Delft University of Technology
Department of Applied Sciences

Quantum Device Crossbars as a Statistical Probe of Disorder in SiGe Heterostructures

by

Cornelius Alexander Carlsson

Principal Supervisor: Dr Giordano Scappucci
Daily Supervisor: Davide Degli Esposti

A thesis submitted in partial fulfilment
of the Requirements for the Degree of

MSC, APPLIED PHYSICS

22 April, 2023

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor Giordano Scappucci for his incredible support and flexibility in choosing my own research direction while still navigating me toward a practical end goal. You opened so many doors for me and were never afraid to throw both opportunities and challenges my way. Your scientific wisdom, coupled with great openness and humour, challenges academic convention and I am really glad to have conducted my thesis in your group. Secondly, my daily supervisor Davide Degli Esposti, for his outstanding mentorship, availability for discussion and eagerness to help me whenever possible. Thank you for keeping me on track and inspiring a good sleep cycle as well. I would also like to thank the remaining members of the Scappucci Lab: Alberto Tosato for his fruitful discussions and help with measurement automation, Luka Bavdaz for his insights into our measurement data and grid design, Lucas Stehouwer for his experience with heterostructure growth and Hall bar measurements, and Davide Costa for joining me during the early training phases in the group and input on our measurements. I would also like to thank the wider Spin Qubit division at QuTech for being part of my thesis journey, both academically and socially: Pablo Cova Fariña, for being a great office mate and source of scientific inspiration, Irene Fernández for her hugs and sharing her passion for techno music with me, Brennan Undseth for often granting me access to the office, helping with coding riddles and that Snickers bar, Eline Raymenants for her genuineness, Daniel Jirovec for his puns, as well as Job van Staveren for explaining to me the electronics behind our cryogenic multiplexer. I would like to give a special thanks to Hanifa Tidjani for her infectious positivity, care, affection and supply of delicious evening meals - I couldn't have endured this thesis without you, so thank you. I would like to thank my closest friends from my master's programme, as well as those I made during my thesis, whose camaraderie made this experience evermore enjoyable, including Davit Petrosyan, Clara Grcevic, Georgious Stylianou, Adriá Riera, Casper Mebius, Kenji Capannelli, and Vicky Domínguez Tubío. Thank you to you all. Lastly, I would like to thank my parents, for their love and support in all my academic, social, and professional endeavors. Thank you, it means more than words can describe.

ABSTRACT

Spin qubits in semiconductor quantum dots are a compelling platform for large-scale quantum computation thanks to their small footprint, long coherence times, and compatibility with advanced semiconductor manufacturing. Single electrons may be confined in tensile-strained Si quantum wells in Si/SiGe heterostructures or at the semiconductor/oxide interface in SiMOS, and holes in compressively strained Ge quantum wells in Ge/SiGe heterostructures. Despite the many milestones achieved by the community, a dominant material platform still needs to emerge among these, and the competition is open.

Some factors these platforms compete on include T_1, $T_2^\star$ and charge noise at 1 Hz, which are often correlated with material disorder qualifiers, such as mobility, percolation density and quantum lifetime. Uniquely, in Si/SiGe and Ge/SiGe heterostructures, strain fluctuations in the virtual substrate below the quantum well give rise to non-uniformities that emerge as a periodic roughness at the surface, called a "crosshatch" pattern. With the size of multi-qubit devices crossing the crosshatch wavelength of $\sim$1 µm, a more precise understanding of the effects of these strain fluctuations is necessary.

In this thesis, we propose a crossbar grid architecture as a testbed to investigate the disorder at length scales comparable to the cross-hatch and beyond. To this end, we design, fabricate, and characterize the first grid of single-hole transistors (SHTs) in Ge/SiGe heterostructures.

The grid integrates the shared control of 648 SHTs, of which we successfully operate 647, highlighting the robustness of our design and fabrication flow. We characterize the device uniformity in transport using turn-on thresholds and maximum currents, and the dimensional uniformity with SEM images. We find small variations of 5.7% of mean values for both dimensional and transport metrics for specific measurement conditions. We highlight how device leakage, turn-on thresholds, and maximum currents are sensitive to our gate bias conditions and find that, under high bias conditions, the leakage pattern appears to align with the cross-hatch. We encounter device drift as the most significant problem during our measurements, preventing us from tuning the SHTs into the Coulomb blockade regime. We explain device drift by the filling of charge traps in or near the dielectric/semiconductor interface. We deem this as also being responsible for a negative current observed before device turn-ons. We propose two schemes to help mitigate device drift, which exploit the kinetics of trap filling in both cases. Under these new conditions, we find reduced skewness of data, which helps expose the true uniformity of our devices. In these cases, maximum currents are measured within only 3% of the mean. Moreover, we propose a new way to gauge the energy density and kinetics of charge traps via an effective voltage mapping technique. Finally, in the quest to improve device stability, we characterize a second grid exposed to UV light, yet we do not find a substantial performance improvement.

This work demonstrates the feasibility of using a crossbar architecture to achieve a statistical characterization of the material and single-device performance and highlights the importance of improving device stability in Ge/SiGe to aid the future development of reliable quantum devices.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

Abbreviations

2DEG	Two-Dimensional Electron Gas
2DHG	Two-Dimensional Hole Gas
Ge	Germanium
HB	Hall Bar
LD	Loss-DiVincenzo
MD	Misfit Dislocation
MOS	Metal-Oxide-Semiconductor
QD	Quantum Dot
SET	Single Electron Transistor
SHT	Single Hole Transistor
Si	Silicon
SOI	Spin Orbit Interaction
TD	Threading Dislocation

Symbols

μ	Hole mobility
ρxy	Transverse resistivity
σ_{xx}	Longitudinal conductivity
B_z	Perpendicular magnetic field strength
$I_{sd_{max}}$	Maximum measured drain current
I_{sd}	Measured drain current
n	Hole carrier density
n_p	Percolation threshold
V_G	Gate voltage (plunger voltage in the case of SHT)
V_{Ghold}	Plunger gate holding voltage
$V_{barrier}$	Barrier gate voltage(s)
V_{screen}	Screening gate voltage
V_{sd}	Source-drain bias
V_{th}	Turn-on threshold voltage

1 INTRODUCTION

1.1 The Rise of Quantum Computing

A computer, according to the Oxford dictionary, is "an electronic machine that can store, organize and find information, do processes with numbers and other data, and control other machines." However, the development of computers, as we know them today, long predates the discovery of electricity. The earliest known analog computer has been dated ∼100 BC [1]. Retrieved from a shipwreck off the coast of Greece in 1901, the Antikythera mechanism was housed in a wooden-frame case and comprised numerous bronze gearwheels. It had an application-specific design to predict astronomical positions and eclipses. The first example of a general-purpose computer was invented much later by Charles Babbage, in 1837. This device, known as the Analytical Engine, was capable of integrating memory, executing loops, and incorporating arithmetic logic. In this way, it could tabulate logarithms and trigonometric functions, which appear across all areas of science and engineering.

