Fin-Width Scaling of Highly Doped InGaAs Fins

Xin Zhao**, Member, IEEE, Alon Vardi**, Member, IEEE, and Jesús A. del Alamo**, Fellow, IEEE

Abstract—We study the impact of fin-width scaling on transport in highly doped InGaAs fins and the effect of digital etch (DE). Our experiments suggest the existence of a 10-nm-wide “deadzone” on each side of the fin that does not contribute to the transport. The extent of the deadzone cannot be mitigated by DE nor sidewall passivation. Simulations suggest that the Fermi-level pinning and its associated subsurface depletion region alone cannot explain the relatively wide deadzone that is measured. We propose an explanation based on the combination of Fermi-level pinning and mobility degradation as the fin width scales down. This leads to an apparent wider deadzone than accounted by the Fermi-level pinning alone.

Index Terms—III–V, Fermi-level pinning, FinFETs.

I. INTRODUCTION

InGaAs is a promising channel material candidate for CMOS technologies beyond the 10-nm node [1]–[10]. In this dimensional range, only high aspect ratio (AR) 3-D transistors with a fin or nanowire configuration can meet the scaling goals. Recently, InGaAs FinFETs with sub-10-nm fin width have been demonstrated [11]. At the point of technology deployment, fin widths in a few nanometers range are required [12]. Therefore, it is critical to understand the width scaling properties of InGaAs fins.

Fin-width scaling studies of heavily doped InGaAs fins revealed what appeared to be a nonconducting “deadzone” located on the fin sidewalls [13]. This was shown to affect both the fin resistance and the contact resistance in a similar way. The width of the deadzone was found to depend on the process conditions and spanned from 10 to 30 nm. In a future scaled InGaAs FinFET, the access and contact regions will be heavily doped. Although the width of these regions will be wider than the channel to minimize series resistance, the scaling behavior studied here would have very serious negative implications to device performance given the overall footprint constraint.

In an effort to understand the origin of this apparent deadzone so that it can be removed, we investigate here the impact of digital etch (DE) on the fin-width scaling properties of reactive ion etching (RIE)-etched heavily doped InGaAs fins. DE is a highly controlled chemical etching process, in which the two elemental steps of chemical etching, oxidation and oxide removal, are carried out separately and are, therefore, self-limiting [14]. DE has shown to reduce the surface roughness and RIE etching damage [15]. It can also be used to accurately trim nanostructures to a few nanometers regime while preserving vertical walls and sharp corners [16]. This makes this technology of great importance in the manufacturing of future nanoscale InGaAs FinFETs.

In this work, counterintuitively, we have found that DE has a minimum impact on the extent of the deadzone. On the contrary, we find that fin-width scaling degrades fin conductivity through a combination of Fermi-level pinning and mobility degradation.

II. FABRICATION PROCESS

We have fabricated transmission-line model (TLM) test structures to measure the resistance of contacted InGaAs fins in parallel [Fig. 1(a) and (b)]. The sample consists, from top to bottom, of 10-nm n* -In0.7Ga0.3As, 30-nm n* -In0.53Ga0.47As, 4 nm of InP, and 300 nm of In0.52Al0.48As on an InP substrate. Both InGaAs layers are Si-doped to 3 × 10¹⁹ cm⁻³ (nominal).
The fabrication flow follows a fin-first process designed to obtain high-quality fin sidewalls and minimizes the deadzone. This to some extent comes at the expense of contact resistance, which is the best in a contact-first approach [13].

The process begins with CVD of 2 nm of a Si3N4 adhesion layer followed by 100 nm of spun-on hydrogen silsesquioxane (HSQ). The fins are then patterned by electron–beam lithography and etched by an inductively coupled plasma (ICP) etcher utilizing a SiCl4/BCl3/Ar chemistry [15]. Chamber pressure was maintained at 0.2 Pa. We use the ICP power of 20 W and 280-V substrate bias voltage. After the fin hard-mask removal, two cycles of DE are performed to smooth the sidewalls. Our DE consists of sequential oxidation by O2 plasma followed by oxide removal in 1:1 H2SO4:H2O [14]. Each etch cycle removes 1 nm on each side while preserving the fin shape. Etch rates of both dry and DE are identical for InGaAs of different compositions.

