Novel Heterogeneous Integration Technology of III–V Layers and InGaAs FinFETs to Silicon

Xing Dai, Binh-Minh Nguyen, Yoontae Hwang, Cesare Soci, and Shadi A. Dayeh*

Heterogeneous integration of III–V compound semiconductors to Si substrates is regarded as a necessary step for advancing high-speed electronics and hybrid optoelectronic systems for data processing and communications, and is extensively being pursued by the semiconductor industry. Here, an innovative fab-compatible, hybrid integration process of III–V materials to Si, namely InGaAs thin films to insulator-on-Si, is reported, and the first III–V FinFET devices on Si are demonstrated. Transfer of crystalline InGaAs FinFETs on insulator on Si is systematically investigated for a broad range of channel lengths and Fin perimeters with excellent switching characteristics. This demonstrates a viable approach to large-scale hybrid integration of active III–V devices to mainstream Si CMOS technology, enabling low-power electronic and fully-integrated optoelectronic applications.

1. Introduction

In the past three decades, III–V compound semiconductor materials have been widely used in high-speed integrated circuits for communication systems, enabling life-changing applications such as cell phones, broadband wireless, satellite communications, space, radar, fiber optic communications, and radio telescopes.[1] Silicon transistors on the other hand have reduced in size by over a million times since their invention in the late ‘40s, thereby increasing density and functionality while simultaneously reducing chip power consumption and production costs; meanwhile, their operation speed increased by over a billion times thanks to excellent processing control and recent rich advancements in strain engineering,[2] metal gate stack with high-k dielectrics,[3] and transistor architecture,[4] up to the development of Si FinFET devices adopted in current CMOS technology. It is expected that downsizing of Si-based technology will eventually reach its physical limits below 10 nm technology node, a bottleneck that research now tries to overcome by replacing the channel material. A potential approach is to combine the advantages of III–V FET technology with the well-established Si CMOS platform.[5] Several strategies for large scale III–V to Si heterogeneous integration are currently being pursued either through monolithic integration, where III–V materials are epitaxially grown on Si substrates, or through hybrid integration, where extrinsic III–V layers are bonded to Si substrates. Significant progress has been accomplished recently in the monolithic integration with buffer layers thinner than 2 µm on 4° offcut (100) Si.[6] Using epitaxial necking or aspect ratio trapping, the thickness of these buffer layers can be considerably reduced as successfully demonstrated for GaAs,[7] and InP[8] among others. While these are tremendously significant advances, a defect-free III–V channel on Si using these approaches has not been currently achieved. On the other hand, hybrid integration with processes including Smart Cut,[9] and direct covalent wafer bonding based on oxide eutectics,[10] require extremely smooth surfaces (typical surface roughness of ≤0.5 nm per bonded surface), which overall results in a tight process and low yield. Other hybrid integration approaches utilize intermediate bonding layers such as eutectics,[11] and polymer layers,[12] where the former is not CMOS favored due to the low melting temperature (with usually non-Si-compatible Au composition) and the latter has poor thermal conduction and has limited thermal budget for post-processing (e.g., post transfer metal alloy contact formation).

Recent advances in direct layer transfer upon epitaxial lift-off (ELO) of thin compound semiconductor films using sacrificial layers have shown enormous potential for large-scale integration of III-V on heterogeneous substrates for photovoltaic,[13] LED[14] and FET[15] applications. Among the main advantages, research demonstrated that ELO and bonding allow large area transfer of various III–V crystalline compounds on Si, without limitations imposed by lattice mismatch constraints and without requiring complex processing steps.[16,17] Here we develop a new approach to demonstrate a reliable III-V transfer process based on solid-state diffusion of Ni, deposited on a dielectric
layer that caps the III-V surface, into an exposed Si surface to form an interfacial NiSi bonding layer; further, we demonstrate that this method yields III-V bonded layers of high quality by fabricating high-performance InGaAs FinFET devices.

In the quest for higher density of integration, current CMOS transistor technology has evolved from planar device architectures to the so-called FinFET geometry, which takes its name from the shape of the active channel resembling a vertical, three-dimensional Fin. The main benefit of this geometry is that the gate can be formed conformally around the three sides of the Fin (tri-gate) to achieve full control of the potential of the conducting channel. The multi-gate structure can effectively improve the transistor electrostatics in terms of lowering supply voltage and reducing short channel effects (SCE), which are widely observed in planar transistors where the source/drain fields penetrate deep into the transistor channel under the gate reducing its effectiveness in controlling current flow and threshold voltage. Moreover, the FinFET architecture introduces fewer changes to conventional planar transistor design compared to surrounding gate transistors and multibridge-channel MOSFET, two of the other three-dimensional competing technologies. Similar FinFET structures have already been adopted in high-mobility III-V transistors, for instance made of InGaAs grown epitaxially on InP substrates with high-k gate dielectrics, which showed improved of the short channel performance.

