A Si-Compatible Fabrication Process for Scaled Self-Aligned InGaAs FinFETs

A. Vardi. J. Lin, W. Lu, X. Zhao, and J. A. del Alamo

Microsystems Technology Laboratories, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.
e-mail: alonva@mit.edu, Phone: +1 617-225-8605

Keywords: FinFET, III-V CMOS

Abstract

We have developed a scalable gate-last process to fabricate self-aligned InGaAs FinFETs that relies on extensive use of dry etch. The fabrication sequence yields high aspect ratio FinFETs with sub-10 nm fin widths and down to 20 nm gate lengths. The process involves F-based dry etching of refractory metal ohmic contacts that are formed early in the process. The fins are etched using a novel ICP process using $BCl_3/SiCl_4/Ar$. High aspect ratio fins with smooth sidewalls are obtained. To further improve the quality of the sidewalls and shrink the fin width, digital etch is used. Using this process flow, we have demonstrated FinFETs with L_g=20 nm and fin width as small as 7 nm with high yield. Good electrostatic characteristics are obtained in a wide range of device dimensions. In devices with 7 nm fin width, record channel aspect ratio and transconductance per unit footprint are obtained.

INTRODUCTION

InGaAs is a promising channel material candidate for CMOS technologies beyond the 7 nm node [1-3]. In this dimensional range, only high aspect-ratio (AR) 3D transistors with a fin or nanowire configuration can deliver the necessary performance. Impressive fin and nanowire based InGaAs FinFET prototypes have recently been demonstrated [4-6]. However, in most demonstrations to date, the fin width in III-V devices is still typically greater than 15 nm. At the point of insertion in a sub-7 nm node, InGaAs FinFETs with sub-10 nm fin widths and steep sidewalls will be required. This paper describes in detail a self-aligned recessed-gate process for scaled InGaAs FinFETs that emphasizes scalability, performance and manufacturability by making use of dry fin etching, digital etch and Si-compatible materials.

FABRICATION PROCESS

A prototypical starting heterostructure used in our work is sketched in Fig. 1. On an InP substrate, an InAlAs buffer layer is first grown that incorporates a Si δ-doping layer with a sheet concentration of $4x10^{12}$ cm^{-2} placed 5 nm below the channel. The channel is InGaAs lattice matched to InP. The cap consists of heavily-doped Si:InGaAs and an undoped

InP etch stopper. Our heterostructures are MBE grown by IntelliEpi.

Our device fabrication process integrates a number of features developed in our group over the last few years [7-10] and summarized in Fig. 2. The process starts with sputtering of a low-ρ (R_{sh}=5 Ω/□) W/Mo ohmic contact bilayer on the as-grown epitaxial structure. This contact-first approach yields outstanding contact resistance in planar devices [8, 10]. The metal stack is then covered by CVD SiO_2 which is used as a hard mask for gate recess and remains on the final device as a vertical spacer.
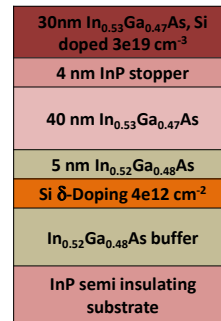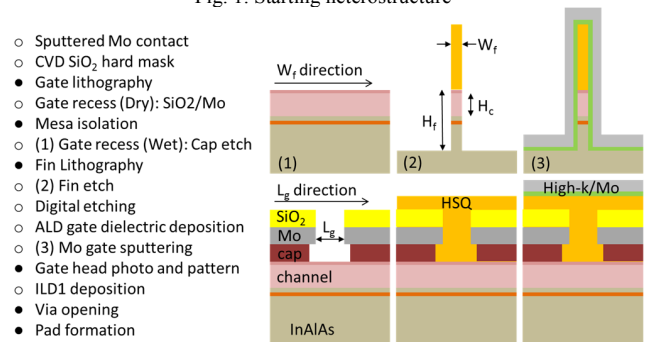

Fig. 1. Starting heterostructure

Fig. 2. (left) Process flow and (right) device schematics at the points in the process marked by numbers

After E-beam gate patterning, 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 (Fig. 3) [8]. F-based RIE stops at the III-V surface. After a mesa definition step, the highly conductive cap is removed using a citric-acid based wet etchant. This process is selective to InGaAs and stops on the undoped InP stopper. The isotropic wet etch pulls back the InGaAs cap about 20 nm, as shown in Fig. 3.

