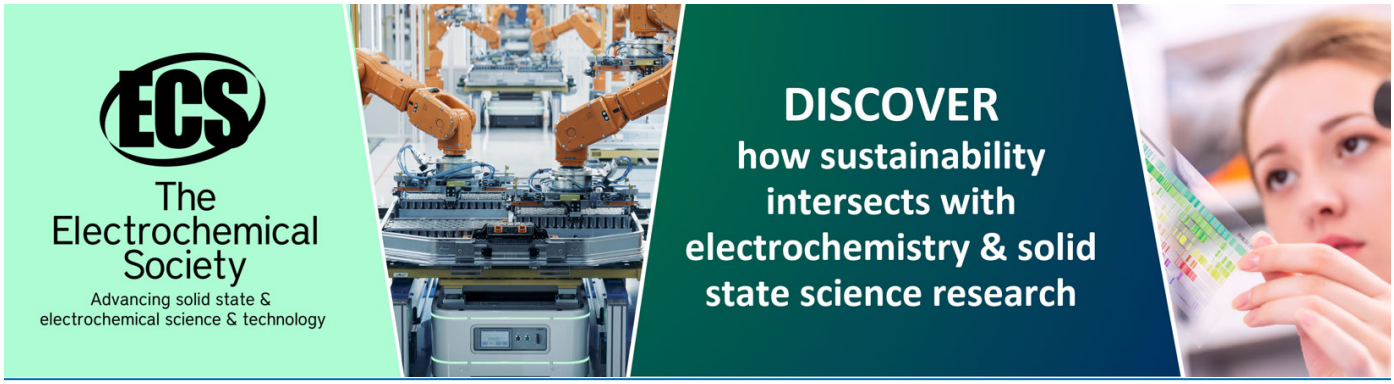
Performance enhancement of GaSb vertical nanowire p-type MOSFETs on Si by rapid thermal annealing



OPEN ACCESS
IOP Publishing
Nanotechnology

Nanotechnology **33** (2022) 075202 (10pp)
https://doi.org/10.1088/1361-6528/ac3689

Performance enhancement of GaSb vertical nanowire p-type MOSFETs on Si by rapid thermal annealing

Zhongyunshen Zhu, Johannes Svensson, Adam Jönsson and Lars-Erik Wernersson

Division of Electromagnetics and Nanoelectronics, Department of Electrical and Information Technology, Lund University, 221 00 Lund, Sweden

E-mail: zhongyunshen.zhu@eit.lth.se

Received 9 August 2021, revised 22 October 2021
Accepted for publication 4 November 2021
Published 24 November 2021

Abstract

GaSb is considered as an attractive p-type channel material for future III-V metal-oxide-semiconductor (MOS) technologies, but the processing conditions to utilize the full device potential such as low power logic applications and RF applications still need attention. In this work, applying rapid thermal annealing (RTA) to nanoscale GaSb vertical nanowire p-type MOS field-effect transistors, we have improved the average peak transconductance ($g_{m,peak}$) by 50% among 28 devices and achieved 70 μS μm^{-1} at $V_{DS} = -0.5$ V in a device with 200 nm gate length. In addition, a low subthreshold swing down to 144 mV dec^{-1} as well as an off-current below 5 nA μm^{-1} which refers to the off-current specification in low-operation-power condition has been obtained. Based on the statistical analysis, the results show a great enhancement in both on- and off-state performance with respect to previous work mainly due to the improved electrostatics and contacts after RTA, leading to a potential in low-power logic applications. We have also examined a short channel device with $L_g = 80$ nm in RTA, which shows an increased $g_{m,peak}$ up to 149 μS μm^{-1} at $V_{DS} = -0.5$ V as well as a low on-resistance of 4.7 k$\Omega\cdot\mu$m. The potential of further enhancement in g_m via RTA offers a good alternative to obtain high-performance devices for RF applications which have less stringent requirement for off-state performance. Our results indicate that post-fabrication annealing provides a great option to improve the performance of GaSb-based p-type devices with different structures for various applications.

Supplementary material for this article is available online

Keywords: GaSb, vertical nanowire, p-type MOSFET, performance enhancement, RTA

(Some figures may appear in colour only in the online journal)

1. Introduction

III-V compound semiconductors are a promising channel material for the next generation high speed complementary-metal-oxide-semiconductor (CMOS) circuits thanks to their high carrier mobilities and injection velocities [1, 2]. High performance n-type metal-oxide-semiconductor field-effect transistors (MOSFETs) based on III-As materials, such as In(Ga)As [3, 4], have successfully demonstrated competitive on-state performance with respect to current Si-based n-MOSFETs. For their p-type counterpart, in spite of the high hole mobility in antimonides such as GaSb and InGaSb [2, 5], fabrication of GaSb-based p-MOSFETs with competitive performance is still challenging. III-V p-MOSFETs are mainly limited by gate-stack properties [6] and high-resistive, non-ohmic contacts [7] resulting in unbalanced

performance in all-III-V CMOS technology [8–10]. Although one alternative with co-integration of conventional SiGe p-type and InGaAs n-type MOSFETs has been proposed [11, 12], the difficulty in material chemistry during processing may limit this type of device integration. Moreover, p-type GaSb play an essential role in heterostructure tunneling FETs [13] for emerging low power application. Thus, the motivation of further investigation and development of III-V p-type transistors still remains. Earlier studies based on (In)GaSb MOSFETs with various device structures [5, 14–17] have shown the challenge to retain the off-state performance when scaling from a long channel device to a short channel device ($L_g < 500$ nm). Thus, the trade-off between on- and off-state performance during gate length scaling therefore needs to be solved.

Recently, in a vertical nanowire (VNW) architecture using a gate-all-around (GAA) geometry, we have demonstrated GaSb VNW p-MOSFETs with improved electrostatic control and high transconductance [18, 19], as compared to other reported GaSb-related MOSFETs. Although an on/off current ratio of about 700 was demonstrated within 1.5 V gate voltage modulation in a 60 nm channel VNW devices [18], a further enhancement is doubtlessly required to approach low power logic applications. Rapid thermal annealing (RTA) is an important technique widely used in semiconductor device fabrication to improve the performance by influencing the material properties, such as dopant activation and defect passivation, as well as contacts improvement [20]. By annealing in a forming gas (N_2/H_2), the electrical properties at the interface between the semiconductor and gate oxides can be improved by the RTA process thanks to the passivation of interface defects (D_{it}), as reported for InGaAs n-channel devices [21]. Several investigations regarding contact improvement using RTA in n-MOSFETs have been also demonstrated, applied on both conventional planar devices [22] and nanoscale VNW MOSFETs [23]. In terms of GaSb-based technologies, it has been proven that the sheet resistance of Ni–GaSb can be lowered by RTA [24, 25] which may offer a possibility to improve the contact. However, for using RTA on the device level, only planar GaSb long channel (gate length $L_g \approx 5\ \mu$m) MOSFETs were reported by comparing individual device performance before and after annealing [17, 26], and it is not clear how the annealing affects the electrostatics and the contacts in short L_g devices.