2. Results and Discussion

To exploit the advantages of FinFET transistors in a manufacturable platform, we combined our III-V transfer process based on NiSi bonding to realize, for the first time, InGaAs FinFET devices on Si with sub-20 nm channel width. This first batch of undoped InGaAs FinFETs showed excellent features in terms of low off-current, steep turn-on characteristics, and immunity to SCE, which are comparable to those of well-developed III-V FinFETs on InP substrates. A key advantage of our fabrication workflow is that it’s fully compatible with CMOS technology and could straightforwardly be deployed to allow multiple substrate reuses. The heterogeneous integration technology and device fabrication process are schematically illustrated in Figure 1. The starting substrate comprises of an undoped InGaAs layer (50 nm) grown on bulk semi-insulating InP wafer. Dielectric stacks of HfO2 and SiO2 were then deposited, followed by the e-beam evaporated Ti/Ni metal layers. Our bonding scheme utilizes the NiSi interface alloyed at temperatures as low as 300 °C to fuse the InGaAs/InP wafer to a Si host substrate. Detailed transmission electron microscopy (TEM) and electrical characterization confirmed the formation of nickel monosilicide interfacial bonding layers. For the purpose of demonstrating this new bonding approach, the bulk InP substrate was thinned by mechanical lapping and completely removed by HCl:H2O (3:1) wet etching. Note, however, that sacrificial layers such as thin AlAs/InAlP layers could be simply included in the initial heterostructures growth on the donor InP substrate to implement ELO and reuse of the InP substrate for subsequent epitaxial growths and transfers.

Transfer yield of the process was greater than 90% even for the largest 2 cm × 2 cm InGaAs area we tested, where only the edges were found to be prone to exfoliation limited by edge effects on the cut 2 cm × 2 cm samples. After transferring the InGaAs layers onto the Si substrate, device fabrication of advanced III-V electronic and optoelectronic devices becomes feasible which is demonstrated here for the case of the first InGaAs FinFETs on Si. The 3D device fabrication proceeded by defining arrays of 5 or 10 parallel InGaAs “Fins” by electron-beam lithography and a Cl2/N2 inductively coupled plasma etching of the InGaAs Fins. After the InGaAs dry etch, a cyclic O2 plasma, dilute buffer oxide etch (BOE) and deionized water cleaning were performed to remove the e-beam resist and etch mask (see experimental methods) and reduce the surface roughness of the InGaAs Fins. During this process, the height of the Fins is reduced from that of the original InGaAs layer thickness of 50 nm to ≈35 nm. These devices feature Ni S/D contact region, a 5 nm HfO2 gate dielectric deposited by atomic layer deposition (ALD) with an equivalent oxide thickness (EOT, the thickness of an SiO2 layer that would result in the same capacitance as the high-k material) of 1.11 nm, and an accurately aligned Ni gate electrode. State of the art wet processing yielded an undoped InGaAs Fin well over 90% with a threshold voltage of ~350 mV, drain current of 3.5 μA/μm, and subthreshold swing of 60 mV/decade. For an in-depth discussion of the fabrication process, refer to the “Experimental Methods” section of the paper.
chemical treatments and in-situ ALD plasma treatments with extensive and systematic capacitance-voltage characterization of metal-insulator-InAs/InGaAs structures were conducted to reduce the leakage current and improve the interface quality (Experimental Section and Supporting Information, procedure and Figures S1–S7). The resulting devices consisted in arrays of InGaAs FinFETs with 5 or 10 parallel Fins with channel length $L_{ch}$ varying from 35 nm to 950 nm.