With the cap removed, a thin layer (2-3 nm) of Si_3N_4 is deposited by CVD as an adhesion layer for the subsequent fin etch mask.

Fins are e-beam patterned using 40 nm thick Hydrogen Silsesquioxane (HSQ). The overall result is a composite hard mask made out of CVD SiO_2 and HSQ that defines the fins and the S/D area. The fins are then etched in inductive coupled plasma (ICP) using a $BCl_3/SiCl_4/Ar$ chemistry [11]. This yields fins as narrow as 15 nm with an aspect ratio of ~10 (Fig. 4). The fins are highly vertical in the top ~70 nm. To further thin down the fins and smooth the sidewalls, we perform several cycles of digital etch (DE) [12] using dry oxidation and H_2O:H_2SO_4 oxide removal (etch rate ~1 nm/cycle for one side, ~2 nm/cycle for the fin width) (Fig. 5).

After the last DE cycle, a fresh semiconductor surface is exposed in a final H_2SO_4 cleaning step. This is immediately followed by ALD of 2.2 nm of HfO_2 as gate dielectric and Mo gate metal sputtering (Fig. 6). The device is completed by gate metal definition, using SF_6:O_2 dry etch. Note that the gate covers completely the fins and overlaps with the source and drain regions. To reduce leakage through the pads and substrate, another layer of SiO_2 is deposited. The process is completed after via opening and pad formation.

The entire front-end fabrication (before pads) closely follows Si CMOS-compatibility requirements and is completely lift-off free and Au-free. The overall process has a very low thermal budget with a maximum temperature of 300ºC from the SiO_2 CVD deposition. Our fabrication flow enables the formation of self-aligned FinFETs with gate lengths as short at 20 nm.
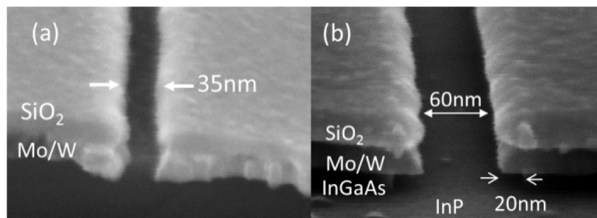

Fig. 3. Gate recess – (a) dry etch of SiO_2 and Mo in a device with L_g=35 nm, (b) wet etch of InGaAs in L_g=60 nm device. After wet etch the cap is pulled back ~20 nm.
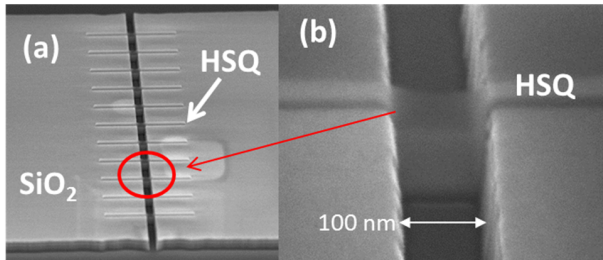

Fig. 4. (a) Array of fins etched in the device recessed area; (b) higher magnification of a single fin.

A key aspect of our process is that the HSQ that defines the fin etch is kept in place. This makes our FinFETs double-gate MOSFETs with carrier modulation only on the sidewalls. While theoretically inferior to Trigate designs, practically, the greater simplicity of the process allows us to aggressively scale all device dimensions and implement a robust self-aligned flow with a high yield. This ultimately results in significantly better performance than prior InGaAs Trigate MOSFET demonstrations, as discussed below. In addition, for high channel height to fin width aspect ratio (AR=H_c/W_f), a top gate yields diminishing returns [13].

Devices with fin widths ranging from 7 to 22 nm, gate lengths from 20 to 600 nm, with a fin pitch of 200 nm were fabricated with high yield. FIB cross-section of a finished device (L_g=20 nm) along the fin length direction and between the fins are shown in Fig. 6 (a) and (b) respectively. Due to the underlap of the cap, the separation between source and drain between the fins is ~10 nm wider than the length of the fins. FIB cross sections of devices with W_f=7 nm and 12 nm are show in Fig. 6 (c).
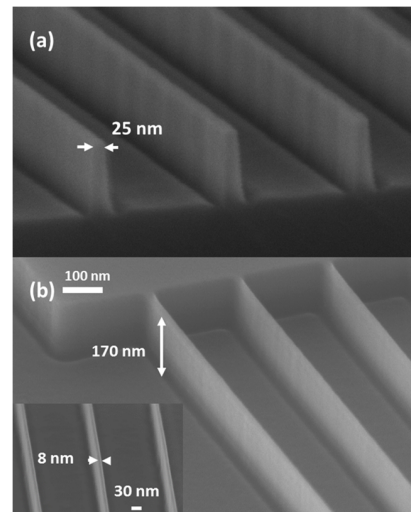

Fig. 5. Fin etch test structures: (a) cleaved fins as etched; (b) tilted view and top view (inset) of fins after 3 cycles of digital etch.
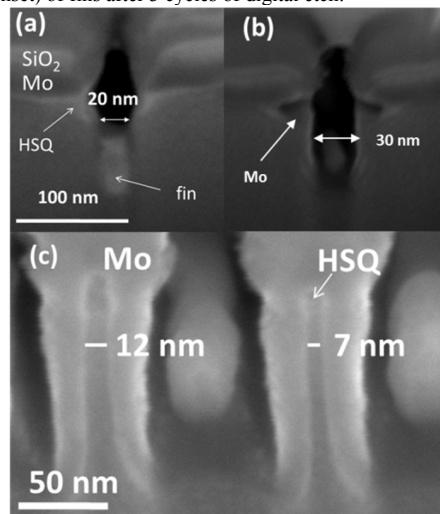

Fig. 6. FIB cross-section of finished device from source to drain: (a) along fin and (b) between fins; (c) Cross-section FIB of two completed fin test structures.

The electrical characteristics of a typical device with L_g= 30 nm and W_f=7 nm (AR= H_c/W_f=5.7), are shown in Fig. 7. Well-behaved characteristics and good sidewall control are demonstrated. The device R_{on} is 320 Ω·µm and a peak g_m of 900 µS/µm is obtained at V_{DS}=0.5 V (Fig. 8). Consistent with the double-gate nature of our devices and common practice, all figures of merit have been normalized by the conducting periphery which, in our case, is two times the channel height. The subthreshold characteristics of the same device (Fig. 7b) indicate a saturated subthreshold swing, S_{sat}, of 100 mV/dec and DIBL of 90 mV/V at 0.5 V. In devices with L_g=30 nm and W_f=22 nm, a peak g_m of 1500 µS/µm is obtained. For L_g=2 µm and W_f=22 nm, S_{lin} at V_{DS}=50 mV is as low as 68 mV/dec (Fig. 9), indicating a high quality interface between the semiconductor sidewall and the high–k gate oxide.
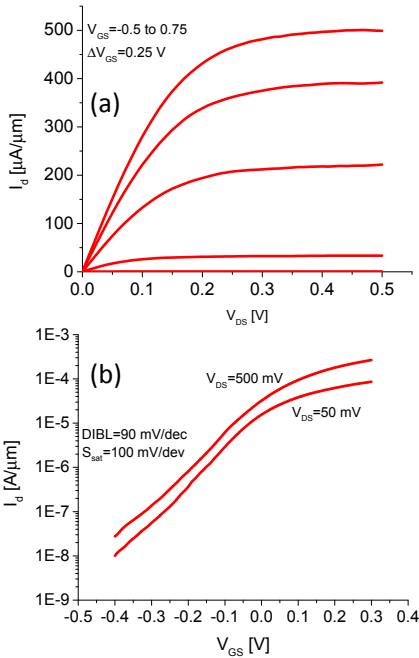

Fig. 7. Output (a) and subthreshold (b) characteristics of a FinFET with W_f=7 nm and L_g= 30 nm.
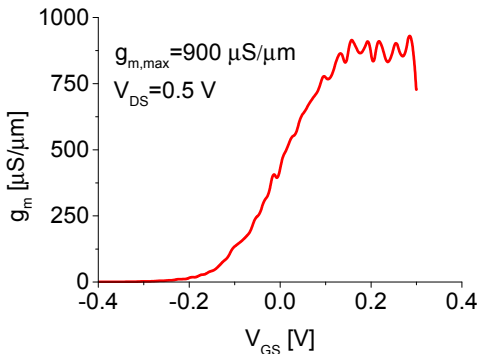