In this work, we for the first time apply post fabrication RTA to nanoscale GaSb VNW transistors with two different device structures (sample 1:200 nm L_g with Ni top contact; sample 2:80 nm L_g with W top contact) and achieve substantial improvements in transistor performance in both samples. Section 3.1 mainly discusses the on-state performance in the different samples with varying RTA temperatures. Section 3.2 focuses on the influence of RTA on the on-resistance (R_{on}) while section 3.3 discusses the annealing effects on the electrostatics and off-state performance. Finally, a further discussion and the corresponding benchmark will be shown in section 3.4 Our statistical results in this study reveal that RTA can be used as a device performance booster for GaSb p-type transistors in order to reach a specific application.

2. Experimental methods

InAs-GaSb NWs were grown on a prepatterned Si substrate with a 260 nm n^+-InAs buffer layer by metal-organic vapor-phase epitaxy (MOVPE) via vapor–liquid–solid process. InAs buffer layer was used for the integration of III-V materials on Si substrates. The NW growth started with a short n-type InAs stem doped with Sn to not only enable nucleation of GaSb NW growth, but also form the source of the device, which has an InAs/GaSb broken band tunneling junction to assist the carrier transport [27]. Subsequently, undoped GaSb with an estimated background doping of $\sim 10^{16}\,cm^{-3}$ and p-type GaSb NWs doped with Zn (molar fractions: $\chi_{TMGa} = 4.9 \times 10^{-5}$, $\chi_{TMSb} = 5.6 \times 10^{-5}$, DEZn/TMGa = 0.39) were grown for the channel and the drain, respectively. Here, two samples with different structures as well as process schemes were fabricated and compared. Sample 1 (S1) was grown as the above structure and fabricated with a gate-first process starting from the NW bottom. In contrast, a gate-last process starting from the top contact of the NW transistors was employed for sample 2 (S2) which has a 2 nm thick Sn-doped InAs shell as an interfacial layer for the top (drain) contact.

The device fabrication for S1 was initialized by the digital etching while the first step for S2 was the drain contact which requires several additional steps discussed below. Firstly, a 20 nm thick Al_2O_3 was deposited by atomic layer deposition (ALD) as the first spacer which was then removed in the drain region of the NW MOSFET. The length of the drain was defined by a back-etch process with S1813 (photoresist). Next, the drain contact was formed by sputtering 20 nm W which was dry etched anisotropically leaving metal only on the NW sidewalls. The S1813 was further thinned down to define the gate length. Then the Al_2O_3 on the channel region was wet etched and S1813 removed. Next, the diameter of the GaSb channel was reduced by ~10 nm and ~20 nm in S1 and S2, respectively, using repeated digital etching with oxidation inside an oxygen chamber followed by an oxide etch in HCl:IPA (1:10) for 30 s. In the case of S2, the oxidation for the first cycle of digital etch was carried out in an ozone ambient for 30 s at 50 °C to fully oxidize the InAs shell. A bilayer high-k film with 1 nm Al_2O_3/3 nm HfO_2 (EOT ≈ 1 nm) was deposited using ALD. For S1, an extra 20 nm Al_2O_3 layer was grown after the high-k film as the first spacer whose height was defined by the same back-etch process using S1813. The remaining fabrication steps were identical for two samples. A 60 nm W was sputtered for the gate metal and the excess gate-metal was subsequently removed using S1813 mask and dry etching. The gate length in S1 was defined in this step. Both the NW diameter of the channel and the gate length in S1 and S2 were verified by the scanning electron microscopy (SEM) image shown in figures 1(a) and (b), respectively. The samples were finalized by the second spacer deposition and the contact metallization.

Figures 1(c) and (d) illustrate the schematics of the final NW transistor in S1 and S2, respectively. In order to enable post processing annealing of the samples at high temperatures, all the spacers for isolating the terminals are replaced with Al_2O_3 rather than polymers which were usually used as

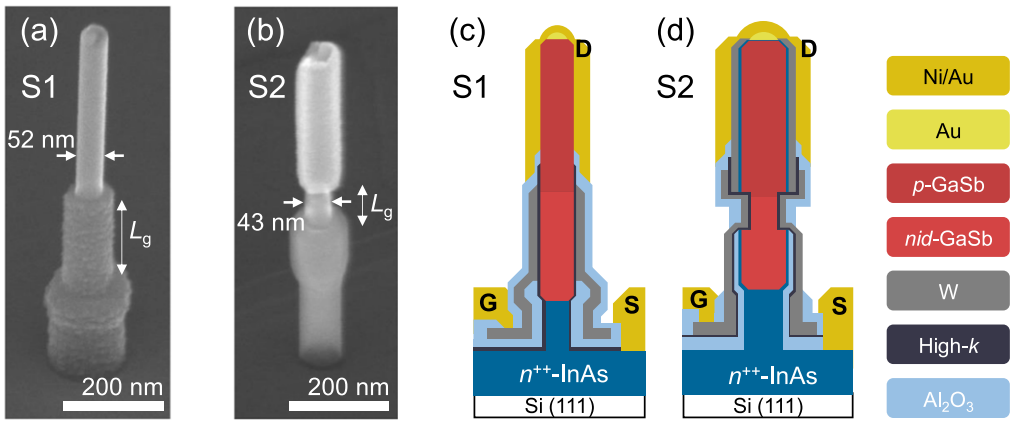


Figure 1. SEM images of the real NW devices: (a) a device in S1 with $L_g = 200$ nm after gate length definition; (b) a device in S2 with $L_g = 80$ nm after high-*k* deposition. The marked diameters include the channel diameter and the high-*k* thickness (4 nm). Schematics of the GaSb single NW transistor are displayed in (c) S1 and (d) S2. G, S and D denote gate, source and drain, respectively. *nid* represents non-intentionally doping.

Table 1. Differences in device structures between two samples.

No.	Process sequence	Top contact	L_g	Channel diameter	RTA
Sample 1 (S1)	Gate-first	Ni/p-GaSb	200 nm	44 nm	200 °C–350 °C
Sample 2 (S2)	Gate-last	W/n-InAs/p-GaSb	80 nm	35 nm	250 °C–350 °C

the second spacer in our previous work [10, 28]. After initial electrical characterizations for the devices, an RTA process was performed in wall-mounted rapid thermal processing (RTP) system, RTP-1200-100, from UniTemp GmbH, with a forming gas (N_2/H_2, 95%/5%) for 2 min at temperatures from 200 °C to 350 °C for S1 and 250 °C to 350 °C for S2. Two step temperature ramping scheme (the temperature first increases to 50 °C less than the setpoint in 30 s and stabilizes for 20 s. Then the temperature continues increasing by a ramping rate of 1.67 °C s^{-1} to the target.) was employed in order to avoid the temperature overshooting. When it reaches the desirable temperature, 2 min is waited before cooling down without supplying any power on the heater. Devices were electrically characterized sequentially after each RTA process. The main information and differences between two samples are summarized in table 1. All the devices in this work are based on single NW transistors. In S2, the highly doped GaSb/InAs core/shell structure in the drain combined with the W–InAs contact can help to lower the resistivity [29, 30].