Illustrated in Figure 2a is the cross-sectional TEM image (along A-A’ cut in Figure 1) of the whole structure of three finished InGaAs FinFET devices on a stack of HfO$_2$/SiO$_2$/NiSi/Si. Due to the formation of NiSi ($\approx$4.7% lattice mismatched to Si), the NiSi layer contains multiple grains and stacking faults, the presence of which is unavoidable but does not compromise the integrity of the III-V layers atop, or possible CMOS devices in the vicinity. It is important to note that no defective regions have been identified by TEM in the Si sample itself after the NiSi bonding procedure. The (001) Si/NiSi interface is typically non-flat and can become smooth and flat with the use of (111) surfaces and thin Ni layers. Figure 2b depicts a high resolution TEM (HRTEM) image of a tri-gated InGaAs FinFET (13 nm top width and 36 nm slanted sidewall lengths) retaining single crystal quality (Figure 2c) with 5 nm conformally covered HfO$_2$ gate dielectric on the Fin top and sidewalls. The bottom HfO$_2$ layer is used as a post-anneal HF-etch stop layer in our process and displays a polycrystalline structure (Figure 2d) due to NiSi high temperature bonding process. The HfO$_2$ crystallization led to electrical shorts in devices made without SiO$_2$ layers where current paths between the FET device leads on top and through the HfO$_2$ to the NiSi layer and then to the electrode leads were created. The leakage through HfO$_2$ necessitated an insertion of a SiO$_2$ layer underneath it for electrical isolation from the underlying NiSi layer. Cross-sectional TEM images of five final FinFET devices in Figure 2e–i demonstrate FinFET perimeter variation of 60 nm, 85 nm, 100 nm, 130 nm and 150 nm with slanted sidewalls due to erosion of the etching mask during the plasma etch.

Transistor performance of all devices was measured after 5 min rapid thermal annealing at 200 °C in forming gas (H$_2$/N$_2$ mixture) ambient. The transfer characteristics of a representative device with 10 InGaAs FinFET channels with $L_{ch} = 390$ nm, $P = 60$ nm are shown in Figure 3, where the current is normalized by the perimeter of the Fins, measured by TEM, and by the number of channels. This device exhibited an $I_{on} = 18$ mA mm$^{-1}$ at $V_{DS} = 0.5$ V and $V_{GS-V_T} = 0.5$ V, where the threshold voltage $V_T = 0.78$ V was extracted by the linear extrapolation of the $I_{DS-V_{GS}}$ at the maximum slope (peak transconductance, $g_m$). The on-current of these devices was limited by the series contact resistance due to the Schottky barrier nature between the Ni S/D contacts and the undoped InGaAs, which was also responsible for the increment of $I_{DS}$ in the negative $V_{GS}$ region from hole transport. An inverse subthreshold slope (SS$^-$) of 165 mV dec$^{-1}$ and a maximum current sweep ratio ($I_{max}/I_{low}$) of $3.2 \times 10^6$ at $V_{DS} = 0.05$ V were observed. The interface quality for these FinFET devices on Si compares well to those on InP and lags behind some others with Al$_2$O$_3$ interfacial...
gate dielectric layers.\[16b\] Further optimization of the HfO\(_2\)/InGaAs interface and reduced Fin widths would result in lower SS\(^{-1}\). The gate leakage was as low as 10\(^{-6}\) pA/\(\mu\)m for all the \(V_{DS}\) biases used in this study up to \(V_{DS} = 1\) V.

Figure 4 compares the normalized output and transfer curves for devices with the same perimeter of \(P = 60\) nm and with various channel lengths of \(L_{ch} = 890\) nm, 690 nm, 500 nm, and 160 nm. Shorter channel length devices exhibited higher drive currents (Figure 4d), more severe SCE, larger off currents and worse SS\(^{-1}\) (Figure 4h), whereas the on-state performance did not scale linearly with \(L_{ch}\) due to the large S/D contact resistance for these undoped InGaAs devices. With decreasing Fin perimeter for the same channel length (\(L_{ch} = 690\) nm), \(I_{DS}\) enhancement was observed as shown in Figure 5a–d. Such enhancement has been attributed to increased quantum confinement and increased carrier mobility resulting from reduced interface scattering.\[23\] In addition, the transfer curves indicated an improved SS\(^{-1}\) and on/off characteristics with decreasing perimeter as shown in Figure 5e–h.