The arrival of electricity ushered in the era of digital computing. Rather than information being stored mechanically, the presence and absence of electrical current, controlled via applied electric fields, represents the binary logic states. The original vacuum tube technology used by early digital computers relied on the thermionic emission of electrons. This was superseded by the invention of the transistor at Bell Laboratories in 1947. Integrated circuits arrived in the late 1950s, followed by the first-ever microprocessor in 1971. This Intel 4004 microprocessor featured 2,300 transistors [2] and formed the basis for today's most powerful chips. The Apple M1 Max, released in 2021, is equipped with a staggering 57 billion transistors [3].

Quantum computing is yet another paradigm shift in computing. Here, information is encoded in the quantum state of objects. For example, the polarization of photons, the energies of ions, or the flux through superconducting rings. What motivates the development of such computers is the fundamental ability of quantum states, $|0\rangle$ and $|1\rangle$, to be superposed and entangled with one another. This property enables the representation of exponentially more logical states, and can therefore be used to explore large data sets and parameter spaces which are not accessible using classical machines. In 2019, Google published a groundbreaking result in which they used these principles to demonstrate quantum supremacy – a dramatic reduction in the time required to solve a task deemed intractable on hardware running any classical algorithm. The task, pseudo-random sampling to reconstruct a speckle pattern, took only 600 seconds, whereas it would have taken approximately 30,000 years on the most advanced classical hardware. Moreover, the power consumption of the quantum processor was 26 kW, whereas it was estimated that performing the same task classically would cost 50 trillion core-hours and

consume one petawatt hour of energy. All this was achieved with just 53 qubits [4].

1.2 The Case for Spin Qubits

One of the most promising quantum computing platforms uses the spin degree of freedom of particles to encode information. This offers several key advantages to scaling over other technologies. The first is their compatibility with existing semiconductor manufacturing techniques. With architectures closely mimicking that of conventional CMOS technology, leveraging industry-scale processes will enable high device yield and uniformity [5]. Second, they have a small footprint. With a device pitch as small as 80 nm [6], this poses a great opportunity for the integration of millions of quantum bits (qubits) on the same chip.

To mitigate errors due to thermal excitation, it is desired to initialize qubits in their energy ground state. This is typically achieved through thermalization to a cold reservoir where the thermal broadening of the Fermi-distributed qubits is smaller than their transition energy. Due to spin qubits' small size, their power dissipation is also low. Hence it is comparatively easy to remain within this thermal budget, which is typically in the millikelvin range. As of 2020, full two-qubit logic with electron spins was also demonstrated up to 1.5 K [7] and quantum coherence was shown to be hardly affected in the range 0.4 – 1.25 K [8]. Hence, there is great promise in increasing spin-qubit count without needing a greatly customized or expansive cryogenic infrastructure.

For coherent operation of *superconducting* qubits, a total attenuation of about 60 dB is required between the room-temperature electronics and the chip, while simultaneously providing large signal bandwidths of 4 – 8 GHz [9]. This introduces a cable heat load management bottleneck that is further exacerbated by cross-couplings between the many RF drive lines that address individual qubits [10]. Fabricating spin-qubits on the backbone of CMOS technology largely overcomes these problems due to the ease with which classical control electronics can be co-integrated onto the same chip. With numerous existing cryogenic CMOS (cryo-CMOS) and multiplexer architectures [11–13], the community is poised to co-integrate hardware for large qubit counts. While the current state of the art cryo-CMOS chip bears just two-qubits [14], devices such as in [15] claim to be scalable to 1000 qubit-control electrodes while still capable of being cooled by a commercially available dilution refrigerator.

The last key advantage of spin-qubits is their very long coherence times (T_1). Nowadays, T_1 times can typically reach the second-scale, while dephasing times are routinely on the order of tens of microseconds and have reached up to milliseconds [16]. This compares favourably with superconducting qubits, which have similar T_2 and T_2^*, but orders of magnitude lower T_1. More important, however, is the ratio between coherence times and gate speeds, as this indicates the volume of qubit operations that can be performed. Here, spin qubits dominate by three orders of magnitude [17]. Furthermore, it is crucial for gate fidelities – the closeness of an actual qubit operation to its target unitary operation – to exceed 99%, as this is the threshold for fault-tolerance [18, 19]. With a physical error rate below this threshold, quantum error correction schemes can be run to suppress computing errors. As of 2022, spin qubits have delivered this requirement [20, 21], paving the way toward a scalable quantum computation system.

1.3 The Scaling Challenge

A handful of scalable architectures have been proposed. One of these draws inspiration from existing dynamic random access memory (DRAM) systems to make a multiplexed spin qubit array module that hosts 480 qubits [11]. Alternatively, one could use coherent links to connect spatially separated registers of a small number of dense qubits [22]. These links would facilitate quantum information transfer across the quantum chip, permitting also long-range two-qubit

gates and entanglement. Since this proposal, said links have been successfully realized on the millimeter scale using capacitively coupled superconducting resonators [23, 24], as well as using a series of voltage-controlled gates to physically shuttle the spin over several microns [25].

Despite these efforts, and the many tantalizing attributes of spin qubits, the record number of universally controlled spin qubits on a single processor is still a mere six [26]. Likewise, the largest square array of spin qubits is 2×2 [27]. The reason for this is that noise forces a trade-off between high-fidelity initialization, readout, one-qubit, and two-qubit gate fidelities. As a result, growing the qubit count while achieving high fidelities across all these components remains a major challenge. Generally referred to as *charge noise*, electrical noise is a major contributor to this issue. Charge noise arises from the random motion of charge anywhere in the device: stray voltages, phonons, or charge trapping by defect states. Through coupling to the magnetic field, the motion of random charges also impacts the extent of Zeeman splitting, AC driving fields and/or spin-orbit couplings that are used to control spin qubits and, as a result, severely detriments their coherence times. 'Sweet spots' are operating regimes where a qubit becomes minimally sensitive to the electronic noise from the environment [28]. However, in emergent materials for spin qubits, lattice mismatch introduces a strain field with a periodicity much larger than the size footprint of a single qubit. This issue is exacerbated by the presence of randomly generated dislocations, which are also separated by distances greater than those spanned by quantum processors of the present day.

It is believed that the point of the development cycle for spin-based quantum computers we are at today is where classical computers were in the 70s. To progress the field from proof-of-concept devices to integrated, multi-purpose machines, the impact of material disorder and noise on large-scale quantum devices is an area of research that requires attention.