3. Results and discussion

3.1. Transistor on-state performance improved by RTA

The transfer characteristics of the individual GaSb NW MOSFET in each sample before and after RTA at different temperatures are shown in figure 2. In consideration of the change in threshold voltage (V_T) after RTA (see supplementary material: figure S1 (available online at stacks.iop.org/NANO/33/075202/mmedia)), the gate overdrive voltage, $V_{OV} = V_{GS}–V_T$, is presented in this article instead of the absolute gate bias V_{GS} for better comparison. Generally, as compared to the as-fabricated device performance, the drain current (I_{DS}) as well as g_m increases after RTA and reaches a maximum after annealing at 300 °C in both samples. Further increasing the RTA temperature up to 350 °C, however, degrades the on-state performance in both samples. In S1, an increased on-current (I_{on}, defined at $V_{OV} = -0.5$ V) of 31 μA μm^{-1} as well as a peak g_m ($g_{m,peak}$) of 70 μS μm^{-1} at $V_{DS} = -0.5$ V are achieved in the device after RTA at 300 °C with corresponding increment of ~25% as shown in figures 2(a) and (b), respectively. However, the source depletion of the device in S1 becomes severe after annealing at 350 °C probably due to higher efficiency in hydrogen passivation at the InAs/high-*k* interface than that in GaSb/high-*k* interface so that InAs bands adjacent to the junction move up faster than GaSb bands at high negative gate bias. As a result, the tunneling probability between InAs and GaSb may be reduced, thereby lowering the drain current. Despite a thicker oxide layer including both the high-*k* and the bottom spacer (Al_2O_3) on the InAs segment, substantially lower D_{it} at the interface of InAs/high-*k* likely results in high gating efficiency at high gate voltage. Several reports have shown that the optimal annealing temperature of hydrogen passivation of In(Ga)As/high-*k* interface is close to 350 °C [31, 32]. Therefore, annealing at lower temperatures has relatively limited effects on InAs/high-*k* interface passivation, resulting in less source depletion.

Thanks to a shorter L_g, S2 provides a better on-state performance in both I_{DS} and $g_{m,peak}$, but meanwhile higher

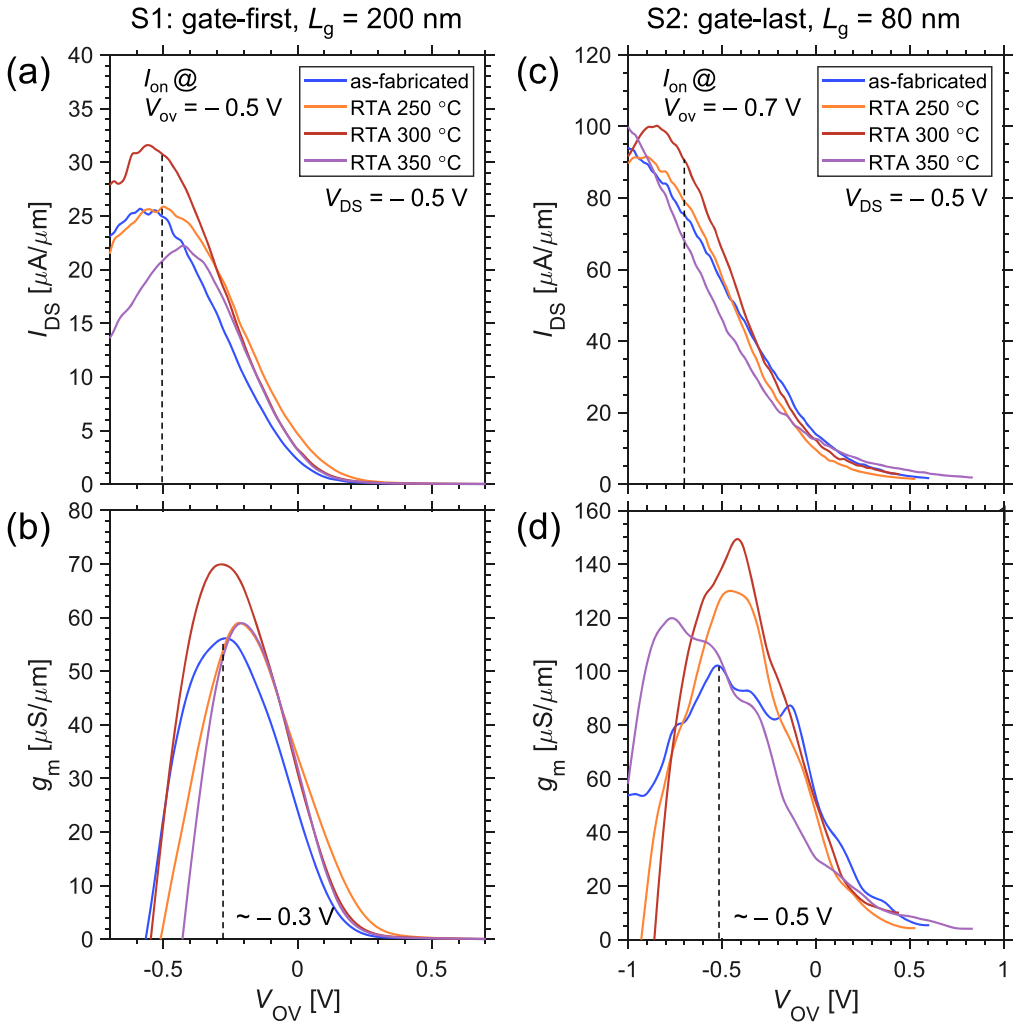


Figure 2. Transfer characteristics of (a), (b) S1 and (c), (d) S2 before (as-fabricated) and after RTA at different temperatures. Here, an overdrive voltage $V_{OV} = V_{GS}–V_T$ is used to align $V_{GS} = V_T$ to 0 in all the cases. $I_{DS}–V_{OV}$ are shown in (a) and (c) while $g_m–V_{OV}$ are shown in (b) and (d). Ion is defined at $V_{OV} = -0.5$ V and -0.7 V for S1 and S2, respectively. The reason of using different V_{OV} for I_{on} definition is to keep the identical gate bias (0.2 V) away from the position of as-fabricated $g_{m,peak}$ in both samples as shown in figures.

overdrive bias (~0.2 V higher in V_{OV}) is needed to achieve $g_{m,peak}$ as compared to S1, seen in figures 2(b) and (d). Thus, to reflect this difference, we defined I_{on} at relatively higher V_{OV} (-0.7 V) for S2. Here, as the comparison of I_{on} only occurs before and after RTA in the same sample, I_{on} may well be defined independently between two samples. In S2, despite only 20% increase in I_{on}, $g_{m,peak}$ is enhanced by almost 50% up to 149 μS μm^{-1} after RTA at 300 °C.