The scaling metrics for InGaAs FinFETs on insulator-on-Si with various gate lengths and Fin widths are summarized in Figure 6, including SS\(^{-1}\), threshold voltage, \(V_T\), and drain induced barrier lowering, DIBL (resulting in \(V_T\) reduction with a \(V_{DS}\) increase in transistors with short channel lengths). Figure 6a shows that SS\(^{-1}\) was degraded for all \(V_{DS}\) values with reducing \(L_{ch}\) from 900 nm to 150 nm, and a higher SS\(^{-1}\) was observed for higher \(V_{DS}\), as expected, due to SCE (Figure 6a). The lowest SS\(^{-1}\) of 150 mV dec\(^{-1}\) was obtained at \(V_{DS} = 0.01\) V for the device with \(L_{ch} = 390\) nm and \(P = 60\) nm. For a variable perimeter with same \(L_{ch} = 450\) nm depicted in Figure 6b, SS\(^{-1}\) improved dramatically from 420 mV dec\(^{-1}\) to 160 mV dec\(^{-1}\) with decreasing perimeter from 150 nm to 60 nm at \(V_{DS} = 0.01\) V. Channel length scaling that is immune to SCE can be achieved by narrowing the Fin width. Figure 6c exhibits the influence of \(L_{ch}\) (35 nm to 700 nm) and perimeter (60 nm to 150 nm) change on \(V_T\), where \(V_T\) was obtained by the linear extrapolation method discussed above at \(V_{DS} = 0.5\) V. A slight increase of \(V_T\) with \(L_{ch}\) was observed for all perimeters from 60 nm to 150 nm, but \(V_T\) was found to increase dramatically with decreasing perimeter, pulling the device into a desired enhancement mode operation. With \(L_{ch}\) of \(= 400\) nm, and as \(P\) increases from 85 nm to 150 nm, a negative shift in \(V_T\) of 1.4 V was required to deplete the channel. This trend was confirmed for perimeters of \(P = 85\) nm, 100 nm, 130 nm, and 150 nm, and Silvaco Atlas 3D simulations provided supporting evidence for this ‘body-thickness’ effect (Figure S8, Supporting Information). For \(P = 60\) nm, \(V_T\) was found to be close to that of \(P = 100\) nm, both of which had a circular cross-sectional perimeter at the top portion of the Fin, compared to all other Fins. However, process variations cannot be excluded to interpret this surprising similarity in \(V_T\). Figure 6d shows that the...
DIBL increased with shorter $L_{\text{ch}}$ due to SCE but this effect was suppressed for smaller $P$. The DIBL degrades from 8 mV/V to 410 mV/V when $L_{\text{ch}}$ was reduced from 925 nm to 180 nm for a fixed $P$ of 130 nm. On the other hand, when $P$ was decreased from 130 nm to 60 nm with a fixed $L_{\text{ch}}$ of 400 nm, DIBL decreased from 265 mV/V to 115 mV/V. These metrics indicate...

Figure 5. Influence of FinFET perimeter ($P$) on device performance for a fixed channel length, $L_{\text{ch}} = 690$ nm. Illustrated are a–d) the output and e–f) transfer characteristics for devices with $P = 150$ nm, 130 nm, 100 nm, and 60 nm. e–h) Larger on-currents and better transfer properties are measured for smaller perimeters. Scale bars of all inset SEM images are 1 µm.

Figure 6. Evaluation of short channel effects in InGaAs FinFETs on Si. a) $SS^{-1}$ scaling properties as a function of channel length, $L_{\text{ch}}$, for $P = 60$ nm and for different $V_{\text{DS}}$ biases. b) $SS^{-1}$ scaling properties as a function of perimeter $P$ for a fixed $L_{\text{ch}} = 450$ nm and for different $V_{\text{DS}}$ biases. c) Scaling properties of the threshold voltage, $V_T$, as a function of $L_{\text{ch}}$ and for different $P$ demonstrating a stronger effect of the $P$ on $V_T$ with relatively insignificant change with $L_{\text{ch}}$, for the same $P$. d) DIBL scaling properties as a function of $L_{\text{ch}}$ down to 35 nm and with various $P$. 
that devices with longer $L_d$ and smaller $P$ will have improved SS$^{-1}$, DIBL, and $I_{on}/I_{off}$ without suffering much from SCE. More sophisticated processing, including retrograde channel doping,$^{[22]}$ decreasing EOT$^{[24]}$ (Supporting Information), gate first, self-aligned $S/D^{[21a]}$ and utilizing gate all around 3D structure,$^{[16b]}$ could help in improving the gate control of the channel and further suppressing the SCE.

3. Conclusion

We demonstrated a novel heterogeneous integration technology for thin III–V layers to Si by NiSi formation for advanced active electronic and optoelectronic III-V devices on Si. We reported the fabrication of the first InGaAs FinFETs on insulator on Si by such fab-compatible process using this new integration platform, and systematically investigated the scaling metrics of InGaAs FinFETs with high-k dielectric and metal gate and provided insights into improving their performance. This demonstration highlights the potential of high performance InGaAs FinFETs on Si for ultimately scaled III-V logic technology, and paves the way for incorporating a variety of III-V electronic and optoelectronic devices on a Si CMOS platform.