Immediately after the second DE cycle, 30 nm of Mo is sputtered and Ti/Au (10/200 nm) pads are patterned, evaporated, and lifted off. Our sputtering system yields a ∼1:2 sidewall/top Mo thickness ratio. Flowable oxide (FOX), which is a negative-tone electron beam resist, is next patterned on the top of the fin contact and access area [Fig. 1(b)]. The Mo along the central portion of the fins is then etched using SF6 RIE [Fig. 1(c)]. Since the fin contact area is protected by FOX, at the end of the process, we obtain Mo contacts which wrap around the fins, as shown in Fig. 1(d).

At this point in the process, an initial device characterization is performed, as detailed below. Subsequently, three additional DE cycles were carried out. After each half cycle, the test structures were measured again. An exception to this is the first oxidation step after which no measurements were taken. The experiment was concluded by passivating the fins using 5 nm of Al2O3 deposited by atomic layer deposition (ALD) at 250 °C. After this, a final evaluation was performed.

Devices with initial fin width \( W_f \) = 40 to 90 nm and fin length \( L_f \) = 0.2 to 1.5 μm were fabricated with fin contact lengths (Fig. 1) between 0.5 and 1 μm. Relatively wide fins are used in this study because fins narrower than 20 nm open up due to the presence of the deadzone. In all cases, the contact length was larger than the contact transfer length (estimated to be ∼ 100 nm). All dimensions were measured by SEM. Each device consisted of ten normally identical fins in parallel. The term “initial” or “INIT” refers to the devices before the final three cycles of DE and ALD deposition were performed. Fin pitch was 300 nm for all devices.

### III. TEST STRUCTURE CHARACTERIZATION AND RESULTS

TLM measurements in a Kelvin configuration were performed on a set of six devices with identical dimensions except for different \( L_f \). The resulting classic TLM plots, when appropriately normalized by total fin width, yield the sheet resistance, \( R_{sh} \), from the slope of the line and twice the contact resistance, \( 2R_c \), from the extrapolation of the measured resistance to \( L_f = 0 \). As in [13], linearity in TLM plots enables accurate sheet resistance and contact resistance extraction.

The results of a typical data set are shown in Fig. 2. In each step, the measured sheet resistance is normalized to our best estimation of the physical fin width, \( W_f \). From detailed calibrations, we assume that after the oxidation steps (O1), the fin width does not change, while after wet oxide removal (S1), it narrows by 2 nm.

Fig. 2(a) shows that in all instances, \( R_{sh} \) increases as \( W_f \) scales down. This is a manifestation of improper fin width scaling that suggests that a portion of the fin does not contribute to conduction. For all fins, \( R_{sh} \) tends to increase with the number of DE cycles. Fig. 2(b) shows \( R_{sh} \) normalized to its initial value, \( R_{sh,0} \). Within the same DE cycle, \( R_{sh} \) is lower after the DE oxide growth step suggesting that the oxide passivates the fin surface to some extent, which is also the case after the final ALD step.

To further evaluate the scaling characteristics of the fin resistance, we plot the sheet conductance of the fin, \( G_{sh} = 1/R_{sh} \) as a function of \( 1/W_f \) [13]. The results for a typical set of devices from one chip are summarized in Fig. 3(a). For clarity, only S1 steps are shown. In all cases, a straight line is obtained that intersects the \( x \)-axis. This suggests that \( G_{sh} \) scales as \( W_{eff} \), with \( W_{eff} = W_f - 2x_d \), as if there is a deadzone of width \( x_d \) associated with each sidewall that does not contribute to conduction.

The deadzone width \( x_d \) is obtained from the extrapolation of the straight line to \( G_{sh} = 0 \). The same plot gives the bulk sheet resistance \( (R_{sh,b}) \) from the extrapolation of the straight line to \( 1/W_f = 0 \). Fig. 3(a) also shows the value of \( G_{sh} \) obtained from Van der Pauw (VdP) measurements on a separate test structure with surface treatment identical to the fin structures. As the number of DE cycles increases, \( G_{sh,VdP} \) decreases. This is, in principle, expected as the thickness of the conductive layer (\( H_c \)) is reduced throughout the experiment. However, \( G_{sh,VdP} \) does not perfectly match the extrapolation of the straight lines fitted to the experimental data. This will be discussed in Section IV.
Fig. 3. (a) Fin-width scaling analysis of experimental fin sheet conductance at different points in the DE sequence, and the 2-D Poisson–Schrödinger simulations with constant mobility for the case of INIT. VdP refers to the value of $R_{sh}$, obtained from VdP measurements on a separate test structure. (b) Evolution of deadzone width, $x_d$, as a function of DE subcycles and final ALD passivation. The red symbols represent the results of experiments, and the blue band represents the depletion region thickness $t_d$ extracted from 2-D simulations.