Figure 3 shows the corresponding statistical results based on 28 single NW devices in both S1 and S2, as a function of RTA temperatures. In the case of S1, the median values of I_{on} and $g_{m,peak}$ determined at different RTA temperatures are compared in figures 3(a) and (b), respectively, showing an unambiguous enhancement with increasing the annealing temperature from 200 °C to 300 °C whereas maintaining almost the same value when further increasing the temperature to 350 °C. In figure 3(b), $g_{m,peak}$ increases 50% on average after RTA at 300 °C compared to as-fabricated devices, reaching a maximum median value of 42 μS μm^{-1}. It is noticeable that when increasing the RTA temperature, the minimum $g_{m,peak}$ increases more than 100% and it follows the same trend as the median value while the maximum $g_{m,peak}$ increases roughly 10%. Therefore, the $g_{m,peak}$ increase mainly results from the enhancement in those devices with low as-fabricated $g_{m,peak}$. However, after annealing at 350 °C, those devices start degrading again, resulting in a reduced minimum value in both I_{on} and $g_{m,peak}$. Thus, the spread of the data set after annealing at 300 °C is narrowed but widened again when annealing at 350 °C due to the device degradation. In addition, a similar feature of I_{on} change with the RTA temperature is shown in figure 3(a).

3.2. Annealing effects on R_{on}

Figure 4(a) compares the output characteristics of the same devices plotted in figure 2 before and after RTA. Here, only the output curve at V_{OV} that defines I_{on} is selected to present for both samples, respectively. For the individual devices in S1 and S2, R_{on} first decreases when annealing at 300 °C, reaching 16.7 k$\Omega\cdot\mu$m and 4.7 k$\Omega\cdot\mu$m in the device of S1 and

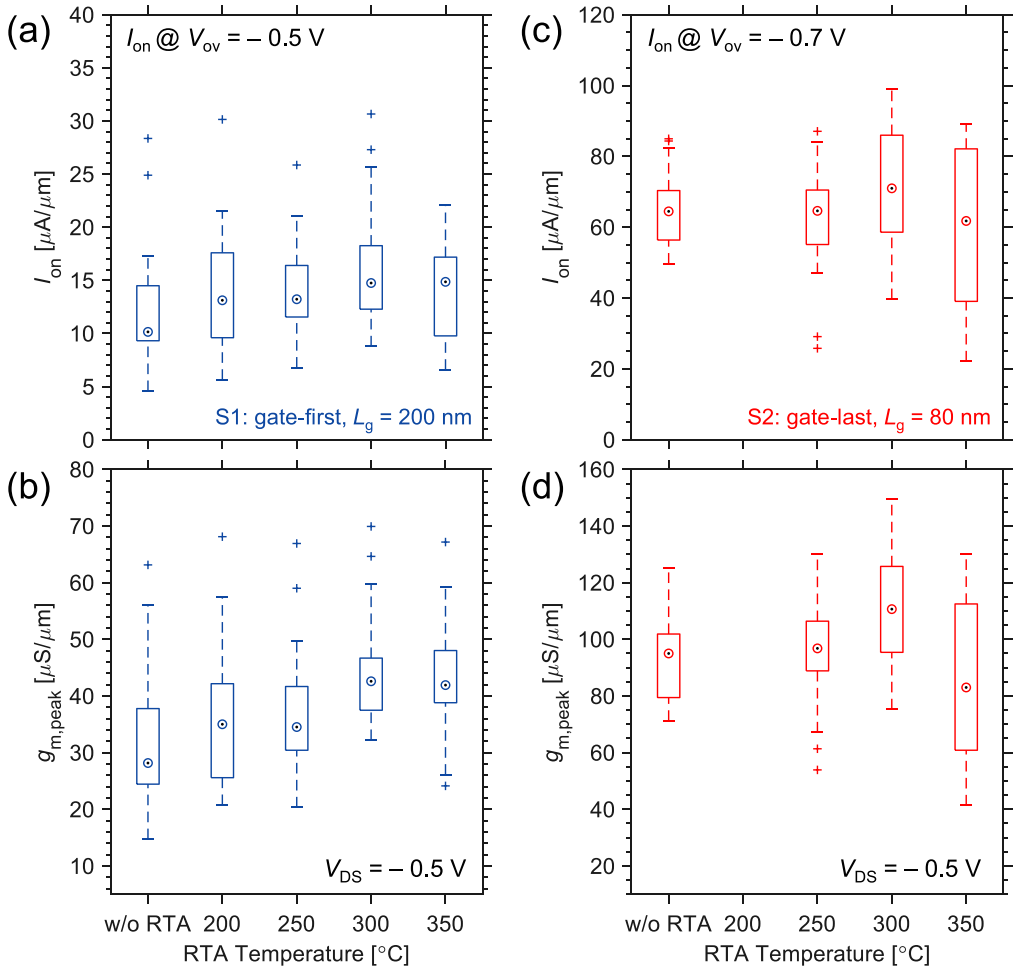


Figure 3. Statistics of (a), (c) I_{on} and (b), (d) $g_{m,peak}$ as a function of RTA temperatures in (a), (b) S1 and (c), (d) S2. Data are shown in a boxplot based on 28 devices in each sample. The same dataset for all other figures in this article. $V_{DS} = -0.5$ V for all plots.

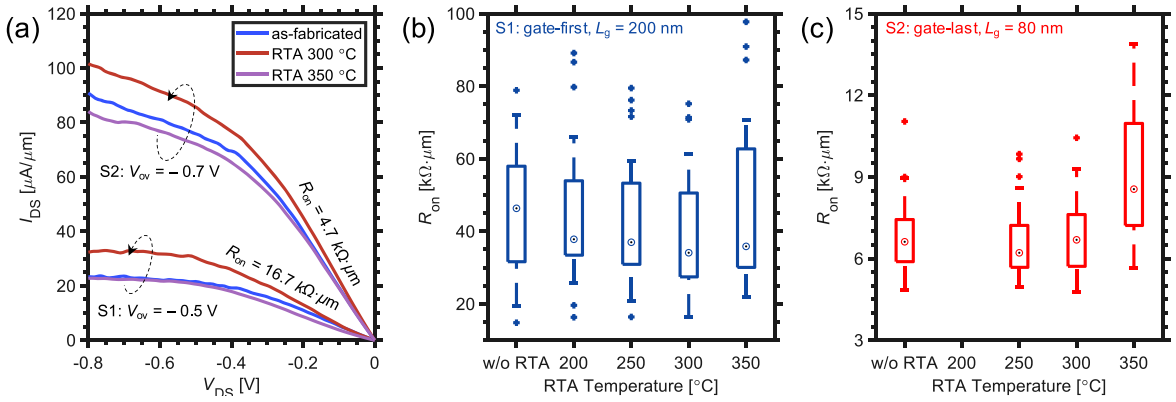


Figure 4. Annealing effects on (a) output characteristics of the same device in figure 2 at the gate bias which is used to define I_{on} in S1 and S2. (b) and (c) show the statistics with boxplots of R_{on} as a function of RTA temperature in S1 and S2, respectively.

