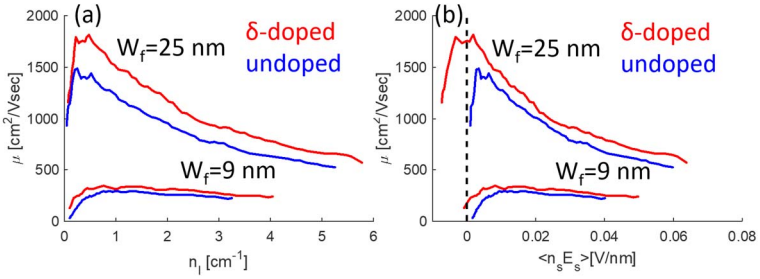

Figure 12: Long channel mobility for undoped and δ-doped samples at W_f=9 and 25 nm, as a function of n_i (b) and average surface field at the fin sidewalls (b)

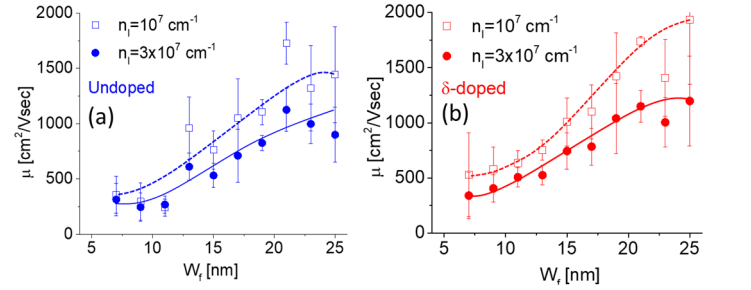

Figure 13: Long channel mobility for undoped (a) and doped (b) samples at two different carrier concentrations

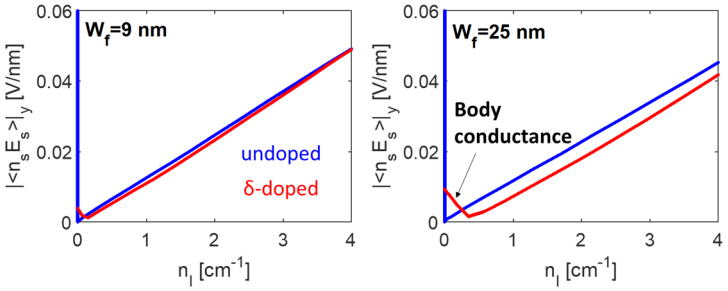

Figure 15: Weighted average of the surface field along the sidewalls as a function of n_i for different levels of δ-doping.

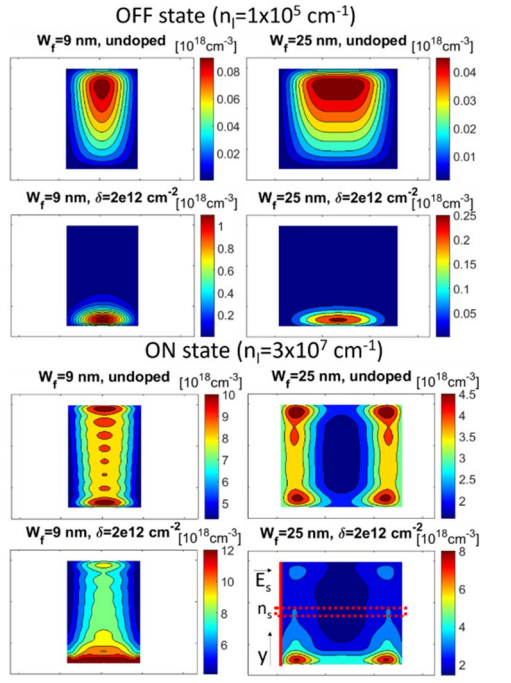

Figure 14: Charge distribution in W_f=9 and 25 nm doped and undoped fins in the OFF state (n_I=1x10⁵ cm⁻¹) and ON state (n_I=3x10⁷ cm⁻¹).

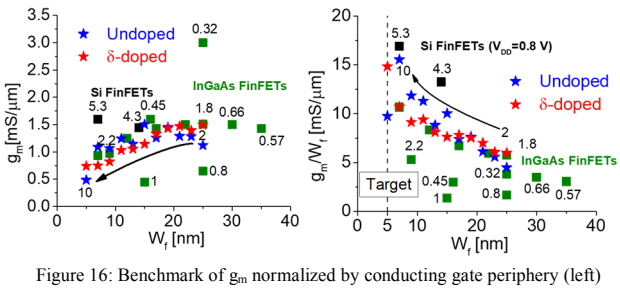

Figure 16: Benchmark of g_m normalized by conducting gate periphery (left) and g_m/W_f (right) as a function of W_f, from this work and the literature (V_{DD}=0.5 V). For reference, state-of-the-art Si FinFETs are also shown (V_{DD}=0.8 V). Numbers represent the fin width to channel height aspect ratio.