Fabrication process and failure analysis for robust quantum dots in silicon

J. P. Dodson, Nathan Holman, Brandur Thorgrimsson, Samuel F. Neyens, E. R. MacQuarrie, Thomas McJunkin, Ryan H. Foote, L. F. Edge, S. N. Coppersmith, and M. A. Eriksson

1Department of Physics, University of Wisconsin-Madison, Madison, WI 53706, USA
2HRL Laboratories, LLC, 3011 Malibu Canyon Road, Malibu, CA 90265, USA
3University of New South Wales, Sydney, Australia

We present an improved fabrication process for overlapping aluminum gate quantum dot devices on Si/SiGe heterostructures that incorporates low-temperature inter-gate oxidation, thermal annealing of gate oxide, on-chip electrostatic discharge (ESD) protection, and an optimized interconnect process for thermal budget considerations. This process reduces gate-to-gate leakage, damage from ESD, dewetting of aluminum, and formation of undesired alloys in device interconnects. Additionally, cross-sectional scanning transmission electron microscopy (STEM) images elucidate gate electrode morphology in the active region as device geometry is varied. We show that overlapping aluminum gate layers homogeneously conform to the topology beneath them, independent of gate geometry, and identify critical dimensions in the gate geometry where pattern transfer becomes non-ideal, causing device failure.

I. INTRODUCTION

Developing a suitable physical system for quantum computation has received much attention in the past two decades. Since Loss and Divencenzo’s proposal [1], significant progress has been made using spins in solid-state systems as quantum bits (qubits). Coherent control of semiconductor quantum dots using spin degrees of freedom was first demonstrated in GaAs/Al_{0.3}Ga_{0.7}As heterostructures [2, 3]. This particular heterostructure found initial success due to the small electron effective mass in GaAs and depletion mode operation of devices. This allows for large, single layer gate geometries to tune devices into the few-electron regime [4]. Although fabrication and characterization of one-qubit devices in GaAs has become routine [2, 3, 5–7], short coherence times of spin-qubits due to the presence of nuclear spins [8] make it difficult to achieve fidelities necessary for fault-tolerant operation [9].

Silicon-based approaches have significantly improved qubit performance in solid-state spin systems in part due to the spinless nucleus of ^{28}Si. Experiments using isotopically purified ^{28}Si have shown average single-qubit control fidelity in excess of 99.9% [10]. However, in Si, the lithographic demands are more stringent because of the larger electron effective mass. Different gate designs have been explored, including open geometries, which use global top gate(s) and several depletion gates to form each quantum dot [11, 12] and tight geometries which use linear, overlapping gates with dedicated accumulation and depletion electrodes for each quantum dot [13, 14]. While both general designs have generated two-qubit devices in Si/SiGe [15–17] and Si-MOS [18, 19], the overlapping gate architecture has clear advantages in scaling to larger systems. A 9-dot array has already been demonstrated [14] and similar architectures can be scaled into much larger arrays. Additionally, quantum dots form in predictable locations with tunnel couplings that can be well-controlled using a single gate [20, 21]. While there are benefits in choosing a linear, overlapping gate architecture, developing a high yield fabrication process is challenging.

In this article, we investigate many yield limiting steps in the fabrication of linear, overlapping aluminum gate quantum dot devices, showing failure analysis of critical interfaces and providing remedies for particular problems experienced in typical fabrication processes. We present results on three main topics: low-temperature oxidation of inter-gate aluminum oxide (AlO\(_x\)), cross-sectional scanning transmission electron microscopy (STEM) analysis of overlapping Al gate geometries, and characterization of interconnects between the device bond pads and active region—the local region surrounding quantum dots. Four low-temperature oxidation techniques are compared for enhancement of the native AlO\(_x\) that electrically isolates subsequent gate electrode layers from each other. STEM analysis investigates test structures with varying dot-to-dot pitch, characterizing how different gate geometries affect gate electrode morphology and the filling of barrier gates in gaps between plunger gates. For the interconnects, we optimize the process flow to allow for thermal annealing of the Al\(_2\)O\(_3\) grown by atomic layer deposition (ALD), incorporation of the presented low-temperature oxidation techniques, and integration of on-chip electrostatic discharge (ESD) protection.

II. FABRICATION METHODS

Layer-by-layer fabrication of overlapping Al gate quantum dot devices is shown in Figure 1. A top down scanning electron microscope (SEM) image of each Al gate layer is shown in Figure 1(a-c) and a completed device active region with all three layers is shown in Figure 1(d). The device consists of three quantum dots with four barrier gates on the bottom side and a single integrated
charge sensing dot on the top side. This particular geometry is shown because it is a unit cell for an architecture that can be linearly tiled [14]. All layers are patterned using an Elionix ELS G-100 electron beam writer on PMMA 495 A4 resist. The Al gate metal is deposited using a Lesker PVD 75 electron-beam evaporator. All electron-beam lithography (EBL) steps using a PMMA resist stack are developed in 3:1 IPA:MIBK and lifted off using Remover PG heated to 75°C for 1 hour, followed by a solvent clean. After lift-off, the Al gates on the first two layers are further oxidized using a plasma ash technique that will be described in detail below. A more thorough fabrication process is described in the Appendix.

Each set of gates takes on a specific role in manipulating the chemical potential of the two-dimensional electron gas (2DEG). The first layer, shown in Figure 1(a), acts to screen stray electromagnetic fields from the second and third layers. The screening gates are used to deplete regions beneath them, defining a long, thin channel in the 2DEG that can be accumulated/depleted by gates from subsequent layers. A 150 nm wide central screening gate is patterned to prevent current injection from the top side to the bottom side. The second layer, shown in Figure 1(b), consists of accumulation gates, including reservoir gates for Fermi level control of electron reservoirs, and plunger gates for tuning electron occupation within quantum dots. Figure 1(c) shows the third layer, which patterns barrier gates designed to deplete the 2DEG and tune the tunnel coupling into and out of quantum dots. The final device with all three layers is shown in Figure 1(d). The width of all plunger/barrier gates is increased to 100 nm before crossing the edge of the screening gates in order to minimize the chance of step coverage failure. In this particular geometry, plunger gates are nominally 70 nm wide with a 120 nm pitch, and barrier gates are 60 nm wide, filling a 50 nm gap. Figure 1(e, f) show the electrostatic control in the few-electron regime between each pair of adjacent quantum dots P1/P2 and P2/P3, respectively, as measured by an integrated charge sensing dot beneath M1. High resolution plots were taken over a 24 hour period to demonstrate stability in the few-electron regime.

III. LOW-TEMPERATURE INTER-GATE OXIDATION

Gate layers are electrically isolated via the AlOₓ that grows natively on the Al gate electrodes. This allows for the omission of blanket dielectric films such as ALD-grown Al₂O₃ typically present between gate layers, which may be causing increased charge noise in Si/SiGe quantum dot devices [22]. Removal of grown dielectrics for electrical isolation comes at a cost though—one must rely on the AlOₓ that grows natively on evaporated Al, which is reported to grow between 1.6–3.0 nm by a variety of techniques [23–26]. This makes high yield fabrication of devices difficult due to gate-to-gate leakage and damage from ESD.

In order to determine if this native oxide is sufficient to prevent leakage, we compare four different techniques used to oxidize Al gates: native oxidation (NO), thermal annealing at 250°C (TO), plasma ashing (PA), and UV-ozone treatment (UV). All samples are natively oxidized at standard temperature and pressure before a 15 minute treatment by each oxidation method. For TO, the anneal is at 250°C and 45% humidity. For PA, the plasma asher used is a YES R3 Downstream Plasma Cleaner, at a pres-
sure of 5 torr with 80 sccm O₂ using a power of 250 W. For UV, a Samco UV-1 model is used with an oxygen flow rate of 0.5 l min⁻¹, giving an ozone concentration of ∼6 g m⁻³ with its substrate platen heated to 250°C.

To characterize how each oxidation treatment affects the electrical isolation properties of Al gate electrodes, we fabricate two-layer test structures designed to replicate the plunger and barrier gate layers of the device shown in Figure 1. The first layer is oxidized using one of the four methods and then the breakdown voltage is measured. Devices are fabricated on [100] Si wafers, electrically isolated from the Si by 100 nm of ALD-grown Al₂O₃. The two layers are patterned using EBL. 30 and 50 nm of Al are deposited for the first and second layer, respectively. The structures have an overlap area of 1 μm x 0.1 μm, similar to the gate overlap in Figure 1(d). Devices are tested cryogenically at 2 K by increasing the differential voltage between electrode and counter-electrode pairs until breakdown is observed, defined here to be when 100 pA of gate-to-gate leakage is measured. Devices are inspected after measurement in an SEM to ensure they have not been destroyed by ESD.

Figure 2(a) shows the results of the breakdown test. For each oxidation method, several samples are measured to determine an average breakdown voltage, shown as a solid data point. Adjacent to solid data points are the breakdown voltage of all devices measured for each method, shown as hollow, semi-transparent points. The average breakdown and standard deviations are summarized in Table I.

The breakdown voltages observed in PA and UV are significantly above the estimated maximum differential voltage needed for overlapping Al gate devices in Si/SiGe (∼1–2 V), whereas it is only modestly higher for NO and TO. The distinct increase in breakdown voltage of the PA and UV techniques suggests increased AlOₓ thickness and/or higher quality AlOₓ. To isolate these variables, we analyze bulk Al films using x-ray photoelectron spectroscopy (XPS) to determine oxide thickness. The XPS spectra are shown in Figure 2(b). The AlOₓ thickness can be extracted from the relative intensities of the oxidic/metallic Al 2p peaks, as detailed in [27]. The spectra shown in Figure 2(b) are a subset of a larger dataset taken to determine the average thickness and fluctuations in thickness for each method. Six XPS spectra are taken at different positions on the bulk film for each oxidation method to obtain statistical fluctuations in the AlOₓ thickness. Figure 2(c) shows the results from all measured samples and the values are summarized in Table I. The thickness for PA, UV and TO all exceed NO. PA displays the fastest oxidation rate. The increase in breakdown fields of PA and UV compared to NO are not statistically significant, therefore the higher breakdown voltage is primarily attributed to the increase in thickness. The oxidation rate of Al for both UV and PA follows a d ∼ √t dependence [24, 25], where d is thickness and t is time, which can be used to further increase the AlOₓ thickness if desired. PA is implemented in the fabrication of the triple-dot device shown in Figure 1.

Table I. Electrical characterization summary of low-temperature Al oxidation methods. Oxide thickness was determined from x-ray photoelectron spectroscopy spectra. The breakdown voltage (BDV) is defined here to be when >100 pA of current is measured between electrode and counter-electrode pairs. Column four reports the ratio of the first two columns, the dielectric strength.