1.4 Project Aims

The purpose of this thesis is to demonstrate the feasibility of a crossbar array architecture for statistically probing material disorder over a hundred-micron length-scale in two systems: silicon and germanium. While charge noise studies have been conducted in both of these systems [29, 30], they are not robust against large-scale material variability as well as fabrication errors, which will become relevant for future quantum processors. In this thesis, we set out to measure transport and charge noise properties in hundreds of devices to acquire statistical insights into their performance.

The structure of this thesis is as follows: Chapter 2 covers relevant theory and the framework under which experimental work is conducted. Chapter 3 explains the design of the grid architecture in germanium and the necessary design changes for implementing the same architecture in silicon. Chapter 4 summarizes the fabrication of the grid, a statistical analysis of its uniformity, and the challenges that arose during its processing. Chapter 5 includes Hall bar measurements and various charge transport characterizations done on the grid. We dedicate a large portion of this chapter to investigating negative currents and strong device drift, which were unexpected behaviors. Lastly, we draw conclusions in Chapter 6 and provide an outlook for further work.

2

BACKGROUND

2.1 Quantum Computation with Spins

A spin-based quantum computer was first proposed by Daniel Loss and David Di Vincenzo in 1998 [31]. The computational basis states $|0\rangle$ and $|1\rangle$ are formed by the projection of an electron's spin quantum number $m_s = \pm 1/2$. The degeneracy between spin-up and spin-down is lifted via the Zeeman effect by applying a static magnetic field $\mathbf{B}_0$,

$$H_{Zee} = g_e \mu_B \mathbf{B_0} \cdot \mathbf{S}_z, \tag{2.1}$$

where g_e is the electron's g-factor, $\mu_B = \frac{e\hbar}{2m_e}$ is the Bohr magneton and $\mathbf{S}_z$ is the electron's spin-half operator along the z-axis. The magnitude of the energy splitting is linearly proportional to the applied magnetic field strength,

$$\Delta E = g_e \mu_B B_0. \tag{2.2}$$

Quantum computation proceeds by gating a tunneling barrier between neighbouring electron pairs. When subject to a low voltage, the electron pairs experience a Heisenberg exchange coupling $J(t)$, as described in the Hubbard model,

$$H_s(t) = J(t)\mathbf{S}_1 \cdot \mathbf{S}_2. \tag{2.3}$$

Here, $\mathbf{S}_i$ is the spin-half operator for an electron on position i. By pulsing this coupling for some duration τ, a coherent interaction, termed a quantum gate, can be achieved. This is analogous to how digital computers operate with Boolean logic gates. To manipulate the state of a single electron spin, a transverse AC magnetic field $\mathbf{H}$ is applied. When the field frequency is tuned to be resonant with the frequency of the spin splitting, electron-spin resonance (ESR) can be used to carry out spin rotations. Continuous AC driving of the qubit will produce *Rabi oscillations*, which are consecutive transitions between the electron's two spin states. Together, this completes a universal set of one- and two-qubit gates for the Loss-Di Vincenzo (LD) qubit.

One extension to the LD qubit uses *holes*, the absence of an electron, as the quantum particle whose spin to manipulate. In candidate qubit materials, such as germanium (Ge), holes offer several key advantages over electron spins, such as larger spin-splitting energies due to larger effective g-factors. Furthermore, an inherently strong spin-orbit interaction (SOI) in Ge enables all-electrical control of qubits, which considerably simplifies quantum computation by eliminating the need for magnetic pulses.

2.2 Quantum Dots in Si and Ge

A quantum dot (QD) is a semiconductor system measuring no more than a few nanometers in diameter and exhibits properties distinctly different from the bulk. Due to its small size, electrons or holes on the QD experience strong confinement, which results in their electronic wave function resembling that of single atoms. To make a spin qubit, the population of electrons or holes on the QD is depleted down to the single particle regime. QDs can be fabricated in a variety of ways. In single molecules or through self-assembly, the material dimensions themselves mark the boundary of the QD. Alternatively, strain, band-gap engineering, gated potentials, or a combination thereof can be used on a larger material sample to artificially create a tight confining potential for the electron or hole to live in. In this thesis, we will mainly be concerned with the latter, specifically, planar gate-defined QDs. Here, a thin slab of material is engineered to support a two-dimensional electron/hole gas (2DEG/2DHG) by gating a metallic *plunger* electrode. This yields vertical confinement of charge along the material growth direction. By means of additional gating *barrier* electrodes, the confining potential can be tuned to create an isolated island of charge.

The first spin qubits in QDs were realized using GaAs/AlGaAs heterostructures [32, 33], thanks to the maturity of molecular beam epitaxy of III-V compounds. Furthermore, GaAs has a small electron effective mass of $0.067m_e$, which necessitates larger confining potentials and thus relaxes lithographic features sizes. However, the absence of zero nuclear spin isotopes in III-V materials makes hyperfine interactions a significant source of decoherence. Moreover, III-V materials are not compatible with the advanced CMOS process flow. For these reasons, research in the past decade has shifted toward silicon (Si), the backbone of the semiconductor industry and a material with spinless isotopes. Natural silicon contains only 4.7% of the spin-full ^{29}Si isotope (with the remainder being ^{28}Si), and this can be reduced down to ppm concentrations through isotopic purification [34]. Ge, also a group-IV material, enjoys these same benefits.

Besides the more exotic geometries that utilize Si and Ge, such as (core-shell) nanowires [35, 36], FinFETs [37] and 'SLEDGE' [38], three planar heterostructures dominate spin-qubit research: Si Metal-Oxide-Semiconductor stacks (SiMOS), Si quantum wells (QW) sandwiched by Si-rich SiGe (Si/SiGe) and Ge quantum wells sandwiched by Ge-rich SiGe (Ge/SiGe). These are depicted in Fig. 2.1

In all of SiMOS, Si/SiGe, and Ge/SiGe, the vertically confined 2DEG/2DHG originates from the equilibrium band offsets of a few hundred meV at the semiconductor heterojunctions. Since there is no chemical doping involved, there is no initial intrinsic band-bending. In the case of SiMOS and Si/SiGe, a positive bias on the plunger pulls electrons toward the top interface. This causes a downward bending of valence and conduction bands. At sufficiently large positive bias, the conduction band crosses the intrinsic Fermi-level, and free electrons can accumulate underneath the gate at the heterojunction. Conversely, in Ge/SiGe, a negative bias is applied such that an upward band bending can take place, and holes can accumulate once the valence band crosses the Fermi level. The modulation of electrical conductivity by introducing free charges is known as the *field-effect*.

In general, a large and burried band offset is desired as accumulated regions then become better isolated from the overlying dielectric, making them less susceptible to impurities and defects. Moreover, the increased lattice strain caused by a large band offset in the case of Ge enhances the extent of heavy-hole light-hole splitting (HH-LH) [39]. In the case of Si, a greater band offset also causes the confining potential to take on a sharper geometry. This has the consequence of more Ge atoms being probed by the electron wave-function at the interface, thereby increasing valley splitting [40]. Both large HH-LH splittings and valley splittings are desirable for coherent qubit operation due to the suppression of decoherence pathways via energy-level mixing.