S2, respectively, but increases again after RTA at 350 °C. Similar as the individual devices, the corresponding statistical result of R_{on} in S1 shows decreasing R_{on} as increasing the RTA temperature until 300 °C, presented in figure 4(b). In spite of the large variation in R_{on}, it is found that the median R_{on} of all devices annealed at 300 °C is still reduced 26% as compared to that without annealing. In contrast to the clear drop in R_{on} after RTA observed in S1, no noticeable change of R_{on} is observed in S2 in the same RTA temperature interval, shown in figure 4(c). Instead, almost identical median value and device variation are obtained with increasing RTA temperature until 300 °C. Among the 28 studied devices in S2, there are 12 devices with a higher R_{on} after RTA at 300 °C and 16 devices having a lower value. Although the maximum reduction reaches 33%, most of the devices have a change within ±10%, leading to a similar

median value after RTA. As compared to S1, a 7-times lower mean R_{on} (6.5 kΩ·μm) is achieved before annealing in S2, while the gate length is only 2.5 times shorter. Hence, the top contact is likely improved with W–InAs/GaSb configuration, thereby contributing to significant R_{on} reduction in S2.

It is reasonable to assume that the access resistance will remain constant when annealing at such moderate temperatures (200 °C–350 °C). The access resistance is mainly determined by the geometry and the carrier density, which relates to the doping concentration in this case. Since the epitaxial growth temperature of the NWs is much higher than the annealing temperatures, the doping profile should be unchanged after RTA. Therefore, we believe that the R_{on} reduction in S1 after RTA can be mainly interpreted as the improvement of the top contact in the NW transistors. Annealing can promote Ni to alloy with GaSb, leading to more Ni–GaSb alloy formed at high temperature thus lowering the sheet resistance of Ni–GaSb alloy layer [7, 24]. Extrapolating to the transistor level, the contact resistance can be lowered by forming higher conducting Ni–GaSb alloy after annealing. However, higher annealing temperatures may degrade the contact attributed to the presence of a new phase Ni_2Ga_3 (forming at 369 °C according to the phase diagram [33] with higher resistivity in the alloy while the desired NiGa (Sb) phase is only formed at relatively low annealing temperature [25]. Consequently, a slight increase in R_{on} occurs after RTA at 350 °C in S1.

Despite unchanged median value and variation in statistical result of R_{on} in S2, a small variance within 10% is still found in most of the devices. R_{on} increasing or decreasing in this case could be attributed to the relatively slight change in the top contact which is W–InAs/GaSb in S2. W is used as a non-alloy ohmic contact for InGaAs transistors [4] and behaves similarly as Mo [28] for contacting InGaAs. Moreover, a recent study on InGaAs VNW devices with Mo–InGaAs (similar as W–InGaAs) non-alloy contact shows a smaller change in R_{on} with varying RTA temperatures as compared to Ni–InGaAs transistors [23]. Thus, W–InAs/GaSb likely provides a rather thermally stable contact below 300 °C as compared to Ni–GaSb. However, we observed a dramatic increase in not only the median value but the spread of R_{on} after RTA at 350 °C, which may highly relate to the deterioration in channel interface thus leading to a dominantly large channel resistance as compared to considerably low contact resistance in S2. The annealing effects on the interface will be discussed in the next section

3.3. Annealing effects on electrostatics and off-state performance

Figure 5(a) presents the statistics of saturation *SS* (SS_{sat}, at $V_{DS} = -0.5$ V) with varying RTA temperature. For S1, SS_{sat} continuously reduces when increasing the annealing temperature even up to 350 °C. Specifically, the reduction in SS_{sat} is mainly determined by the improvement of devices with originally high *SS*, thus a narrower distribution is achieved, reaching a lowest $SS_{sat} = 166\,\mathrm{mV\,dec^{-1}}$ in a device as shown in figure 5(b). A similar trend based on statistics is observed for SS_{lin} (see supplementary material: figure S2(a)) and a minimum value of $144\,\mathrm{mV\,dec^{-1}}$ is obtained in the same device shown in figure 5(b), which is the lowest subthreshold swing among reported GaSb-based p-MOSFETs [8, 17, 19, 34]. In accordance with the equation [35] $SS \approx (kT/q) \cdot \ln(10)(1 + qD_{it}/C_{ox})$, where k is the Boltzmann constant, T the temperature, q the electron charge and C_{ox} the capacitance of the high-k oxides which is assumed to be invariant with annealing [36]. D_{it} refers to the trap density of the MOS interface, usually describing the interface quality. The lower D_{it}, the better interface quality. Based on this equation, the change of *SS* can be determined by D_{it} only. Thus, we estimated $D_{it} \approx 3.6 \times 10^{13}\,\mathrm{eV^{-1}\,cm^{-2}}$ at $V_{DS} = -0.05$ V after annealing by calculating a coaxial oxide capacitance of $C_{ox} \approx 7\,\mathrm{aF\,nm^{-1}}$. By considering the same device before annealing, D_{it} approximates to $4.4 \times 10^{13}\,\mathrm{eV^{-1}\,cm^{-2}}$, thus indicating a reduction of 22%. Therefore, the reduction in *SS* can originate from the MOS interface improvement by lowering D_{it} attributed to the H_2 passivation capability during RTA.

In contrast, S2 exhibits an opposite trend of SS_{sat} (also see SS_{lin} trend in supplementary material: figure S2(a)) with RTA temperature, showing a significant increase after annealing at 300 °C–350 °C. Although S2 has a higher SS_{sat} than S1 before RTA, a similar drain-induced barrier lowering is found in S1 and S2, suggesting a good electrostatic control without short channel effect when scaling down L_g from 200 to 80 nm. Therefore, the high *SS* before RTA in S2 mainly results from the increase of D_{it} at the channel interface as compared to that of S1 fabricated with the gate-first process. For the gate-last process in S2, many additional steps are required prior to the digital etching and subsequent high-k deposition, resulting in unexpected process-induced contaminations. One likely contamination of the channel can be oxygen vacancy from residual GaSb oxides present from the digital etching. The first 30 s ozone treatment at 50 °C is likely to oxidate through the InAs shell and deeply penetrate into GaSb to form a thick GaSb oxide layer [37] which may be insufficiently etched by HCl:IPA for 30 s in the current process. The presence of oxygen vacancy at the interface may generate more traps at the GaSb/high-k interface thereby increasing *SS*. Although these oxygen vacancies might be reduced by annealing in an ambient with hydrogen, other possible contamination originating from the previous process would be active to react with the channel surface when annealing, probably leading to more interface states. Thus, the dramatic increase of *SS* after RTA is likely related to these interface states activated by annealing. As a result, D_{it} increases substantially after RTA at 300 °C or higher, leading to a poor off-state performance. Further detailed material characterizations for the channel interface in our NWs, such as x-ray photoelectron spectroscopy and transmission electron microscopy, are needed to verify of our hypothesis.