4. Experimental Section

Wafer Bonding: The undoped 50 nm thick InGaAs was grown on semi-insulating InP (001) substrate by molecular beam epitaxy at Intellegent Epitaxy Inc. In preparation for wafer bonding, the native oxide layer on the surface of the InGaAs was removed by a diluted BOE. An HfO$_2$ layer (15 nm) was deposited by atomic layer deposition (H$_2$O source kept at 20 °C and Tetraakis(dimethylamino)hafnium (TDMAH) source kept at 75 °C) atop the InGaAs/InP sample at 200 °C, followed by a SiO$_2$ dielectric layer (200 nm) deposition by plasma enhanced chemical vapor deposition for electrical isolation. A Ti adhesion layer (15 nm) followed by a Ni bonding layer (100 nm) were deposited by e-beam evaporation. This material stack (Ni/Ti/SiO$_2$/HfO$_2$/InGaAs/InP) was then brought in contact with a Si sample with similar area (typically 1 cm x 1 cm) that is cleaned by “Piranha” solution (H$_2$SO$_4$ : H$_2$O$_2$ 3:1) and HF. 400 °C rapid thermal annealing in forming gas ambient (N$_2$ : H$_2$ 85% : 15%) for 10 mins was then performed on this stack with pressure applied on both sides by a silver metal paperclip, which led to the formation of a bonding NiSi interface between the III-V stack and Si.

InGaAs FinFET Fabrication: After bonding of the III–V stack to Si, the InP wafer was thinned down by mechanical lapping on a polishing cloth to a thickness of ~70–100 µm which was then selectively etched by HCl:H$_2$O (3:1) leaving the 50 nm InGaAs atop HfO$_2$/SiO$_2$/NiSi/Ti/Si. The device fabrication utilized a 100 Kev electron-beam lithography (JEOL JBX-6300FS) system to pattern arrays of 5 or 10 parallel InGaAs Fins, whose height was also reduced to 35 nm. S/D contact removal the HSQ etch mask and reducing the surface roughness of the InGaAs Fins. After the InGaAs Fin etch, a cyclic O$_2$ plasma, and HF:DI water cleaning was performed to remove the HSQ etch mask and reduce the surface roughness of the InGaAs Fins, whose height was also reduced to 35 nm. S/D contact regions were then patterned, followed by a BOE dip, and transferred immediately to an e-beam evaporator system for Ni deposition followed by lift-off. The native oxide of the InGaAs Fins was etched in diluted BOE, rinsed in deionized water, followed by 10 min 10% (NH$_4$)$_2$S dip for surface passivation.$^{[35]}$ A 5 nm thick HfO$_2$ layer was then deposited by ALD at 200 °C and were characterized versus different deposition and surface treatment parameters to provide the best interface quality (Supporting information). Overlapping gate electrodes were finally aligned and patterned by electron beam lithography, followed by Ni metal deposition. The resulting devices consisted in arrays of InGaAs FinFETs with 5 or 10 parallel channels with channel length $L_d$, varying from 35 nm to 950 nm. The sample fabrication and morphology were characterized by an FEI SEM.

Transmission Electron Microscopy: The cross-sectional TEM samples were prepared using an FEI Nova 600 Nanolab dual beam system, consisting of SEM and focused ion beam (FIB). Pt was in-situ deposited along the targeting lamellae for sample surface protection purposes. Ion beam (gallium) was utilized to cut two trenches from each side, followed by mounting the sample on the nanomanipulator needle by Pt deposition. The sample was then lifted out and mounted on a TEM grid after a “U” cut on the cross-section. Fine cleaning further thinned down and polished the lamella in to an electron transparent thickness. HRTEM characterization was conducted in a 300 keV FEI Tecnai F30 system.

Electrical Characterization: Current-voltage ($I$–$V$) and capacitance voltage ($C$–$V$) characteristics were measured by an Agilent B1500A semiconductor parameter analyzer in fast scan mode. No significant influence of the integration time on the IV characteristics was observed.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The authors are grateful to Dr. John Nogan and Mrs. Denise Webb of Sandia National Laboratories for their assistance with facility operation and useful discussions and for the insightful comments from Prof. S.S. Lau at UC San Diego. This work was performed, in part, at the Center for Integrated Nanotechnologies, an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science. Los Alamos National Laboratory, an affirmative action equal opportunity employer, is operated by Los Alamos National Security, LLC, for the National Nuclear Security Administration of the U.S. Department of Energy under contract DE-AC52-06NA25396. X.D., B.M.N., Y.H., and S.A.D. acknowledge support from a faculty start-up fund at UC San Diego (ECER117) and Laboratory Directed Research and Development (LDRD) program at Los Alamos National Laboratory project numbers (20120747PRD1, 20100061PRD2), and partial support from the National Science Foundation, Division of Materials Research, DMR-Award No. 0902277. X.D. and C.S. acknowledge support from NTU-NAP start-up grant (M4080511).

Received: January 12, 2014
Revised: March 4, 2014
Published online:
