A Self-aligned Gate-last Process applied to All-III-V CMOS on Si

Jonsson, Adam; Svensson, Johannes; Wernersson, Lars-Erik

Published in:
IEEE Electron Device Letters

DOI:
10.1109/LED.2018.2837676

2018

Document Version:
Peer reviewed version (aka post-print)

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
A Self-aligned Gate-last Process applied to All-III-V CMOS on Si

Adam Jönsson, Johannes Svensson, and Lars-Erik Wernersson

Abstract—Vertical nanowire n-type (InAs) and p-type (GaSb) transistors are co-processed and co-integrated using a gate-last process, enabling short gate-lengths ($L_g=40$ nm) and allowing selective digital etching of the channel. Two different common gate-stacks, including various pre-treatments, were compared and evaluated. The process was optimized to achieve high n-type performance while demonstrating p-type operation. The best n-type device is scaled down to 12 nm diameter and has a peak transconductance of 2.6 mS/µm combined with a low $R_{on}$ of 317 Ω µm while the p-type exhibit 74 µS/µm. In spite of increased complexity due to co-integration, our n-type InAs transistors demonstrate increased drive-current, 1.8 mA/µm, compared to earlier publications.

Index Terms—Vertical, nanowire, III-V, MOSFET, CMOS, InAs, GaSb

I. INTRODUCTION

Complementary metal-oxide-semiconductor (CMOS) circuits based on all-III-V channel materials require further development to achieve competitive p-type performance [1], [2]. Although antimonide-based materials such as GaSb have demonstrated high hole mobility [3], transistor performance is, in part, limited by the gate-stacks [4]–[6]. It has been suggested that combinations of a more conventional p-type SiGe channel combined with n-type III-V InGaAs channel is a viable alternative to current CMOS technology [7], [8]. This type of material integration is, however, not straightforward, mainly due to strong material selectivity during processing.

With continued transistor scaling, deteriorated electrostatics leads to various short channel effects [9]. Vertical nanowires with a gate-all-around (GAA) geometry is one alternative with beneficial performance at the 5 nm node [10]. The vertical nanowire geometry in particular decouples the gate length and contact geometry from the footprint area [11].

In this paper, GaSb p-type and InAs n-type MOSFETs are co-integrated in a vertical nanowire gate-all-around structure, utilizing a common gate-stack. The process uses hydrogen silsequioxane (HSQ) spacers with adjustable thickness for development of a self-aligned, gate-last, process compatible with vertical antimonide-based structures.

All presented devices are scaled to sub-100 nm gate lengths and, for the n-type devices, nanowire diameters down to 12 nm are demonstrated. Improved drive-currents are also realized as compared to previous iterations of vertical InAs transistors, namely a 400% (gate-last) [12] and 150% (doped channel) [13] increase is shown with $I_{DS} = 1.8$ mA/µm ($V_d$ and $V_g = 0.7$ V).

![Fig. 1. Schematic illustration of crucial steps in the full process flow attached with an overview of fabrication. The illustrations show n-channel devices, notice that the process allows for varying gate placement along the nanowire. *Post metal anneal (PMA) is performed only on specific samples.](image)

II. DEVICE FABRICATION

Fig. 1 schematically illustrates the process flow for co-integration of p-type GaSb and n-type InAs MOSFETs. The devices are fabricated on 1 cm² p-type silicon (111) substrates with a 260 nm epitaxially grown InAs layer [14].

InAs-GaSb nanowires are subsequently grown from 15-nm-thick Au seed particles defined by electron beam lithography (EBL). Sections of the InAs segment is n-doped by Sn and the top of the GaSb segment is p-doped by Zn, see Fig. 1. In the case of GaSb growth, significant background doping is present attributed to point defects ($10^{16}$) [15]. By utilizing the Gibbs-Thomson effect during VLS growth, the GaSb growth rate can be significantly reduced for smaller gold particle sizes [16].

This work was supported in part by the Swedish Research Council, and in part by the Swedish Foundation for Strategic Research, and the European Union H2020 Program INSIGHT (Grant Agreement No. 688784).

All authors were with the Department of Electrical and Information Technology, Lund University, SE-221 00 Lund, Sweden (e-mail: adam.jons@cit.lth.se).
(diameter < 30 nm) which enables length control of the two separate materials in the heterojunction nanowires [17]. Therefore, 24 nm and 28 nm diameter Au dots are used for growing n-type wires and 44 nm for p-type wires. The Au dots are patterned in structures with a pitch of about 300 nm.

In the first step, the top contact is defined with an HSQ mask, whose thickness is determined by the dose in an EBL process [9]. The thicknesses of the HSQ is gradually varied from 60 to 300 nm to define gate position in separate devices. The InAs shell on the protruding part of the nanowires, that will later constitute the top contact, is digitally etched by oxidation and citric acid treatment. An ammonium sulfide treatment is also used for some of the samples, to passivate the III-V surface [18], processing differences are outlined in Table I. Subsequently, 20 nm tungsten (W) is sputtered followed by a 3 nm TiN atomic-layer-deposited (ALD) film. The metal is removed from all planar surfaces by an ICP-RIE SF$_6$/Cl$_2$/Fs based dry-etch process leaving metal on the nanowire sidewalls defining the top metal contact.