| Method | Thickness (nm) | BDV (V) | Diel. Str. (MV cm⁻¹) |
|--------|---------------|---------|---------------------|
| PA     | 4.12 ± 0.12   | 6.36 ± 1.12 | 1.54 ± 0.27        |
| UV     | 3.49 ± 0.09   | 5.23 ± 0.31 | 1.50 ± 0.10        |
| TO     | 2.88 ± 0.04   | 3.65 ± 0.59 | 1.27 ± 0.21        |
| NO     | 2.00 ± 0.07   | 2.91 ± 0.24 | 1.46 ± 0.13        |
Figure 3. Failure analysis for varying gate widths. (a) False-colored scanning transmission electron microscopy (STEM) image of an overlapping Al gate structure. Al plunger gates (blue) are deposited nominally 50 nm thick and 70 nm wide on a Si substrate with gate-to-gate pitches varying from 90–150 nm. Al barrier gates (red) are deposited subsequently, nominally 65 nm thick with widths varying from 40–100 nm. The test structure is capped with 20 nm of ALD-grown Al$_2$O$_3$ (gray) and a protective platinum layer (black). (b) Schematic cross-section of imaged overlapping Al gate structure. The sidewalls are assumed to be $\sim 60^\circ$ as is observed on average in the STEM image. (c) Zoom in of Figure 3(a). The 70 nm gap between plunger gates is filled homogeneously by the barrier gate. (d-f) Schematic showing before (d), during (e) and after (f) the Al e-beam evaporation. The sloped sidewalls observed for Al gates is due to accumulation of Al on PMMA sidewalls during evaporation [23].

ALD-grown Al$_2$O$_3$ as a gate dielectric [22, 28].

IV. STEM FAILURE ANALYSIS

To further investigate potential failure modes in overlapping Al gate devices, we fabricate test structures using PA between gate layers and take cross-sectional STEM images (Figure 3). Two-layer test structures are fabricated to investigate filling of the gaps between plunger gates (first layer) by barrier gates (second layer) for varying gate widths.

The test structures are fabricated using the same procedure described for the device shown in Figure 1(a-d), except the first layer is omitted. The gate geometry can is shown in Figure 3(a), where the plunger gate layer consists of eight 70 nm wide, 50 nm thick gate electrodes with gate-to-gate pitches varying from 90–150 nm (increasing right to left) in steps of 10 nm. This leaves nominal gaps of widths 20–80 nm for the barrier gates to fill. The barrier gates are evaporated 20 nm wider than each gap, ranging from 40–100 nm (increasing right to left). The thickness of the barrier gate layer is nominally 65 nm. A schematic of the expected cross-section is shown in Figure 3(b).

Several striking features are revealed from the STEM images. The sidewall slope of the plunger and barrier gate electrodes is found to be between 45–60$^\circ$. This is consistent with AFM profiling shown in Ref. [29] and can be explained by the process illustrated in Figure 3(d-f). During evaporation, Al accumulation narrows the opening of the PMMA mask. In extreme cases, the opening closes completely, leaving gate electrodes thinner than intended. This symptom is visible on the far right of Figure 3(a), and can lead to yield problems associated with step coverage failure.

On the far right of Figure 3(a), the plunger gates themselves are deformed and reduced in thickness. This is attributed to resist wall collapse [30], since the thickness is skewed to one side. Top-down SEM images (not shown) are consistent with such resist collapse. Resist wall collapse is an issue in these devices because the barrier gates are relatively thick (65 nm), to guarantee high-yield connection over the many underlying gates in this device. As a consequence, the PMMA resist is also chosen to be thick (180 nm), for high-yield liftoff.

The barrier gates fill and contact the underlying surface across all regimes, independent of the gap width. The only case in this test structure where a barrier gate is not in contact with the substrate is on the far right side, where the intended gap of 20 nm completely closed off. In the right regime, the barrier gate layer fails for a different reason than the plunger gate layer. The barriers do not hold their intended thickness and shape, due to the process illustrated in Figure 3(d-f). Given the 45–60$^\circ$ sloped sidewalls, the thickness of the gate cannot reliably exceed the width using this recipe. A way to design around this is to pattern barrier gates as wide as possible when devices contain small gaps between plunger gates.

V. DEVICE INTERCONNECTS

Another challenge in fabricating quantum dots in Si/SiGe heterostructures is the design considerations needed for developing on-chip interconnects between
bond pads and the active region. High temperature processes are desirable at different stages of device fabrication, but often times they can cause failure of interconnects. Below, we discuss critical interfaces affecting the thermal budget at various stages of fabrication, how to design around the thermal budget, and integration of on-chip ESD protection.