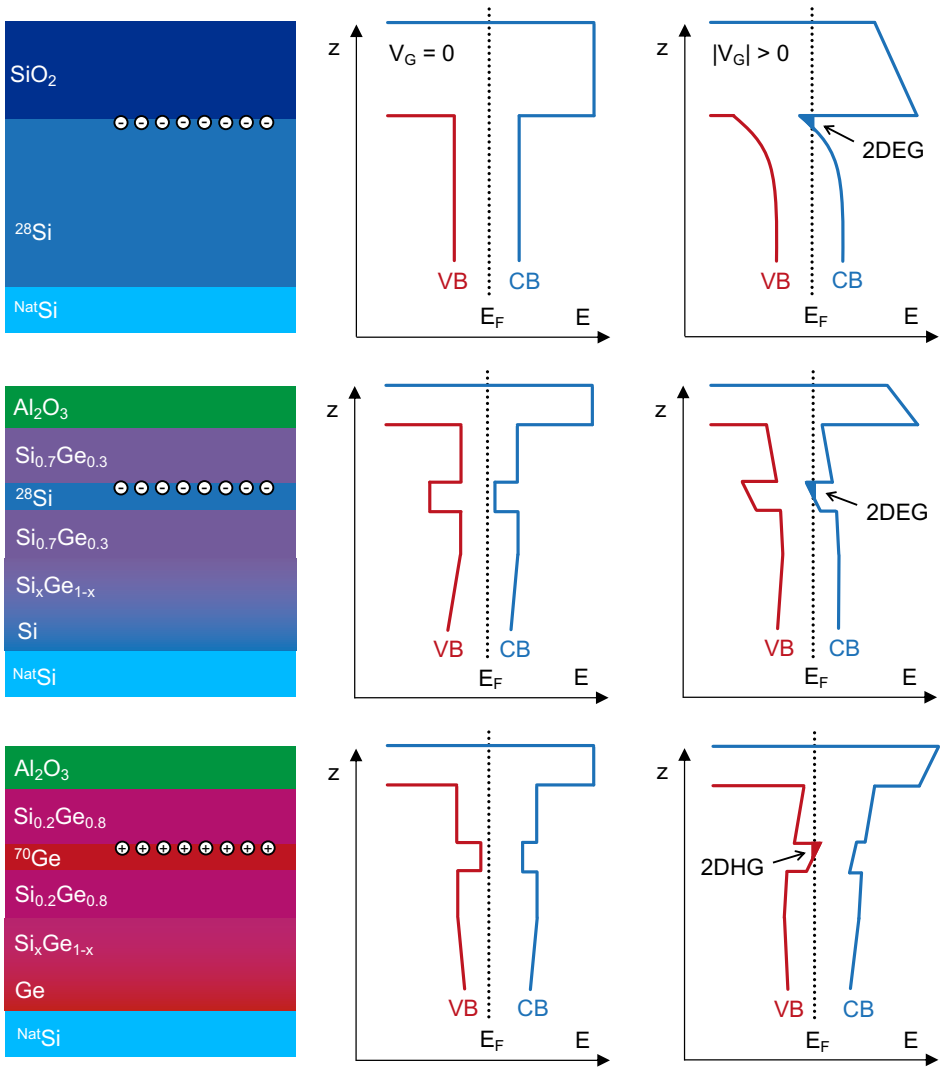


Figure 2.1: Schematic comparing the three competing material stacks for hosting electrons and holes in semiconductor quantum dots. We report the material stack, and the band diagram with no voltage and with a non zero voltage applied (left to right) for SiMOS, Si/SiGe and Ge/SiGe heterostructures (from top to bottom).

2.3 Single Electron/Hole Transistors

By tunnel-coupling two charge reservoirs on either side of a QD, a current can be made to flow across the QD when its electrochemical potential aligns favourably with those of the reservoirs. This alignment can be achieved by shifting the energy spectrum on the QD with its plunger gate bias, as well as applying a voltage between the two reservoirs, termed *source* and *drain*. In the opposite regime, potentials can be tuned so that current is entirely suppressed, i.e. the energy cost of adding an extra charge onto the QD exceeds the available thermal energy. A device that operates in this way is known as a single electron transistor (SET) in the case of electron transport, and a single hole transistor (SHT) in the case of hole transport. Placing barrier gates between the source and the QD, and the QD and drain, gives tunability over tunnel couplings and enables single charges to be loaded, held, and unloaded from the QD. This forms the basis for common single-shot qubit read-out techniques such as energy selective measurement (ESM) [32], more commonly known as "Elzerman readout", as well as time selective measurement (TSM) [41]. Hence, SETs/SHTs form the basis for today's spin-qubit computing architectures, and

understanding their behaviour serves excellently toward scaling quantum computing technology. The theory of SETs laid out in the following section is based on Refs. [42–44], and can be extended to SHTs by flipping gate polarity.

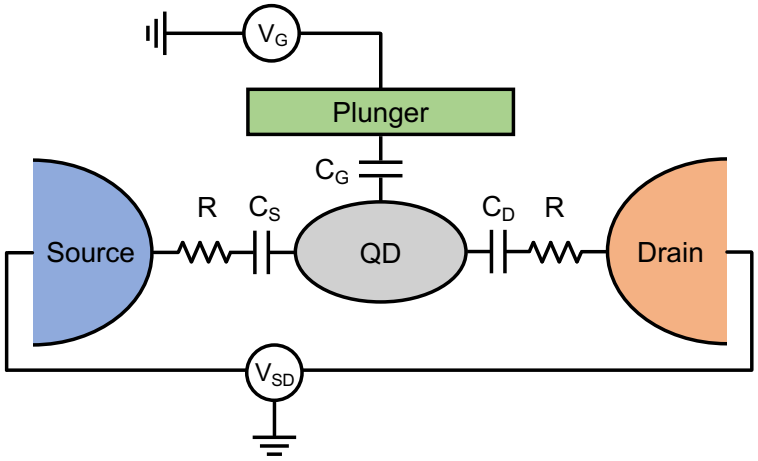


Figure 2.2: Schematic of the single-electron transistor.