The same HSQ mask is also used as a bottom spacer by thinning it with a diluted HF (1:1000) wet etch so that the nanowire part that constitutes the channel protrudes. This method allows for sub-100 nm gate-length definition. The exposed semiconductor surface is digitally etched by oxidation followed by either citric acid or an HCL:IPA (1:30) treatment. The number of digital etch cycles are adjusted to remove the overgrown doped InAs shell, restoring the core channel material for the p- and n-type wires, that finishes the gate recess.

To optimize the processing conditions, two types of high-k are compared, a conventional bi-layer Al$_2$O$_3$/HfO$_2$ [19] and a pure HfO$_2$ dielectric. The HfO$_2$ only gate-stack is preferred for aggressive EOT scaling [20]. High temperature (≥250°C) is necessary when forming the interface layers for more effective self-cleaning [19], [21]. The different high-k with corresponding pre-treatments are described further in Table I, representing the best optimization for GaSb p-type (Sample A) and InAs n-type (Sample B) device performance.

The gate metal is defined by sputtering 60 nm W followed by a back-etched S1813 spacer as mask for an SF$_6$ dry-etch process to remove the top of the gate W and vertically align the gate to the top contact. An organic spacer is then deposited, followed by sputtering of the top metal electrode consisting of W/Ni/Au.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Sample A</th>
<th>Sample B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top contact formation$^1$</td>
<td>Ozone oxidation + citric acid 60 s + W/TiN</td>
<td>Ozone oxidation + citric acid 60 s + 1.500 + (NH$_4$I)$_3$S:HfO 3 min</td>
</tr>
<tr>
<td>Gate-stack</td>
<td>HCl:IPA + 5x TMAI</td>
<td>Citric acid + 8x (TMAI pulse and H$_2$ plasma)</td>
</tr>
<tr>
<td>Pre-treatment pulses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-k $^2$</td>
<td>56/36 cycles Al$_2$O$_3$/HfO$_2$: EOT = 0.85 nm</td>
<td>40 cycles HfO$_2$ (+PMA): EOT = 0.76 nm</td>
</tr>
<tr>
<td>n-type doping</td>
<td>nid-channel</td>
<td>n-channel</td>
</tr>
<tr>
<td>$L_0$</td>
<td>50 nm</td>
<td>40 nm</td>
</tr>
<tr>
<td>$g_{m,peak}$</td>
<td>1.2 mS/μm</td>
<td>2.6 mS/μm</td>
</tr>
<tr>
<td>$SS_{ns}$</td>
<td>74 mV/dec</td>
<td>191 mV/dec</td>
</tr>
<tr>
<td>$R_m$</td>
<td>1039 Ω·μm</td>
<td>317 Ω·μm</td>
</tr>
<tr>
<td>p-type $L_0$</td>
<td>70 nm</td>
<td>100 nm</td>
</tr>
<tr>
<td>$g_{m,peak}$</td>
<td>74 μS/μm</td>
<td>11 μS/μm</td>
</tr>
<tr>
<td>$SS_{ns}$</td>
<td>273 mV/dec</td>
<td>622 mV/dec</td>
</tr>
<tr>
<td>$R_m$</td>
<td>5.9 kΩ·μm</td>
<td>20 kΩ·μm</td>
</tr>
</tbody>
</table>

Description of processing variations for two different samples highlighting the critical steps involved. An overview is given for the important performance metrics. $g_{m,peak}$ is extracted at drain bias $V_{ds} = 0.5$ V and $SS_{ns}$ at $V_{ds} = 0.05$ V.

$^2$Contact annealing, post top metal definition, has been performed at 350 degrees C for 60 minutes (Fig. 1). Al$_2$O$_3$ (TMAI) and HfO$_2$ (TDMAHf) is deposited at 300 °C and 120°C respectively. $^3$In situ remote plasma treatment of the surface is carried out inside the ALD chamber and HfO$_2$ (TDMAHf) is deposited at 250°C. Additional post metal anneal (PMA) at 350°C is also performed after gate-definition on sample B.

![Fig. 2. Falsely colored SEM-images depicting single nanowires inside a p-type and n-type structure for Sample A (Table I), highlighting the different HSQ thickness. Combined transfer and output characteristics is presented for a selected p- and n-type device. The n-type (nid-channel) device consists of 184 nanowire array with a pitch of 300 nm, diameter of 20 nm, and $L_0 = 50$ nm. The p-type consists of 144 nanowires with a pitch of 350 nm, diameter of 40 nm, and $L_0 = 70$ nm. Increased resistance is shown in $R_m$ by switching drain and source to realize a top grounded configuration.](image)
Notice that the InAs-GaSb combination introduces a tunneling source contacts with broken bandgap, for the p-type device, where reported values suggest that it may have sufficiently low resistivity (axial Esaki diodes: \(-260 \, \Omega \cdot \mu\)m) [22], [23]. Thus the InAs-GaSb structure, due to resemblance with state-of-the-art Tunnel-FETs [24], [25], lends itself for future CMOS Tunnel-FET implementations including co-integration of MOSFETs and Tunnel-FETs.

### III. DEVICE CHARACTERIZATION

Transfer and output characteristics for sample A with a bi-layer gate-stack is presented in Fig. 2. Transconductance \(g_{m,\text{peak}}\) values of 1.2 mS/\(\mu\)m for n-type and 74 \(\mu\)S/\(\mu\)m for p-type was achieved with minimum subthreshold slope \(SS\text{lin}\) of 74 and 271 mV/dec. Both devices exhibit enhancement mode operation with \(V_T = 0.08\) and -0.02 V. The limited n-type off-state performance is partly caused by drain tunneling due to the narrow bandgap of InAs [26]. However, for the p-type GaSb device the large diameter of the gate-segment in conjunction with a non-optimal high-\(k\) interface degrades the off-state. The 24 times higher drive current for the n-type as compared to the p-type MOSFET is expected due to the large difference in charge carrier mobility [2]. Further, significant gate-length scaling and strained GaSb channels are needed to enable the p-type MOSFET semi-ballistic operation [27], [28].

The resistance originating from the top contact is reduced for the n-type device in sample B (Table I) by adding a sulfur passivation step during the top contact definition, which increases the on-state performance, \(g_{m,\text{peak}} = 2.6\) mS/\(\mu\)m (n-channel), Fig. 3. This is further quantified by switching the source and drain electrodes during measurement that displays a symmetrical behavior with low \(R_{ds}\) values of 317 and 338 \(\Omega \cdot \mu\)m, in contrast to the n-type device in sample A (Fig. 2 inset). This device also utilizes a pure HfO\(_2\) gate-stack with low EOT of 0.76 nm, see Table I. The off-state and electrostatics are limited, which can be attributed to a doped channel and a high-temperature deposition of HfO\(_2\) at 250\(^\circ\)C [29]. The hysteresis window corresponding to the gate-stack is evaluated to 20 mV, on a representative device, at the minimum SS point (\(V_{ds} = -0.2\) to 0.7 V). An additional device with nid-channel, based on the same type of nanowire that has a gate shifted upwards along the nanowire, is also available on the same sample and is presented in Fig. 3. Here, improved electrostatics is shown with \(SS\text{lin}\) decreased from 191 (n-channel) to 80 mV/dec (nid channel). Even though H\(_2\)-plasma on InAs may result in increased surface roughness we find that high transconductance values still can be obtained. The n-type devices, on sample B, with peak transconductance \(g_{m,\text{peak}}\) of 2.6 (n-channel) and 1.7 mS/\(\mu\m (nid-channel), at \(V_{ds} = 0.5\) V, indicate added source resistance due to varied HSQ spacer thickness. The spacer is varied from 10 (n-channel) up to 50 nm (nid-channel), see Fig. 3. Notably the saturation current for the device with doped n-channel is 1.8 mA/\(\mu\m at \(V_{gs} = 0.7\) V. The improved drive current, compared to previous InAs transistor iterations, can be attributed to the addition of doping at the bottom nanowire-segment [12] and n-doped shell growth [13]. The p-type device is demonstrated although with reduced performance and electrostatics as compared to the n-type device quantified by \(g_{m,\text{peak}} = 11\) \(\mu\)S/\(\mu\m (V_{ds} = -0.5\) V) and \(SS\text{lin} = 622\) mV/dec, see Table I.

### IV. CONCLUSIONS

A new process for n-type (InAs) and p-type (GaSb) MOSFET co-integration compatible with highly sensitive antimonide-based materials has been developed. Aggressively scaled devices with sub-100 nm gate-lengths and n-type diameters down to 12 nm have been demonstrated using gate-stack variation, giving a highest \(g_{m,\text{peak}}\) of 2.6 mS/\(\mu\m, see Table I. P-type devices, demonstrating \(g_{m,\text{peak}}\) of 74 and 11 \(\mu\)S/\(\mu\m, are also co-integrated on the same samples. Further optimization of the GaSb gate-stacks combined with aggressive diameter scaling will increase the balance for the III-V CMOS. The HSQ-spacer utilization allows for future integration of p- and n-type on the same nanowire, ultimately reducing the diameter difference between the devices. The process also enables possibilities for co-integrating MOSFETs with state-of-the-art Tunnel-FETs.

### V. ACKNOWLEDGEMENTS

I extend my gratitude to Dr. Peralagu at The University of Glasgow for helpful feedback regarding process development.