First Transistor Demonstration of Thermal Atomic Layer Etching: InGaAs FinFETs with sub-5 nm Fin-width Featuring in situ ALE-ALD

Wenjie Lu\textsuperscript{1}, Younghee Lee\textsuperscript{2}, Jessica Murdzek\textsuperscript{2}, Jonas Gertsch\textsuperscript{2}, Alon Vardi\textsuperscript{1}, Lisa Kong\textsuperscript{1}, Steven M. George\textsuperscript{2}, and Jesús A. del Alamo\textsuperscript{1}

\textsuperscript{1}Microsystems Technology Laboratories, MIT, Cambridge, MA, 02139, USA; \textsuperscript{2}Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO, 80309, USA

E-mail: wenjie@mit.edu

\textbf{Abstract}—For the first time, thermal atomic layer etching (ALE) on InGaAs-based III-V heterostructures is demonstrated. Also, we report the first transistors fabricated by the thermal ALE technique in any semiconductor system. We further highlight one unique advantage of thermal ALE: its integration with atomic layer deposition (ALD) in a single vacuum chamber. Using \textit{in situ} ALE-ALD, we have fabricated the most aggressively scaled self-aligned In\textsubscript{0.53}Ga\textsubscript{0.47}As n-channel FinFETs to date, featuring sub-5 nm fin widths. The narrowest FinFET with $W_f = 2.5$ nm and $L_g = 60$ nm shows $g_m = 0.85$ mS/\mu m at $V_{ds} = 0.5$ V. Devices with $W_f = 18$ nm and $L_g = 60$ nm demonstrate $g_m = 1.9$ mS/\mu m at $V_{ds} = 0.5$ V. Subthreshold swings averaging $S_{in} = 70$ mV/dec and $S_{sat} = 74$ mV/dec across the entire range of $W_f$ at minimum $L_g = 60$ nm have been obtained. These are all record results. The transistors demonstrated here show an average 60\% $g_m$ improvement over devices fabricated through conventional techniques. These results suggest a very high-quality MOS interface obtained by the \textit{in situ} ALE-ALD process.

\section{I. INTRODUCTION}

As CMOS technology continues to scale down and device structures become more three-dimensional, manufacturing challenges compound. In recent years, 3D MOSFETs with sub-10 nm physical dimensions have been demonstrated in various material systems, such as Si, SiGe, and III-V’s [1-5]. Further scaling progress demands fabrication technologies with Ångstrom-scale precision and fidelity. This is out of reach for mainstream plasma etching and wet digital etch techniques.

Atomic layer etching (ALE) is a novel technique that removes materials using sequential self-limiting processes [6-8]. There are two types of ALE. One uses energetic ions or neutrals, commonly assisted by plasma, and the etching is usually anisotropic. The other type is based on the chemical ligand-exchange, and it enables isotropic etching. This is usually referred to as “Thermal ALE”, and its reaction sequence closely resembles that of an ALD process. Thermal ALE is still in its youth, and reports on thermal ALE are limited to etching of dielectrics, metals and some nitriles [9-11]. To our knowledge, there are no device demonstrations to date.

In this work, we report on the development of the first thermal ALE process for InGaAs-InAlAs heterostructures. InGaAs is a promising channel material for CMOS scaling and memory applications [12, 13]. The performance of advanced InGaAs FinFETs is still lacking, partly due to limitations in the MOS stack quality [14]. Thermal ALE is a breakthrough technology that can address these problems. In this work, we demonstrate: (1) precise and highly controllable etching rate at Ångstrom/cycle-scale, (2) plasma-free conformal sidewall etching resulting in low damage and smooth surfaces, (3) material selectivity to enable fabrication of gate-all-around (GAA) structures, and, most importantly, (4) integration of ALE and ALD in an \textit{in situ} process that completely prevents air exposure of the gate oxide-semiconductor interface. These unique attributes enable innovative transistor designs with remarkable ON and OFF-state performance.

We illustrate the device worthiness of our ALE technique by fabricating the most aggressively scaled InGaAs FinFETs to date with fin widths as narrow as 2.5 nm. Record device characteristics highlight the extraordinary device potential of the \textit{in situ} thermal ALE-ALD process.