2.3.1 Constant Interaction Model

The simplest model that captures blockade effects and the energy spectrum of the QD is the *constant interaction model*. An SET containing a QD that is capacitively coupled to a plunger, source, and drain electrode is considered. In this model, all Coulomb interactions between electrons in the QD, and of electrons between the QD and the reservoirs, are parametrized by a single capacitance C_Σ. This capacitance is fixed and is independent of the number of electrons already on the QD,

$$C_\Sigma = C_S + C_D + C_G. \tag{2.4}$$

It is also assumed that these electron interactions bear no contribution to the single-electron energy level spectrum on the QD. Under these assumptions, the total charge on the quantum dot Q is given by,

$$\begin{aligned} Q &= C_S(V_Q - V_S) + C_D(V_Q - V_D) + C_G(V_Q - V_G) \\ &= C_\Sigma V_Q - C_S V_S - C_D V_D - C_G V_G \end{aligned} \tag{2.5}$$

where V_Q, V_S, V_D and V_G are the voltages on the QD, source, drain and plunger, respectively. By re-arranging the above formula, one can define the voltage on the QD as,

$$V_Q = \frac{Q}{C_\Sigma} + V_{ext}, \tag{2.6}$$

with V_{ext} being the contribution to voltage coming from the QD's surroundings,

$$V_{ext} = \frac{C_S V_S + C_D V_D + C_G V_G}{C_\Sigma}. \tag{2.7}$$

Using these definitions, the total electrostatic energy $U(N)$ needed to charge the QD with N electrons is,

$$U(N) = \int_{Q=0,V_{ext}=0}^{Q=-eN,V_{ext}} V_I \, dQ - Q \, dV_{ext} = \frac{e^2 N^2}{2C_\Sigma} + eNV_{ext}. \tag{2.8}$$

The total energy on the QD is then the electrostatic energy plus the already-occupied single-particle energies.

$$E(N) = U(N) + \sum_{i=1}^{N} \epsilon_i \tag{2.9}$$

The electrochemical potential, defined as the work done by adding a single electron to the QD that already holds $N-1$ electrons, becomes

$$\mu(N) \coloneqq E(N) - E(N-1) = \frac{e^2}{C_\Sigma}\left(N - \frac{1}{2}\right) - eV_{ext} + \epsilon_N. \tag{2.10}$$

Comparing Eq. 2.8 & 2.10, it is clear that the electrostatic energy has a quadratic dependence on N, while the electrochemical potential scales linearly, both with respect to N and V_{ext}. This implies that the electrochemical potential on the QD forms a roughly equally spaced "ladder" and is only ever rigidly shifted by varying the plunger voltage. The energy that separates these levels is called the *addition energy* $E_{add}(N)$, and depends on a purely electrostatic charging energy $E_c = e^2/C_\Sigma$, as well as a discrete energy spacing ϵ_N arising from the chemical term.

$$E_{add}(N) \coloneqq \mu(N) - \mu(N-1) = \frac{e^2}{C_\Sigma} + \epsilon_N \tag{2.11}$$

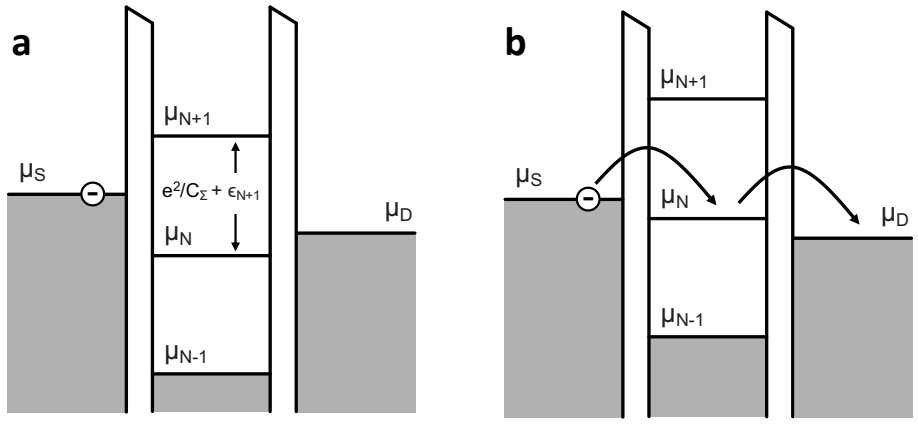


Figure 2.3: Diagrams of the electrochemical potential levels of a quantum dot in the low-bias regime. a) If no level in the dot falls within the bias window, the electron number is fixed. b) If electrochemical potentials are aligned, the number of electrons on the dot can fluctuate by one and a tunneling current develops between source and drain.

Under a small source-drain voltage $V_{SD} = V_S - V_D$, a step is created in the electrochemical potential between the reservoirs, equal to $\mu_S - \mu_D = -|e|V_{SD}$. This is called the bias window. In the *weak bias regime*, the bias window is narrower than the addition energy of the QD ($-|e|V_{SD} < E_{add}$). Consequently, the gate voltage can be tuned such that the chemical potential μ_N inside the dot aligns with that of the source and drain in a descending fashion ($\mu_S \gtrsim \mu_N \gtrsim \mu_D$). Under these circumstances, a tunneling-current can flow and the number of electrons on the dot will fluctuate between N and $N-1$. If the gate voltage were to be increased such that $\mu_{N+1} > \mu_S$, $\mu_N < \mu_D$, the number of electrons on the dot remains fixed and no current will flow. This is known as *Coulomb blockade*. By sweeping the gate voltage, a series of conductance peaks, called *Coulomb peaks* can be observed, each corresponding to electron transport through consecutive, single levels on the QD. These are called *Coulomb oscillations* and are depicted in Fig. 2.4a.

2.3.2 Conduction at High Bias

It is important to realize that these charging effects are only observed if the number of electrons on the dot remain quantized at all times. This is enforced by the tunnel barrier having some minimum resistance R to prevent leakage. The charging time for the QD is given by the time constant $\tau = RC$, from which it follows that the minimum energy uncertainty will be $\delta E = \frac{\hbar}{\tau}$. As a result, a well defined number of electrons on the QD exists if the following condition is met

$$\delta E < E_c = \frac{e^2}{2C_\Sigma}. \tag{2.12}$$

A simple substitution reveals that, in other words, the barrier conductance should be smaller than the conductance quantum,

$$\frac{1}{R} = G < \frac{e^2}{2\hbar} \sim \frac{2e^2}{h}. \tag{2.13}$$

If V_G is biased appreciably, or the gating potential on the barriers is reduced, the potential landscape of the QD broadens to a point where it is no longer strongly confined by the barriers. As a result, the barrier conductance exceeds the conductance quantum and quantized transport is lifted. This regime mimics the operation of conventional field-effect transistors, where a continuous conduction channel is created between the source and drain. The point in voltage at which this occurs is called the *turn-on threshold* V_{th}. In this scenario, the current is spread over a continuum of energy states and grows continuously with further biasing of V_G due to the introduction of additional free charge carriers. This is quantified via the *transconductance* as $G_S = I_{sd}/V_G$. Eventually, however, a current saturation point I_{sat} is reached once the channel width becomes fully developed. At this point, the current is limited by the intrinsic carrier *mobility*, which we discuss in the following section.