For the interconnect design, we etch a mesa between the bond pads and active region, shown in Figure 4(b), preventing gate-to-ohmic leakage in the event that wire bonds punch through the substrate into the 2DEG. The yield of devices can be severely limited by damage due to ESD after the fabrication of the active region. To remedy this, on-chip ESD protection is implemented, shown in Figure 4(a). Shorting wires are patterned before fabrication of the active region in the same lithographic step as bond pads, preventing build-up of charge between gates. An equipotential for all gates can be maintained through device packaging by wire bonding, grounding through a printed circuit board, and physically scribing away the shorting wires on-chip afterwards. Alternatively, the leads can be electrically disconnected after fabrication using an etch step, however, this does not protect the device during packaging.

Consideration of the thermal budget is important when determining a process flow for quantum dot devices. Significant interdiffusion of the Si/SiGe interface at the quantum well can occur above 800°C [31], lowering the valley splitting [32]. This makes growth of high quality SiO$_2$ [33, 34] on Si/SiGe heterostructures difficult, so instead ALD is used to grow amorphous Al$_2$O$_3$ or HfO$_2$ for gate-to-2DEG isolation. After the field/gate oxide is grown, thermal annealing can be used to reduce interface trapped charge [35], which has been shown to reduce threshold voltages and increase transconductance in Si-MOS quantum dots [13].

The presence of interfaces (Figure 4(c), I-1 and Figure 4(d), I-2) further limits the thermal budget of devices. We also note that the presence of thin Al films imposes its own set of thermal constraints, including dewetting and void formation, shown to occur at 400°C [23] and between 300–500°C [36], respectively. To maximize the thermal budget of devices, we choose palladium instead of gold for interconnect metallization. Au is often used, but formation of AuAl$_2$ (purple plague), a non-conductive alloy, is observed in thin Au-Al films at temperatures as low as 217°C on a minutes time scale [37]. Additionally, at the ohmic metallization site, the Au/Si interface experiences an interstitial diffusion process that affects thin film electrode morphology at temperatures as low as 200°C [38]. We observe a similar process when annealing a Au/Ti/Si structure analogous to I-2 at 250°C for 30 minutes. When using Pd, the thermal budget is found to increase. Formation of Pd-Al alloys in thin films near 300°C [39] can cause electrical discontinuities, which we verified with a 30 minute anneal at 300°C on a 30/17/3 nm Al/Pd/Ti gate stack. Additionally, we observe diffusion processes at 400°C occurring at the Pd/Ti/Si interface, affecting ohmic gate morphology. The consequence of these critical temperatures is that they cannot be exceeded by processes such as post-metallization anneals or inter-gate oxidation with the expectation of high yield. However, the UV and PA methods presented in this paper do not exceed any of these critical temperatures when using Pd as an interconnect/ohmic gate metal (Figure 4, (d) Schematic cross-section (not to scale) of a device interconnect at the ohmic site. An interface (I-2) between Ti/Pd and the P$^+$ doped Si/SiGe heterostructure forms. Diffusion processes are observed to begin at ~400°C, which can affect the interconnect morphology and electrical connectivity.}

![Figure 4. Interconnect and ESD protection architecture.](image)
inset), and thus are good choices for oxidation of inter-
gate oxide.

VI. CONCLUSION

We have demonstrated an improved fabrication process and performed failure analysis on many critical interfaces in overlapping Al gate quantum dot devices. The main takeaway from the AlO₃ characterization is that the temperatures induced by PA and UV are compatible with the overlapping Al gate quantum dot device fabrication presented, mitigating failure due to gate-to-gate leakage and ESD. Optimization of oxidation techniques may improve oxide quality and allow further tuning of the AlO₃ thickness. For UV oxidation of Al, relative humidity, time of oxidation, temperature, and partial pressure of oxygen all affect the growth rate [25, 26, 40]. For PA, time of oxidation, excitation frequency, power, and partial pressure of oxygen affect the growth rate [24, 41].