\section{II. THERMAL ATOMIC LAYER ETCHING}

Fig. 1 shows the schematic of the viscous flow ALD reactor in which both the thermal ALE and ALD processes are performed. Fig. 2 shows the InGaAs/InAlAs heterostructure used to develop the thermal ALE process. It consists of 30 and 40 nm In\textsubscript{0.53}Ga\textsubscript{0.47}As (with different doping) on an In\textsubscript{0.52}Al\textsubscript{0.48}As buffer layer, on (100) InP substrate. Fig. 3 shows the sequence of a complete cycle of thermal ALE of InGaAs. The first step is surface fluorination using HF-pyridine. The second step is a ligand-exchange process to remove the metal fluoride layer. For this we use dimethylaluminum chloride (DMAC) [15] at a partial pressure of 40 mTorr. The volatile metal etch products are then purged away, and the sequence is repeated in cycles. The entire process is performed at 300 °C. The reactor has a baseline vacuum of 5-10 mTorr and a working pressure of 1 Torr with N\textsubscript{2}: flow. Fig. 4 displays X-ray reflectivity scans of the substrate before and after 200 and 450 cycles of thermal ALE, showing an average etch rate for InGaAs of 0.21 Å/cycle for the first 200 cycles and 0.16 Å/cycle for the last 250 cycles.

Detailed ALE etch rate calibrations are obtained from fins and vertical nanowires (VNW) etched by RIE in a BCl\textsubscript{3}/SiCl\textsubscript{4}/Ar ICP plasma [16]. After RIE, the samples are etched by thermal ALE for 250 cycles. We observe (Fig. 5) an average radial etch rate for In\textsubscript{0.53}Ga\textsubscript{0.47}As and In\textsubscript{0.52}Al\textsubscript{0.47}As of −0.2 and −0.6 Å/cycle, respectively, with smooth substrate and sidewall surfaces.

The 3:1 etching selectivity between InGaAs and InAlAs can be exploited to create suspended InGaAs GAA fins or...
vertical nano-sheet MOS structures. The heterostructure is shown in Fig. 7a. Fig. 6 shows TEM images of 50 nm tall InGaAs fins obtained by 250 cycles of ALE, with Al2O3 deposition in situ in the same reactor, finished by W metal ALD in a separate reactor. Fins with Wf < 24 nm are fully suspended, consistent with the ALE selectivity measured above. Fins as narrow as 3-4 nm are obtained. Fig. 6 shows a remarkably sharp interface between InGaAs and Al2O3.

III. InGaAs FinFET Fabrication

The starting heterostructure and cross-section schematics of the finished devices are illustrated in Fig. 7. The channel layer consists of 50 nm thick In0.53Ga0.47As lattice matched to an InAlAs buffer on an InP substrate. A δ-doping layer is placed 5 nm below the channel (Nd = 4·10¹² cm⁻²). A 30 nm n⁺-In0.53Ga0.47As cap is placed above a 4 nm InP etch stopper.

Fig. 8 outlines the fabrication process. The process starts with Mo/W sputtering for the ohmic contacts. Then, CVD SiO2 is deposited as contact spacer and hard mask to etch the gate foot. This is defined by e-beam lithography. After SiO2 and Mo/W RIE and mesa lithography, the heavily doped InGaAs cap is recessed. This is performed in two steps. First, timed RIE is used to remove most of the InGaAs cap. This is followed by a 5 s citric acid:H₂O₂ wet etch to expose the InP etch stopper with minimal lateral etching. This results in a highly self-aligned geometry, with less than 5 nm extrinsic region to minimize series resistance (Fig. 9a).

The process follows with fin patterning by e-beam lithography using HSQ as the mask. 220 nm tall fins are etched in BCl3/SiCl4/Ar ICP plasma [16]. Then, 4 cycles of alcohol-based digital etch, using methacnolic H2SO4 and O₂ plasma, are carried out to shrink the fin width [17]. Following this, the samples are introduced into the ALE/ALD reactor. First, 162 cycles of thermal ALE are performed at 300 °C, as described in the previous section. After this, the substrate temperature is reduced to 250 °C, followed by ALD deposition of 3 nm of H₂O₂ (EOT = 0.8 nm). In a separate ALD reactor (due to limited precursor lines), 30 nm of W are deposited at 130 °C as the gate metal. Fig. 9b shows devices after gate stack formation. Fins with final Wf < 10 nm are suspended. The process continues with the gate head being defined by e-beam lithography and W patterning by SF6/O₂ RIE. A backend process composed of inter-level dielectric (ILD) deposition, via etch and pad metallization completes the device fabrication.