The off-state performance in S1 is further studied. The device with lowest *SS* after RTA at 350 °C also reveals a high on/off current ratio (I_{on}/I_{off}) over 1000 with attractive $I_{on} = 23\,\mu\mathrm{A\,\mu m^{-1}}$ and a low $I_{off} = 19\,\mathrm{nA\,\mu m^{-1}}$ at

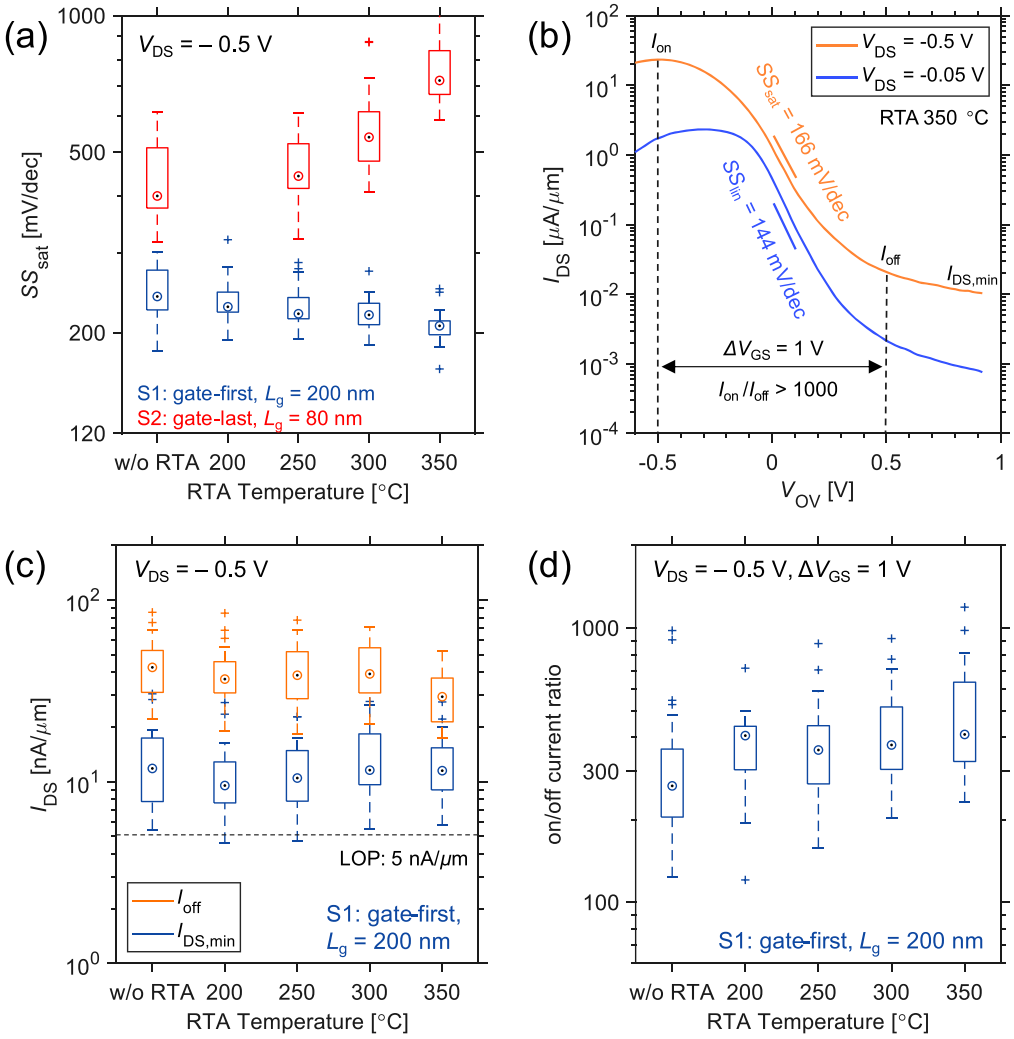


Figure 5. Annealing effects on electrostatics and the off-state performance. (a) Statistics of SS_{sat} in both samples. *Y*-axis is plotted in logarithm. (b) Transfer characteristics of the device with lowest *SS* and highest I_{on}/I_{off} at $V_{DS} = -0.5$ V in S1. I_{on} is defined at $V_{OV} = -0.5$ V while I_{off} is defined at $V_{OV} = 0.5$ V. Statistics of (b) I_{off} and minimum I_{DS} at off-state in the range of $0\text{ V} < V_{GS} < 1\text{ V}$, as well as (c) on/off current ratio I_{on}/I_{off} in S1.

$V_{DS} = -0.5$ V, as demonstrated in figure 5(b). In figure 5(c), a stable $I_{off} = 40\text{ nA }\mu\text{m}^{-1}$ retains until annealing at 300 °C and further reduces to $30\text{ nA }\mu\text{m}^{-1}$ after RTA 350 °C. In addition, the minimum drain current ($I_{DS,min}$) fluctuates with RTA temperatures in a small range around $\sim 10\text{ nA }\mu\text{m}^{-1}$. The lowest $I_{DS,min}$ in all studied devices at off-state reaches $4\text{ nA }\mu\text{m}^{-1}$ after annealing at 200 °C and 250 °C, which fulfills the I_{off} specification of the International Technology Roadmap for Semiconductors low operation power application ($5\text{ nA }\mu\text{m}^{-1}$) [38]. In addition, approximately 50% increase in I_{on}/I_{off} is attained after RTA at 350 °C attributed to both I_{on} increase and I_{off} reduction with respect to as-fabricated performance as shown in figure 5(d). On the other hand, only I_{on} increase accounts for the gradual increase in I_{on}/I_{off} when the annealing at 200 °C–300 °C. The results of S1 show the best balance in on- and off-state performance among recently reported GaSb-related transistors [10, 18, 39], indicating an attractive potential for low power logic applications. In contrast, the off-state performance degrades significantly after RTA process in S2 as both *SS* and I_{off} (see supplementary material: figure S2(b)) increase with annealing temperatures.

Table 2. Summary of the RTA effects on S1 and S2 by comparing the difference before and after RTA at 300 °C.

Ave. change	$g_{m,peak}$	I_{on}	R_{on}	SS_{lin} / SS_{sat}
S1: gate-first	+50%	+47%	−26%	−10%/−9%
S2: gate-last	+20%	+9%	0%	+30/+31%

3.4. Discussion and benchmarking

Based on the analysis in the previous sections, it is evident that RTA at 300 °C provides the best performance improvement with annealing in both S1 and S2. Thus, table 2 summarizes the improvements and changes in each sample after RTA at 300 °C with respect to the as-fabricated performance. For S1, the improvements of not only the top contact but the channel interface after RTA lead to the reduction in both R_{on}

Table 3. Benchmarking devices in S1 with other published GaSb-based p-type MOSFETs at $V_{DS} = -0.5$ V. Blank spaces are due to incomplete data.

	S1 in this work			IEDM18 [19]	TED2020 [18]		IEDM17 [39]	IEDM15 [8]
Structure	VNW-1	VNW-2	VNW-3	VNW	VNW-1	VNW-2	FinFET[a]	LNS[b]
L_g (nm)	200	200	200	60	110	60	20	500
Diameter (nm)	44	44	44	22	24	24	10	20
SS_{lin} (mV dec^{-1})	144	157	162	175		224	260	217
SS_{sat} (mV dec^{-1})	166	189	175	305	216			188
$I_{DS,min}$ (nA μm^{-1})	9	9	4	~3000	~4	28	~2000	5
I_{on} at $V_{GS} = -0.5$ V (μA μm^{-1})	23	31	16	98	3	20	100	~10
I_{on}/I_{off} with $\Delta V_{GS} = 1$ V	1190	980	1210	~30	~750	330	50	~2000

[a] InGaSb channel.
[b] LNS denotes lateral nanosheet.

and *SS*, thus in turn contributing to an enhancement of $g_{m,peak}$ by 50% on average. Surprisingly, for S2, $g_{m,peak}$ achieves 20% increment after annealing whereas R_{on} keeps invariant and *SS* deteriorates due to more unexpected contamination in S2. In spite of an invariant R_{on} on average, slight change still exists in individual devices (see supplementary material), where those devices with reduced R_{on}, the $g_{m,peak}$ increase linearly depends on the reduction in R_{on}. Therefore, the reduction of R_{on} could be the unique contribution to the $g_{m,peak}$ increase, leading to less improvement in percentage of $g_{m,peak}$ for S2 as compared to S1. The detailed correlation of $g_{m,peak}$, R_{on} and *SS* regarding on individual devices in both samples is shown in supplementary material: figures S3 and S4.