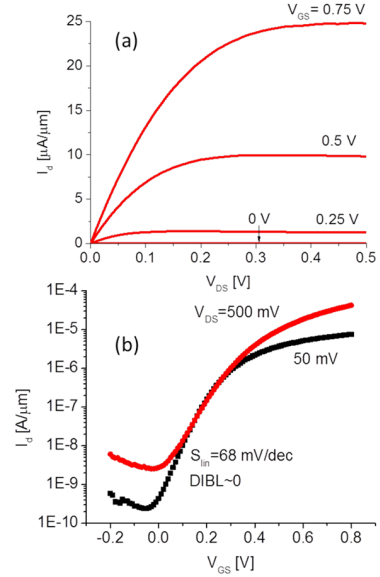Fig. 8. Saturated g_m characteristics of the FinFET of Fig. 7.

Fig. 9. Output (a) and subthreshold (b) characteristics of device with W_f=22 nm and L_g= 2 µm. A nearly ideal S_{lin}=68 mV/dec is obtained.

Fig. 10 shows the scaling of g_m and saturated S with L_g and W_f. While g_m increases as L_g scales down, a clear degradation of g_m is observed with W_f scaling. In addition, the saturated subthreshold swing shows a weak dependence on fin width. These are all manifestations of improper fin width scaling. This might be due to poor sidewall characteristics in the form of excessive interface state density or sidewall roughness, or line edge roughness from the E-beam lithography and RIE processes that are used to define the fins. Understanding and addressing these issues is crucial for the eventual success of this technology.
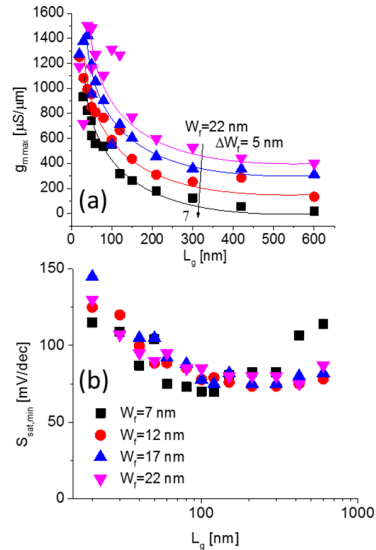

Fig. 10. Scaling of g_m and S_{sat} with L_g and W_f. Non-ideal fin width scaling is evident. The high subthreshold swing for long L_g transistors is due to excessive gate leakage current.

Fig. 11 benchmarks peak g_m of InGaAs FinFETs published to date [14-20] as a function of the fin width. The same data is graphed in two different ways: the conventional approach (Fig. 11a), which is to normalize g_m by the conducting gate periphery; and an alternative method, where the same data is normalized by the fin width (Fig. 11b). The latter approach is relevant, because it relates to transistor density, which, in the end, is what Moore's Law is all about. A winning FinFET should be capable of conducting a lot of current standing on a minimum footprint. Both graphs include estimations (green symbols) from selected state-of-the-art silicon FinFETs (22 nm and 14 nm CMOS from Intel [19,20]), along with the best of our recently published results [21,22]. Next to etch data point the channel aspect ratio (channel height over fin width) is indicated.