Final Wf and Lg of the devices are measured by TEM and SEM, respectively. The final Wf ranges from 2.5 nm to 18 nm, Lg between 60 nm to 1 μm. Fig. 10 shows the TEM cross-section of a finished device with Wf = 2.5 nm. The inset shows a close-up image of the upper portion of a fully suspended InGaAs channel. In the next section, unless indicated otherwise, all metrics are normalized by total conducting gate periphery.

IV. Electrical Characteristics

Fig. 11 and 12 show electrical characteristics of the most scaled InGaAs FinFET with Wf = 2.5 nm and Lg = 60 nm (AR = Hg/Wf = 20). Classic MOSFET behavior is obtained, showing excellent Slin = 62 mV/dec, Ssat = 68 mV/dec, and DIBL = 40 mV/V. A maximum g_m of 0.85 mS/μm is obtained at VDS = 0.5 V. This is the InGaAs FinFET with the thinnest fin width and highest aspect ratio ever demonstrated. Fig. 13 shows electrical characteristics of a FinFET with Wf = 6 nm and Lg = 60 nm. It shows Slin = 61 mV/dec, Ssat = 72 mV/dec, and DIBL = 50 mV/V. In the widest device (Wf = 18 nm, Lg = 60 nm), we demonstrate maximum g_m = 1.9 mS/μm at VDS = 0.5 V (Fig. 14). In all these devices, the OFF-state current is limited by gate leakage.

Fig. 15 summarizes the scaling behavior of peak g_m (VDS = 0.5 V) with Wf and Lg. Compared with InGaAs FinFETs fabricated without ALE (same heterostructure and EOT) [3], a consistent improvement of ~60% in peak g_m is obtained for Lg = 60 nm. Fig. 16 summarizes Slin and Ssat (VDS = 0.05, 0.5 V, respectively) of Lg = 60 devices of different Wf, together with identical InGaAs FinFETs fabricated without ALE, with and without δ-doping under the channel [3]. Average Slin = 70 mV/dec and Ssat = 74 mV/dec are obtained across the entire range of Wf at Lg = 60nm. Fig. 17 shows the scaling of Slin with Lg at Wf = 9-10 nm. A remarkable improvement in S in ALE-fabricated devices is obtained with values in the 60-80 mV/dec range and weak sensitivity to Wf. The extraordinary enhancement in electrostatic control confirms the very high interface quality obtained by in situ ALE-ALD.

Fig. 18 shows DIBL (VDS = 0.05 and 0.5 V) vs. Lg for Wf = 6-7 nm and 18-19 nm. Excellent improvement in short-channel effects and electrostatics are demonstrated. Fig. 19 shows the scaling behavior of R on at fixed VDS = 0.6 V, together with R on of δ-doped and undoped FinFETs without ALE treatment (Lg = 40 – 60 nm). The very tight self-aligned process developed here yields a lower R on that increases weakly as Wf decreases.

Fig. 20 benchmarks peak g_m, normalized by conducting gate periphery, and g_m/Wf, normalized by fin footprint, for InGaAs FinFETs from this work and [3], and from the literature (VDD = 0.5 V), vs. Wf. For reference, g_m of Intel’s Si FinFETs is also shown (VDD = 0.8 V for 1st gen, and VDD = 0.7 for the 2nd and 3rd gen). With the usual caveats when making comparisons of this kind, our InGaAs FinFETs match the performance of Intel’s 14 nm node (Wf = 7 nm), in spite of the lower VDD and longer Lg. At Wf = 2.5 nm, this work shows a record g_m/Wf > 30 mS/μm. Given that this is the first demonstration of III-V MOSFETs by ALE and the first demonstration of working III-V FinFETs at Wf < 5 nm, this work displays the great promise for both ALE technology and III-V devices.

V. Conclusions

We have developed thermal ALE for III-V heterostructures and demonstrated a FinFET fabrication process that incorporates thermal ALE in combination with in situ ALD to form the gate stack. To our knowledge, this is the first demonstration of thermal ALE in a transistor of any kind. We achieve the most aggressively scaled InGaAs FinFETs with Wf = 2.5 nm and a record AR = 20 among all existing FinFETs. In situ thermal ALE/ALD yields remarkable improvements of device performance and electrostatic control. Record g_m/Wf have been obtained in these devices.
Acknowledgement: This work was sponsored in part by DTRA (#HDTRA1-14-1-0057), SRC (#2016-LM-2655), and Lam Research. Devices fabrication was performed at the Microsystems Technology Laboratories and SEBL at MIT. The University of Colorado authors acknowledge support from Intel Corporation through SRC and additional support from NSF (CHE-1609554).