The average change reflects the similar conclusions as discussed previously. For S1, it is difficult to significantly improve the top contact and channel interface at the same since effectively improving MOS interface usually requires a higher annealing temperature (approximately 350 °C) [40, 41], which could degrade the top contact of the device (increasing R_{on}) by changing the Ni–GaSb alloy structures as addressed previously. One feasible option to further improve GaSb NW MOSFETs by annealing is probably to include both forming gas (N_2/H_2) annealing directly after gate metal deposition for the MOS interface at a higher temperature and another annealing process at a relatively lower temperature after fabrication to only improve the contacts [41]. For S2, the key issue is the high *SS* which likely results from the process-induced contamination, being strongly degraded by RTA. However, the impressive improvement in g_m exists in those devices with high as-fabricated performance in S2. Thus, although using the characteristics of S2 for digital applications may be challenging, for some RF applications [28] or the current source of all-III-V platform, a high *SS* can still be acceptable. Further gate-stack development can also help to improve *SS*.

Table 3 and figure 6 benchmark this work with recently published GaSb-related p-type MOSFETs with $L_g < 500$ nm. 3 different VNW devices from S1 after annealing are included in table 3, all showing a good off-state current below 10 nA μm^{-1} simultaneously along with a competitive I_{on}, showing a great balance in on- and off-state performance. We

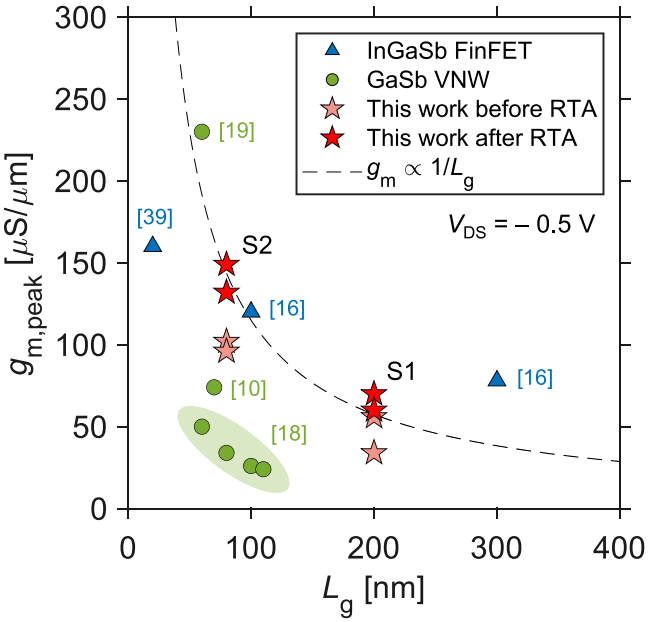


Figure 6. Benchmarking $g_{m,peak}$ versus L_g of the devices in this work and other GaSb-based p-FETs. Stars represent the present work which is in line with state-of-the-art GaSb-based devices after annealing.

have also achieved a record *SS* among all GaSb-based sub-500 nm p-type MOSFETs, verifying a great electrostatic control with the gate-first process. For the benchmarking of $g_{m,peak}$ versus L_g in figure 6, comparing to the performance of not annealed devices, the annealed devices in both S1 and S2 have been improved and both approach the position in line with the gate length scaling in state-of-the-art GaSb-based transistors.

4. Conclusion

We have demonstrated that annealing improves the on-state performance of nanoscale GaSb vertical NW p-MOSFETs on Si with two different structures by RTA. The statistical results show that 50% and 20% increase in $g_{m,peak}$ are achieved in S1 ($L_g = 200$ nm) and S2 ($L_g = 80$ nm), respectively. We also

found that the device structure and processing sequence strongly affect the off-state performance after RTA. For S1, a good off-state is obtained before RTA and remains unchanged after annealing, resulting in an increased on/off current ratio and a low *SS* thanks to the improvement in on-current and channel interface quality, respectively. The results suggest a great balance in on- and off-state performance among all reported sub-500 nm GaSb-based transistors, leading to an attractive potential for III-V based low power logic applications. However, in the case of S2, *SS* strongly increases after RTA probably due to the formation of interface traps during annealing, causing a severe degradation in off-state performance. Despite this, by the means of RTA, a remarkably further enhancement of $g_{m,peak}$ in the devices with as-fabricated high performance is noticed. Thus, RTA can be used as an approach to increase g_m for devices with potential use in III-V RF applications. Our work suggests that the use RTA can be a good strategy to further improve the performance of nanoscale GaSb-based p-type devices with various structures for different applications based on III-V platform technologies.

Acknowledgments

This work was supported by the Swedish Research Council (VR), and the Swedish Foundation for Strategic Research (SSF).

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Zhongyunshen Zhu https://orcid.org/0000-0002-7588-4088
Johannes Svensson https://orcid.org/0000-0002-4673-8225