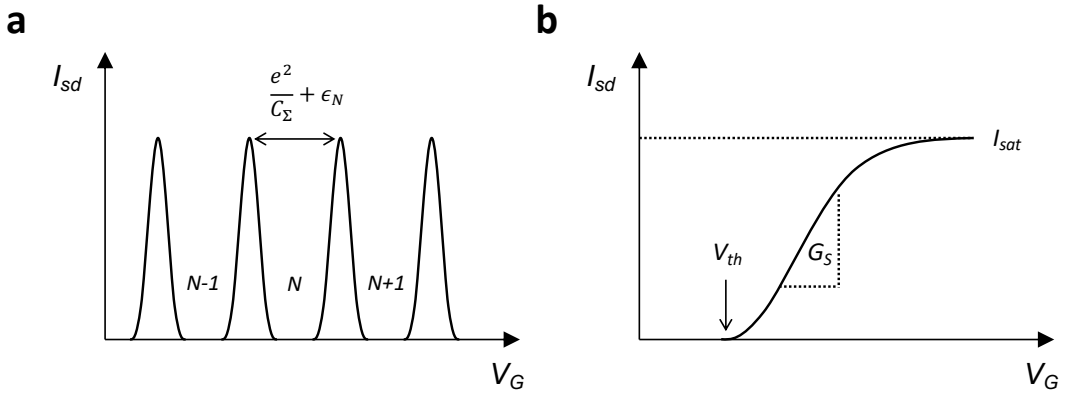


Figure 2.4: Current-voltage plots of a **a)** single electron transistor with strong QD confinement leading to coulomb oscillations **b)** developed conduction channel between source and drain leading to non-discrete transport.

2.4 Charge Transport

A prerequisite to making well-behaved SET/SHTs, and by extension qubits, are high-purity materials with good transport properties. Transport metrics inform us of the disorder level both in the bulk and in lower dimensions, thus making them good predictors of the level of charge noise and energy fluctuations that a qubit may encounter in its lattice environment.

2.4.1 Drude Model

The Drude model [45] is an application of kinetic theory that, from a classical picture, correctly explains the conductive behaviour of metals at ambient temperatures. The model considers a sea of electrons that may instantaneously collide with stationary lattice ions, thereby changing direction. All other interactions apart from the collisions between electrons and ions are neglected. The probability of any one electron randomly colliding with an ion is characterized by a *relaxation time* τ, also called a *scattering time*. On average, electrons travel ballistically for a time τ before colliding. Between collisions, electrons are in thermal equilibrium with their environment and will have a velocity of $\sqrt{3k_BT}$. The average distance traveled by electrons between collisions is the *mean free path* λ. Under these simplifying assumptions, a microscopic formulation of Ohm's law emerges,

$$\mathbf{J} = \sigma\mathbf{E}, \tag{2.14}$$

where $\mathbf{J}$ is the current density, $\mathbf{E}$ is the electric field and σ the conductivity, given by

$$\sigma = nq\mu. \tag{2.15}$$

Here, n is the charge carrier concentration, q is the electron's charge, and μ is the mobility. The mobility quantifies how quickly charges can migrate in response to an external electric field and depends on the relaxation time τ and electron mass m.

$$\mu = \frac{q\tau}{m} \tag{2.16}$$

The more scattering sources there are, the more frequent collisions become the shorter the relaxation time, and the lower the current.

2.4.2 Classical Hall Effect

The Hall bar has become a ubiquitous tool for measuring charge carrier concentrations and mobility. Its geometry is shown in Fig. 2.5. By applying a longitudinal voltage V_{xx} between terminals 1 & 2, mobile charge carriers develop a current I_{xx} through the device. If a perpendicular magnetic field B_z is applied, an additional Lorentz force is induced $\mathbf{F}$ onto the carriers,

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{2.17}$$

where $\mathbf{v}$ is the drift velocity. This results in a curved trajectory of the carriers, causing charge to accumulate and deplete on opposite faces of the Hall bar, a phenomenon known as the *ordinary Hall effect*. This separation of charge generates the electric field $\mathbf{E}$, which opposes further carrier migration. In the steady state, $\mathbf{F} = 0$, and the magnitude of the electric field in the transverse direction is given by

$$E_{xy} = v_{xx}B_z. \tag{2.18}$$

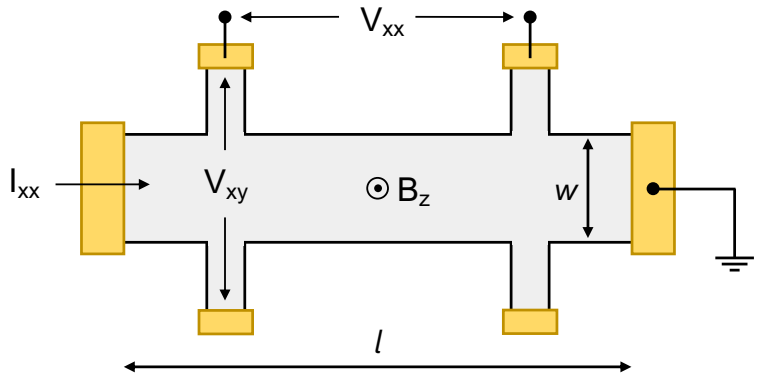


Figure 2.5: Hall bar geometry

Using the expression for current in terms of longitudinal drift velocity v_{xx}, charge carrier density n, width w and thickness t of the Hall bar,

$$I_{xx} = nq(wt)v_{xx}, \tag{2.19}$$

the *Hall voltage* V_{xy} in the transverse direction, between terminals 1 & 3, can be formualted as

$$V_{xy} = E_{xy}w = \frac{I_{xx}B_z}{nqt}. \tag{2.20}$$

Hence, by re-arranging Eq. 2.20, one can express the charge carrier concentration as

$$n = \frac{I_{xx}B_z}{V_{xy}qt} = \frac{B_z}{R_{xy}qt}, \tag{2.21}$$

where R_{xy} is the *Hall resistance*. For a 2DEG/2DHG, there is no associated thickness for conduction, so t can be neglected in Eq. 2.21. As we will find, however, mobility is always independent of t. By Ohm's law, the longitudinal resistance is equal to $R_{xx} = V_{xx}/I_{xx}$, and thus the resistivity for the Hall bar geometry would be, $\rho_{xx} = R_{xx}w/l$. Since the resistivity tensor ρ is something that can be experimentally measured easily using lock-in techniques, it is convenient to write μ in terms of it. By substituting Eqs. 2.21 into Eq. 2.15, and noting that $\sigma_{xx} = 1/\rho_{xx}$ we reach the final expression for carrier mobility,

$$\mu = \frac{\sigma_{xx}}{nq} \equiv \frac{R_{xy}}{\rho_{xx}B_z}, \tag{2.22}$$

By using the field-effect, a top gate voltage V_G on the Hall bar can be used to increase the free charge carrier density n. Within a linear approximation, an effective capacitance can be defined as,

$$C = q\frac{dn}{dV_G}. \tag{2.23}$$

By sweeping the gate voltage, the trend in conductivity σ_{xx} as a function of the charge carrier concentration n can be extracted. Using a percolation transition model [46, 47], the data can be fitted to

$$\sigma_{xx} \sim (n - n_p)^p \tag{2.24}$$

where n_p is the *percolation threshold* and p is the *percolation exponent*. The percolation threshold is a measure of the minimum number of free charge carriers that must be introduced into a system

to overcome its disorder and establish an unobstructed conduction path. n_p marks the transition point between insulator ($n < n_p$) and metal ($n > n_p$). Typically, materials that exhibit higher mobilities have lower disorder and therefore have lower percolation thresholds.