Fig. 1. Schematic of the hot wall viscous flow reactor used for ALE and ALD.

Fig. 2. Starting hetero-structure for the development of InGaAs/InAlAs thermal ALE process. The same structure is used for fin and VNW fabrication by ALE of Fig. 5.

Fig. 3. Schematic presentation of the InGaAs thermal ALE process: (a)-(b) fluorination of InGaAs surface with HF. (c)-(d) ligand-exchange process by DMAC to remove the metal fluoride layer. The volatile etch products are then purged away.

Fig. 4. X-ray reflectivity scan of InGaAs heterostructure after 200 and 450 cycles of DMAC/HF thermal ALE at 300°C. The average etch rate is 0.18 Å/cycle.

Fig. 5. InGaAs/InAlAs VNW and fin structures (a) after RIE (VNW has initial diameter of 35 nm), and (b) after 250 cycles of thermal ALE at 300 °C. The final diameter of InGaAs VNW is 24 nm (r = 0.2 Å/cycle), and that of InAlAs is 4 nm (r = 0.6 Å/cycle). The fin sidewall shows a smooth surface after the ALE process.

Fig. 6. Cross-section TEM images of (a) array of InGaAs fins (Wf = 3-30 nm) fabricated by in situ ALE-ALD process, (b) InGaAs suspended fin with minimum Wf of 3 nm, and (c) close-up image of the fin in (b).

Fig. 7. (a) Starting heterostructure for InGaAs n-channel FinFET fabrication. Cross-section schematics of FinFETs: (b) along the fin length direction and (c) across the fin. For narrow fin widths, the InGaAs channel is fully suspended.

Fig. 8. Process flow for InGaAs FinFET fabrication.

Fig. 9. SEM images of FinFETs after (a) gate recess, and (b) ALE-ALD gate process.
**Fig. 10.** Cross-section TEM image of finished device with $W_f = 2.5$ nm. Inset: close-up image of upper portion of fin.

**Fig. 11.** (Left) output and (right) subthreshold characteristics of the most scaled InGaAs FinFETs with $W_f = 2.5$ nm and $L_g = 60$ nm.

**Fig. 12.** (Left) transconductance and (right) subthreshold swing characteristics of InGaAs FinFETs with $W_f = 2.5$ nm and $L_g = 60$ nm. Maximum $g_m = 0.85$ mS/$\mu$m, and minimum $S_{lin} = 62$ mV/dec and $S_{sat} = 68$ mV/dec are obtained.

**Fig. 13.** (Left) output and (right) subthreshold characteristics of InGaAs FinFETs with $W_f = 6$ nm and $L_g = 60$ nm.

**Fig. 14.** Subthreshold characteristics of wide-fin device with $W_f = 18$ nm and $L_g = 60$ nm. Inset shows peak $g_m = 1.9$ mS/$\mu$m.

**Fig. 15.** (Left) scaling of $g_m$ vs. $L_g$ at various $W_f$, and (right) scaling of $g_m$ vs. $W_f$ at $L_g = 60$ nm of InGaAs FinFETs with and without thermal ALE [3] ($V_{DS} = 0.5$ V).

**Fig. 16.** Scaling of (left) $S_{lin}$ and (right) $S_{sat}$ vs. $W_f$ at $L_g = 60$ and 100 nm, of FinFETs with ALE ($\delta$-doped) and without ALE ($\delta$-doped and undoped) [3].

**Fig. 17.** Scaling of $S_{lin}$ vs. $L_g$ at $W_f = 9$-10 nm, of FinFETs with ALE ($\delta$-doped and without ALE ($\delta$-doped and undoped) [3].

**Fig. 18.** DIBL at $V_{DS} = 0.05$, 0.5 V vs. $L_g$ at (left) $W_f = 6$-7 nm, and (right) 18-19 nm, of FinFETs with ALE ($\delta$-doped) and without ALE ($\delta$-doped and undoped) [3].

**Fig. 19.** $R_{on}$ (at $V_{GS}=0.6$ V) vs. $W_f$ of FinFETs with ALE ($\delta$-doped) and without ALE ($\delta$-doped and undoped) [3].

**Fig. 20.** Benchmark of (left) $g_m$ normalized by conducting gate periphery, annotated by AR ($H_c/W_f$), and (right) $g_m/W_f$, as a function of $W_f$ from this work and the InGaAs literature. State-of-the-art Si FinFETs are also included. $V_{DD} = 0.5$ V for InGaAs FinFETs and 0.7 V for the 2nd and 3rd generation Si FinFETs.