The extracted values of $x_d$ averaged over many sets of structures obtained from four nominally identical chips are summarized in Fig. 3(b) (red squares). We find that the deadzone width has a value of $\sim 10$ nm that is rather unaffected by the number of DE cycles. This is unexpected. A detailed analysis of these results follows in Section IV.

IV. DISCUSSION

Vardi et al. [13] observed that adjusting the fin-etch parameters had a major effect on the deadzone width (between 10 and 30 nm). We then concluded that etching damage is, to some extent, responsible for the deadzone. With a capability of removing fin sidewall material of 1 nm at a time, DE should constitute an efficient way to reduce etching damage [15]. We therefore expect a deadzone width that should shrink after DE. Our experimental observations are inconsistent with this hypothesis.

An alternative explanation is that the deadzone is a manifestation of the existence of a subsurface depletion layer due to the Fermi-level pinning. We can use the self-consistent Poisson–Schrödinger simulations to estimate the magnitude of this effect. Fig. 4(a) shows the 2-D distribution of electron concentration across a fin with $W_f = 36$ nm before and after four DE cycles. The doping level in the fin was set to match the VdP measurements: $n_B = 9.7 \times 10^{13}$ cm$^{-2}$. Surface Fermi-level pinning of 0.2 eV is assumed for In$_{0.53}$Ga$_{0.47}$As and 0 eV for In$_{0.7}$Ga$_{0.3}$As [17]. A depletion region is clearly visible below all surfaces.

To estimate the extent of this depletion region, $t_d$, we first evaluate the electron concentration in the fin cross section (unit of cm$^{-1}$)

$$n_l = \int \int n dS. \tag{1}$$

Then, we compute the average sheet electron concentration in the fin (unit of cm$^{-2}$), defined as

$$n_s = \frac{n_l}{W_f}. \tag{2}$$

We follow with a calculation of the fin sheet conductance through:

$$G_{sh} = q \mu_j n_s \tag{3}$$

where we have assumed that the electron mobility is equal to the VdP measured value ($\mu_j = 1050$ cm$^2$/V·s). This forces the
initial $R_{sh,B}$ to match the VdP measurements. Nevertheless, the choice of mobility does not affect the effective depletion layer width.

We apply this procedure to fins with different values of $W_f$ and $H_e$ that match our best estimate of the actual fin dimensions in our experiment. The depletion layer thickness at the sidewalls ($t_d$) can be estimated by graphing $G_{sh}$ versus $1/W_f$, as shown in Fig. 3(a). A classic plot with a deadzone-like behavior is obtained in all cases. Fig. 3(a) shows one such plot labeled SIM:INIT (blue dots) corresponding to the initial fins right before any DE cycles. The simulated data deviate from the straight line at small $W_f$, as the fin width approaches twice the deadzone width. From the straight portion of this plot, the value of the depletion region thickness at the sidewalls is extracted. We follow this procedure for all fin widths. These values are plotted in Fig. 3(b) in the blue band assuming a surface Fermi-level pinning position for In$_{0.53}$Ga$_{0.47}$As in the range of 0.1–0.3 eV [17]. In all cases, the simulated $t_d$ values are significantly lower than the experimental $x_d$ results.

Fig. 4(b) shows the effective depletion layer width versus pinning energy and doping level. The black line shows the doping concentrations and Fermi-level pinning needed to match the carrier concentration extracted from VdP sheet resistance and Hall mobility. This analysis leads to an estimated actual doping level of around $3.5 \times 10^{13}$ cm$^{-2}$. This is slightly higher than the nominal value though within typical bounds of molecular beam epitaxy growth calibrations. We can see that $t_d$ is rather insensitive to both doping-level and Fermi-level pinning energy. A wide range of doping level from 2.5 to $4.5 \times 10^{19}$ cm$^{-3}$ with the Fermi-level pinning of 0.2 eV results in an estimation of $t_d$ between 4.6 and 5.8 nm. No realistic value of the Fermi-level pinning can correct the large discrepancy observed with our experiments. This shows that the Fermi-level pinning alone cannot account for our observations.