2.5 Disorder

In addition to good transport metrics, which provide information on the *average* level of material disorder, our ability to coherently control qubits relies on a homogeneous energy landscape than spans all involved length-scales of a quantum processor – from the nanoscale of individual qubits, to the micrometers that separate multiple qubits. With this comes the importance of understanding the origin of individual defects.

Disorder in crystalline materials can take on many forms, including point defects (vacancies or ad-atoms), line defects (dislocations), or plane defects (grain boundaries). In the context of spin qubits in QDs, plane defects do not come into play, because through careful regulation of growth rate, precursor mass flows, temperature and pressure, SiMOS, Si/SiGe, and Ge/SiGe can be grown out of single crystal substrates. Epitaxial SiGe is commonly grown via reduced-pressure chemical vapour deposition (RP-CVD) using H_2SiCl_2 and GeH_4 as precursors at 650 – 1050 °C and at pressures of 20 – 350 Torr. Si and Ge QWs can be grown with the same method using SiH_4 and GeH_4.

2.5.1 Crosshatching

Dislocations arise whenever a crystal lattice experiences stress that is too large to be accommodated elastically. In this case, it becomes energetically favourable for atoms to displace from their lattice sites. This displacement can lead to dislocations of two types: *edge* or *screw*. An edge dislocation is best visualized as an abrupt removal of a plane of atoms from the crystal structure. A screw dislocation is best visualized as a crystal that has been sliced part-way, then sheared out of the cut plane (mode III). The *line vector* gives the direction of the dislocation, which is constant for an edge dislocation but traces a helix for a screw dislocation. The *Burger's vector* characterizes the direction and magnitude in which atoms are displaced. For edge dislocations, the line and Burger's vector are always perpendicular, whereas, for a screw dislocation, they are parallel. When dislocations arise from a mismatch in lattice constants, such as in SiGe heterostructures, we refer to them as *misfits* (MDs), if they are edge dislocations, and *threads* (TDs), if they are screw dislocations. This important distinction exists because, as opposed to edge and screw dislocations, which may form anywhere in the material, MDs and TDs form at material interfaces and may extend into both an overgrown film as well as the substrate.

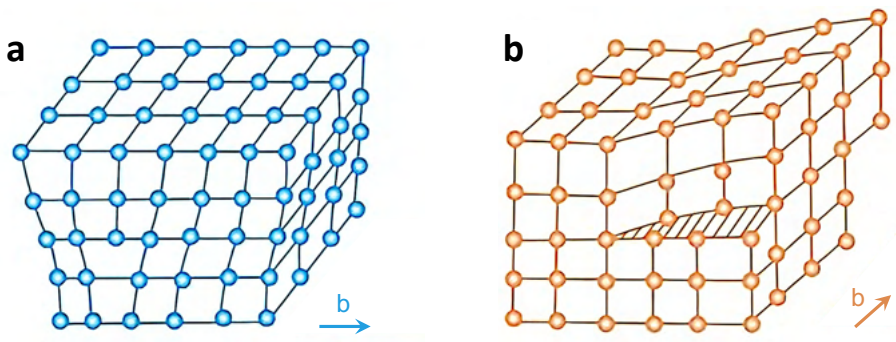


Figure 2.6: Schematics of an **a**) edge dislocation, and **b)** screw dislocation, with Burger's vectors indicated. Adapted from [48].

By a mechanism first described in Ref. [49], a network of MDs at the base of a SiGe buffer layer will lead to an undulated morphology at the surface, known as a *crosshatch pattern*. Assuming the SiGe crystal is grown along the [001] direction on top of a Si wafer, then the formation of an MD will create a locally compressive strain field. The SiGe lattice compensates for this strain energy with an expansion of its lattice constant in the (001) plane. Since the system will tend to conserve its Bravais cell volume, the lattice constant along the [001] direction contracts. This contraction leads to a downward movement of the upper lattice planes. The total displacement at the surface would be an integral of the contraction δ_z with the distance z over the entire buffer thickness. With the concurrent relaxation of atoms in the (001) planes, the ultimate result is a half-sinusoidal morphology at the surface along [110] and [1$\bar{1}$0] directions. Considering the formation of many MDs, this will lead to a periodic surface roughening, and the aforementioned crosshatch pattern [49]. Images made by atomic-force microscopy (AFM) can easily resolve this crosshatch pattern and are shown in Fig. 2.7.

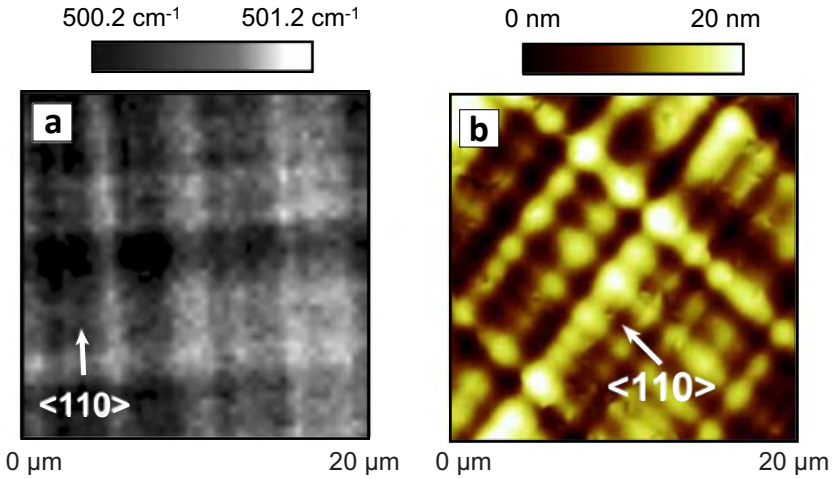


Figure 2.7: a) Raman-mapping image of the Si-Si bond peak and **b)** AFM surface image of strained Si on a 1 µm thick $Si_{0.7}Ge_{0.3}$ buffer layer homo-epitaxially grown on a polished $Si_{0.7}Ge_{0.3}$ substrate. The crosshatch pattern is visible in both images and with similar wavelength. Adapted from [50].