We also fabricated overlapping Al gate two-layer structures and used cross-sectional STEM images to analyze failure modes of different gate geometries. The side walls of the plungers/barriers were found to be between 45–60°, creating potential issues with step coverage near the active region. This can often be designed around by maximizing the barrier gate widths when devices have small gaps between plunger gates. We also identified failure modes associated with critical dimensions in device geometries. Gates with <40 nm gaps between plungers showed abnormalities due to resist wall collapse and thinner than intended deposition thickness.

Finally, we showed an interconnect fabrication process that implements on-chip ESD protection and identified critical temperatures that can result in electrical discontinuities at different stages of device fabrication. By designing a fabrication process compatible with these temperatures, field/gate oxides may still be annealed at 450°C in forming gas before metal deposition, and inter-gate oxide in Al devices can be further oxidized using UV or PA. The result of a process with these changes implemented was shown in Figure 1(e, f), where the quantum dots were notably stable and all gate electrodes worked as intended.

ACKNOWLEDGMENTS

We thank Jason Petta and Lieven Vandersypen for helpful discussions. Research was sponsored in part by the Army Research Office (ARO) under Grant Numbers W911NF-17-1-0274 and by the Vannevar Bush Faculty Fellowship program under ONR grant number N00014-15-1-0029. We acknowledge the use of facilities supported by NSF through the UW-Madison MRSEC (DMR-1720415) and the NSF MRI program (DMR-1625348). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office (ARO), or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

Appendix: Detailed Fabrication Process

The fabrication process flow for devices integrating low-temperature oxidation of Al gates, on-chip ESD protection, thermal annealing of the field/gate oxide, and the use of Pd as interconnect metal is discussed below. Patterning for all photolithography steps is done using a Nikon NSR-20058A i-line stepper. All photolithography steps precede EBL steps and are completed on a 3" Si/SiGe wafer before it is diced into chips for EBL. A base mesa structure, shown in Figure 4(a, b), accommodates a range of device geometries with up to 40 gates. All Al₂O₃ etch steps use 20:1 BOE as an etchant.

Device fabrication begins with a mesa etch using reactive ion etching (RIE) with CHF₃ as the process gas. Before ion implantation, a 10 nm screening oxide (Al₂O₃) is grown using ALD. This helps prevent cross-linked resist from contaminating the substrate when stripping the resist mask post-ion implantation. 31P⁺ is implanted at 25 keV with a density of 5×10¹⁵ ions cm⁻² at a 7° tilt to electrically connect to a 40 nm deep Si quantum well. The implanted region is then annealed at 700°C for 15 seconds in forming gas using a rapid thermal annealer. 15 nm of field oxide and and 5 nm of gate oxide (Al₂O₃) are grown subsequently using ALD. Between oxide depositions, a 15 µm × 15 µm terrace is etched in the field oxide to reduce bulk oxide in the active region. Each layer is annealed individually using a forming gas anneal (FGA) at 450°C. Two anneals, one after each layer, is preferred over annealing both at the same time due to blistering of ALD Al₂O₃ [42, 43].

Metalization of devices begins with a two step process to etch and metallize the ohmic contacts. Positive photoresist with an HMDS adhesion layer is used to prevent undercutting during the etch and a negative resist is used for ohmic metallization. Interconnect jumper pads are then deposited using a 3/17 nm Ti/Pd metal stack. This total thickness must be sufficiently thin compared to the first Al gate layer thickness to prevent step coverage failure. Bond pads and the ESD protection wiring are deposited together using a 20/180 nm Ti/Pd metal stack.

After wafer-level photolithography steps have been completed, the wafer is diced and the active region is fabricated on chips. Three layers of Al gates are deposited using an e-beam evaporator with thicknesses of 30/50/65 nm, increasing from first to last. After the first and second layers are deposited, the native AlO₃ that grows on the gate electrodes is further oxidized as discussed in the main text.