The discrepancy between Fermi-level pinning induced deadzone and the experimental results can be resolved if we question the assumption of constant mobility in the fins. It is well known that in thin highly doped layers, the mobility depends on the sheet carrier concentration $n_s$, with the mobility dropping as the sheet carrier concentration is reduced [18]. The $n_s$ dependence of the mobility arises as a result of screening of ionized dopants by the background electrons and it approximately follows a power law [19]:

$$\mu_f \propto n_s^{-5}. \quad (4)$$

In a heavily doped fin, this means that the reduction in the integrated carrier concentration that accompanies the formation of a depletion layer under the sidewalls should result in a drop in overall mobility, which would increase the fin resistance beyond the reduction of its cross section. This effect would be more prominent in thinner fins since the fraction of material that is depleted is larger. In the presence of a carrier concentration dependence of the mobility, and using (4), the sheet conductance can be expressed as

$$G_{sh} = q\mu_f n_s \propto n_s^{2.5} \propto \left(1 - \frac{2x_d}{W_f}\right)^{2.5}. \quad (5)$$

This new fin-width dependence agrees well with our experiments. In Fig. 5(a), we fit the experimental data in Fig. 3(a) with (5) and $x_d$ as a parameter. The fits are quite good. Fig. 5(b) shows the fitted values of $x_d$. The new model in Fig. 5(a) accurately captures the VdP data as well, which the depletion-based linear model in Fig. 3(a) fails to predict. There is now an excellent agreement between the values of $x_d$ that are inferred from the fitting process and the values of $t_d$ extracted from the F-S simulations, as shown in Fig. 5(b).

The relative insensitivity of the deadzone to DE and passivation that we find in this work is consistent with our mobility degradation/Fermi-level pinning hypothesis. This is because impurity scattering is the main scattering mechanism that limits the mobility and there is little sensitivity of the electrical fin width to the Fermi-level pinning position. This does not preclude a larger deadzone as a result of sidewall damage from RIE, as shown in [13]. In fact, our simulations here suggest that for the parameters selected here, if the deadzone is larger than 10 nm, RIE damage might be responsible and DE should be effective in reducing it down to this value.

Alternative explanations for our observations can be ruled out. We do not anticipate significant dopant loss as the
temperature of processing of the fin etch, Mo strip, etc., is relatively low (<250 °C). Dopant passivation might happen to some extent due to fin damage during these processing steps. However, neither of these two mechanisms can be significant since they are both incompatible with our experimental observation of an insensitive deadzone width in sequential DE experiments.

Aside from the Fermi-level pinning and mobility degradation, donor deactivation as a result of dielectric mismatch between the InGaAs fin and the gate stack has been postulated as potentially contributing to the apparent deadzone [20]. To test this hypothesis, we characterized the temperature dependence of the sheet resistance of fin structures after ALD passivation, as shown in Fig. 6. Contrary to [20], $R_{sb}$ is nearly temperature-independent. In fact, it increases slightly, rather than decreases, with rising temperature for all fin widths, suggesting that donor deactivation does not play a role. In addition, the doping level in our fins is well above the expected value for the Mott transition concentration for InGaAs [21], [22].

This work focuses on the impact of deadzone on sheet resistance, rather than contact resistance, $R_c$. The design of the experiment was not optimized to allow a detailed study of the deadzone effect on the contact resistance. In our process, DE was performed after contact metallization and therefore the fin width under the contact does not change with sequential DE. Nevertheless, our experiments reveal a deadzone behavior of the contact resistance with respect to $W_{fo}$. Fig. 7 shows $1/R_c$ as a function of $1/W_{fo}$. A similar deadzone width of 10 nm is extracted following the same procedure as with the sheet resistance. This is also consistent with earlier results obtained in [13]. We are currently carrying out a systematic study to investigate the impact of process conditions on the deadzone under the contact.

V. Conclusion

We study the effect of DE on the conductivity of highly doped InGaAs fins. Our experimental results show that an effective deadzone of 10 nm in width cannot be further reduced by DE. The discrepancy between the experimentally observed deadzone and the effective depletion layer expected from the Fermi-level pinning is explained through a reduction in mobility as the carrier concentration in the fins drop. Further study is needed to characterize the mobility dependence on fin width. However, it appears that both transport and electrostatics play an important role.

Acknowledgment

Device fabrication was carried out at the Microsystems Technology Laboratories and the Electron Beam Lithography Facility, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA.

References