References

[1] del Alamo J A 2011 Nanometre-scale electronics with III–V compound semiconductors *Nature* **479** 317–23
[2] Borg M *et al* 2017 High-mobility GaSb nanostructures cointegrated with InAs on Si *ACS Nano* **11** 2554–60
[3] Convertino C *et al* 2021 A hybrid III–V tunnel FET and MOSFET technology platform integrated on silicon *Nat. Electron.* **4** 162–70
[4] Kilpi O-P *et al* 2017 Vertical InAs/InGaAs heterostructure metal–oxide–semiconductor field-effect transistors on Si *Nano Lett.* **17** 6006–10
[5] Takei K *et al* 2012 Nanoscale InGaSb heterostructure membranes on Si substrates for high hole mobility transistors *Nano Lett.* **12** 2060–6
[6] Chu R L *et al* 2014 Passivation of GaSb using molecular beam epitaxy Y_2O_3 to achieve low interfacial trap density and high-performance self-aligned inversion-channel p-metal-oxide-semiconductor field-effect-transistors *Appl. Phys. Lett.* **105** 182106
[7] Nishi K, Yokoyama M, Kim S, Yokoyama H, Takenaka M and Takagi S 2014 Study on electrical properties of metal/GaSb junctions using metal-GaSb alloys *J. Appl. Phys.* **115** 034515
[8] Goh K *et al* 2015 Gate-all-around CMOS (InAs n-FET and GaSb p-FET) based on vertically-stacked nanowires on a Si platform, enabled by extremely-thin buffer layer technology and common gate stack and contact modules *2015 IEEE Int. Electron Devices Meeting (IEDM) (7–9 December 2015*
[9] Svensson J, Dey A W, Jacobsson D and Wernersson L-E 2015 III–V nanowire complementary metal–oxide semiconductor transistors monolithically integrated on Si *Nano Lett.* **15** 7898–904
[10] Jönsson A, Svensson J and Wernersson L 2018 A self-aligned gate-last process applied to All-III–V CMOS on Si *IEEE Electron Device Lett.* **39** 935–8
[11] Irisawa T *et al* 2013 Demonstration of InGaAs/Ge dual channel CMOS inverters with high electron and hole mobility using staked 3D integration *2013 Symp. on VLSI Technology (11–13 June 2013*
[12] Deshpande V *et al* 2015 Advanced 3D monolithic hybrid CMOS with sub-50 nm gate inverters featuring replacement metal gate (RMG)-InGaAs nFETs on SiGe-OI Fin pFETs *2015 IEEE Int. Electron Devices Meeting (IEDM) (7–9 December 2015*
[13] Memisevic E, Svensson J, Lind E and Wernersson L 2018 Vertical nanowire TFETs with channel diameter down to 10 nm and point SMIN of 35 mV/Decade *IEEE Electron Device Lett.* **39** 1089–91
[14] Guo L W, Xia L, Bennett B R, Boos J B, Ancona M G and Alamo J A D 2014 Enhancing p-channel InGaSb QW-FETs via process-induced compressive uniaxial strain *IEEE Electron Device Lett.* **35** 1088–90
[15] Yokoyama M *et al* 2014 III–V single structure CMOS by using ultrathin body InAs/GaSb-OI channels on Si *2014 Symp. on VLSI Technology (VLSI-Technology): Digest of Technical Papers (9–12 June 2014*
[16] Lu W *et al* 2015 An InGaSb p-channel FinFET *2015 IEEE Int. Electron Devices Meeting (IEDM) (7–9 December 2015*
[17] Kim S H *et al* 2021 High hole mobility and low leakage thin-body (In)GaSb p-MOSFETs grown on high-bandgap AlGaSb *IEEE J. Electron Devices Soc.* **9** 42–8
[18] Jönsson A, Svensson J, Lind E and Wernersson L E 2020 Gate-length dependence of vertical GaSb nanowire p-MOSFETs on Si *IEEE Trans. Electron Devices* **67** 4118–22
[19] Jönsson A 2018 Balanced drive currents in 10–20 nm diameter nanowire All-III-V CMOS on Si *2018 IEEE Int. Electron Devices Meeting (IEDM) (1–5 December 2018*
[20] Fair R B 1993 *Rapid Thermal Processing: Science and Technology* (San Diego: Elsevier Science) (https://doi.org/10.1016/C2009-0-22393-2)
[21] Shin B, Weber J R, Long R D, Hurley P K, Van de Walle C G and McIntyre P C 2010 Origin and passivation of fixed charge in atomic layer deposited aluminum oxide gate insulators on chemically treated InGaAs substrates *Appl. Phys. Lett.* **96** 152908
[22] Zhang X *et al* 2011 In0.7Ga0.3As channel n-MOSFET with self-aligned Ni–InGaAs source and drain *Electrochem. Solid-State Lett.* **14** H60
[23] Zhao X, Heidelberger C, Fitzgerald E A, Lu W, Vardi A and Alamo J A D 2018 Sub-10-nm-diameter InGaAs vertical nanowire MOSFETs: Ni versus Mo contacts *IEEE Trans. Electron Devices* **65** 3762–8
[24] Zota C B, Kim S-H, Yokoyama M, Takenaka M and Takagi S 2012 Characterization of Ni–GaSb alloys formed by direct reaction of Ni with GaSb *Appl. Phys. Express* **5** 071201

[25] Lin K-L and Chen S-H 2014 Interfacial characterization and electrical properties of Ni–GaSb contacts *Appl. Phys. Lett.* **105** 141603
[26] Zeng Z *et al* 2014 Impacts of annealing processes on the electrical properties of gasb metal-oxide-semiconductor devices *2014 12th IEEE Int. Conf. on Solid-State and Integrated Circuit Technology (ICSICT) (28–31 October 2014*
[27] Memišević E, Svensson J, Hellenbrand M, Lind E and Wernersson L 2016 Scaling of vertical InAs–GaSb nanowire tunneling field-effect transistors on Si *IEEE Electron Device Lett.* **37** 549–52
[28] Kilpi O *et al* 2020 High-performance vertical III-V nanowire MOSFETs on Si with gm >3 mS μm^{-1} *IEEE Electron Device Lett.* **41** 1161–4
[29] Ganjipour B *et al* 2014 Electrical properties of GaSb/InAsSb core/shell nanowires *Nanotechnology* **25** 425201
[30] Lin J, Antoniadis D A and Alamo J A D 2016 InGaAs quantum-well MOSFET arrays for nanometer-scale ohmic contact characterization *IEEE Trans. Electron Devices* **63** 1020–6
[31] Hu J, Philip and Wong H S 2012 Effect of annealing ambient and temperature on the electrical characteristics of atomic layer deposition Al2O3/In0.53Ga0.47As metal-oxide-semiconductor capacitors and MOSFETs *J. Appl. Phys.* **111** 044105
[32] Olausson P, Södergren L, Borg M and Lind E 2021 Optimization of near-surface quantum well processing *Phys. Status Solidi A* **218** 2000720
[33] Okamoto H 2008 Ga–Ni (gallium–nickel) *J. Phase Equilib. Diffus.* **29** 296
[34] Tsai M, Chang Y and Chien C 2018 Enhancing the thermal stability of GaSb Schottky-barrier MOSFET with Pt source/drain *IEEE Electron Device Lett.* **39** 939–42
[35] Xu W, Wang H, Ye L and Xu J 2014 The role of solution-processed high-κ gate dielectrics in electrical performance of oxide thin-film transistors *J. Mater. Chem.* C **2** 5389–96
[36] Modreanu M *et al* 2006 Investigation of thermal annealing effects on microstructural and optical properties of HfO_2 thin films *Appl. Surf. Sci.* **253** 328–34
[37] Tan Z *et al* 2013 Effects of ozone pre-deposition treatment on GaSb MOS capacitors *2013 Spanish Conf. on Electron Devices (12–14 Febuary 2013*
[38] Visciarelli M, Gnani E, Gnudi A, Reggiani S and Baccarani G 2017 Design guidelines for GaSb/InAs TFET exploiting strain and device size *Solid-State Electron.* **129** 157–62
[39] Lu W *et al* 2017 10 nm fin-width InGaSb p-channel self-aligned FinFETs using antimonide-compatible digital etch *2017 IEEE Int. Electron Devices Meeting (IEDM) (2–6 December 2017*
[40] Nainani A *et al* 2011 Optimization of the Al_2O_3/GaSb interface and a high-mobility GaSb pMOSFET *IEEE Trans. Electron Devices* **58** 3407–15
[41] Yuan Z *et al* 2013 Antimonide-based heterostructure p-channel MOSFETs with Ni-alloy source/drain *IEEE Electron Device Lett.* **34** 1367–9