When normalized to gate periphery, the best Si and InGaAs FinFETs show equivalent performance. However, it is important to note that, with the exception of the devices presented in this work (blue stars), the fin width and aspect ratio of InGaAs FinFETs in the literature is far from that of Si FinFETs and from what is required for beyond 7-nm applications.
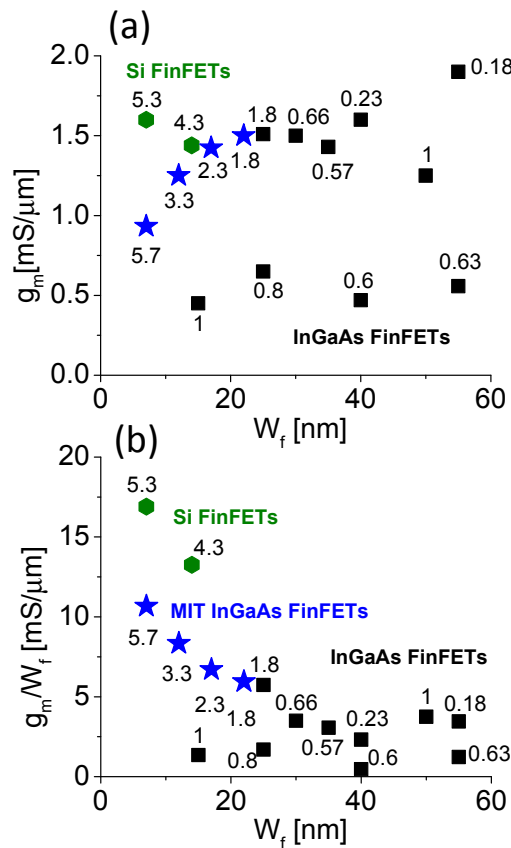
On the other hand, when normalizing to fin footprint, a large gap between Si and InGaAs transistors is revealed. Our devices [19-20] contribute to bridging this gap due to narrow fins and efficient use of sidewall conductivity. Our process potentially allows 2.5 more current per footprint than planar InGaAs MOSFETs. Nevertheless, more work is ahead before InGaAs FinFETs match and eventually exceed the performance of Si FinFETs.

CONCLUSIONS

We demonstrate self-aligned InGaAs FinFETs with extremely thin fins (down to 7 nm), high channel aspect ratios (as high as 5.7), short gate lengths (down to 20 nm) and excellent performance. When scaled by the fin footprint, our transistors improve the state of the art by nearly a factor of three, suggesting effective channel charge control from the sidewalls of very thin, high aspect-ratio fins.

ACKNOWLEDGMENTS

Fig. 11. Benchmark of maximum g_m vs. W_f for InGaAs FinFETs and state-of-the-art Si FinFETs. (a) g_m normalized by conducting gate periphery; (b) g_m normalized by fin footprint. The numbers next to each data point represent the aspect ratio of the conducting channel (height over width).

REFERENCES

[1] J. A. del Alamo, et al., IEDM Tech. Dig., 2013, p.24.
[2] J. A. del Alamo et a., J. Electron Dev Soc. vol. 5, no. 5, pp. 205-214, September 2016.
[3] Riel, H., et al., MRS Bulletin, Vol. 39, pp. 668-677, August 2014.
[4] R. Oxland et al, EDL., vol. 37, no. 3, pp. 261–264, Mar. 2016.
[5] V. Djara et al. VLSI Sym. Kyoto, 2015, pp. T176-T177.
[6] M. L. Huang et al, VLSI Sym. Honolulu, Jun. 2016, pp. 16-17.
[7] J. Lin, et al., APEX, vol. 5, p. 064002. 2012.
[8] J. Lin, et al., IEDM Tech. Dig., 2012, p.757.
[9] J. Lin, et al., IEDM Tech. Dig., 2013, p.421.
[10] W. Lu et al., EDL vol. 35, no. 2, pp. 178-180, February 2014.
[11] X. Zhao and J. A. del Alamo, EDL, vol. 35, no. 5, pp. 521-523, May 2014.
[12] J. Lin et al., EDL, vol. 35, no. 4, pp. 440-442, April 2014.
[13] S.-H. Kim et al., TED, vol. 52, no. 9, pp. 1993, September 2005.
[14] M. Radosavljevic et al., IEDM Tech. Dig. 2011, pp.33.1.1–33.1.4.
[15] T.-W. Kim et al., IEDM 2013, pp. 16.3.1–16.3.4,
[16] S. H. Kim et al., TED vol. 61, no. 5, pp. 1354-1360, May 2014.
[17] N. Waldron et al., VLSI Sym. 2014, pp. 1-2.
[18] A. V. Thathachary et al., VLSI Sym. 2015, pp. T208-T209.
[19] C. H. Jan et al., IEDM 2012, pp. 3.1.1–3.1.4.
[20] S. Natarajanet al., IEDM 2014, pp. 3.7.1–3.7.3.
[21] A. Vardi et al., VLSI Tech. Symp. 2016.
[22] A. Vardi and J. A. del Alamo, EDL, vol. 37, no. 9, pp. 1104-1107, Sept. 2016.