The crosshatch wavelength has a strong linear dependence on the SiGe buffer thickness [51]. As the distance from the MD network increases, the lattice strain is able to relax more, and both the crosshatch roughness and wavelength increase. For 1 µm thick layers of $Si_{0.7}Ge_{0.3}$ grown at 740 °C, crosshatch wavelengths were found to be ∼4 µm and saturated to ∼25 µm at a thickness of 4 µm and higher, indicating a fully relieved strain in the buffer layer [51].

Spatially resolved Raman mapping is a spectroscopic technique that measures vibrational bond energies across a sample surface, and whose energy shift can be correlated with lattice strain. In crosshatched SiGe, the morphology can be entirely removed by chemical mechanical polishing (CMP), however, Raman mappings in these smooth samples still reveal a crosshatch pattern in the strain field running along ⟨110⟩ directions. Regrowth of SiGe on this polished surface causes crosshatch roughness to re-appear, and also in later overgrown Si, as shown in Fig. 2.7. Moreover, the wavelength of the roughness is almost the same as that of the strain distribution seen in Raman mappings [50]. This is strong evidence for strain fluctuations, caused by the underlying MDs, also affecting the local growth of SiGe and Si [51]. Regions of relaxed SiGe provide lower energy sites for the incorporation of Si and Ge atoms, hence the growth rate here is higher.

Both the crosshatch pattern and TDs are important ingredients for controlling the mobility, because any variation in the lattice, including a local strain field, can act as a scattering center. Carriers are most efficiently scattered when the crosshatch wavelength equals the Fermi wave-

length [52], and this has been shown to cause a $\times 3$ reduction in the mobility, even when the roughness amplitude is $\times 10$ smaller [53]. Compositional grading of SiGe buffers is a widely used method that can lead to a reduction in the density of TDs by $4-5$ orders of magnitude [49]. MDs act as nucleation sites for TDs. Therefore, by keeping the composition of the initial $Si_{1-x}Ge_x$ near $x \approx 0$, and incrementally increasing the Ge content up to the functional layer, a sparser network of MDs can be expected, and consequently also TDs. The name given to engineered buffer layers, for the sake of minimizing their dislocation density, is a *virtual substrate*.

2.5.2 Effects on the Quantum Well

MDs have been shown, through both theory and experiment, to also cause lattice plane bending, which emerges as tilt lines in the measured diffraction profile [54]. This can be explained by the larger atomic radius of Ge preferentially binding to lattice sites with lower compressive strain. This leads to a larger in-plane lattice constant during growth, which is compensated for with lattice tilt. By the same mechanism, the MD network can cause fluctuations in the composition of SiGe buffer layers [55].

With advancements in scanning X-ray diffraction microscopy (SXDM), a quantitative description of the local lattice tilt, strain, and composition has become possible. For one, it has been shown that localized strain gradients in SiGe cause a variation in the crystallographic plane orientation in overgrown Si QWs [56]. The angular displacements amount to 0.022° and vary over lateral distances of approximately 1 µm. In turn, the crystallographic tilt affects the strain distribution in the QW. Over a thickness of just 10 nm, a difference in strain of 3×10^{-7} was measured [56]. In another study on just the SiGe buffer layer, $Si_{0.7}Ge_{0.3}$ samples treated both with and without CMP exhibited an average crystallographic tilt of $\sim 0.051^\circ$ [57]. The combined effects of lattice tilt and compositional variations in SiGe were held responsible in Ref. [58] for strain fluctuations seen in their Ge QW, which also spanned 1 µm length scales. This was the same device used to demonstrate universal control of a four-qubit quantum processor [27].

MDs and TDs are highly relevant for the operation of qubits in QDs, because as qubit count increases, processor sizes will quickly surpass 1 µm, making their performance susceptible to strain fluctuations. A recent work operating a 4×4 QD device in Ge already reaches this 1 µm footprint [59]. In Si, local strain variations have been shown to lead to local changes in the conduction-band energy of 14 µeV, approaching the magnitude of valley-splitting energies [56]. The variation in Ge composition, along with the presence of TDs, also lowers the kinetic barrier for forming MDs inside the QW [60]. Consequently, the maximum dislocation-free thickness of the QW deviates below the ideal Matthews–Blakeslee critical thickness. Being forced to grow substantially thinner QWs is undesirable, as it increases the exposure of qubits to the surrounding, disordered environment. Moreover, a thinner QW imposes a strong vertical confinement, which would increase the energy level splitting between adjacent qubits, reducing their wavefunction overlap, and potentially making it less practical to carry out two-qubit gates. In light of these results, tuning and operating future multi-qubit devices reliably could become a great challenge, and motivates the characterization of wafer-scale SiGe devices and exploration of how strain fluctuations can be better controlled.

2.5.3 Point Defects

There exist various types of point defects, characterized by their chemistry, location and/or energy. In spin/electronic applications, we are mainly concerned with defects that lie within the band gap, as these contribute additional states for electrons and holes at energies smaller than the gap energy. As a result, their presence may severely affect carrier mobilities and scattering times in addition to providing alternative transport pathways.

Regions of a device that already possess higher energies, such as from lattice mismatch, provide a lowered energy barrier for point defect formation. This is why oxide and hetero-interfaces typically have higher defect densities than the bulk, with values reaching as high as $10^{15}\,\mathrm{cm}^{-2}$ [61]. For comparison, bulk defect densities usually have densities around $10^{10}\,\mathrm{cm}^{-2}$. In various $Si/Si_{1-x}Ge_x$ heterostructures, it was found that by increasing the Ge content, the density of charges trapped at the heterointerface continually increased [62]. On the other hand, the easier formation of these defects means they usually lie lower in energy than bulk defects, which often introduce costly lattice distortions. Low-energy defects that lay near the band gap edges are termed *shallow defects* while those located nearer the band gap center are termed *deep defects*. The binding energy for shallow defects is on the order of a few tens of millivolts [63], which is why they can easily release charges once they have been captured. This is in contrast to deep defects, which may hold charges over extended periods, severely reducing a material's carrier lifetime. The *capture cross-section*, with units of cm^2, is a measure of the effective area within which a defect can cause a capture reaction to occur. The capture cross-section of deep defects is typically larger than for shallow defects.

Any defect that can capture and release charges is called a *trap*. Traps come in two flavours – *acceptor* and *donor*. Starting from an initially neutral state, an acceptor trap can become negatively charged by accepting an electron (donating a hole), while a donor trap becomes positively charged by donating an electron (accepting a hole) [61]. Reliable qubit operation suffers from fluctuations in trap occupancy, because they introduce unpredictable Coulomb fields that impact the energy landscape and efficacy of pre-tuned voltage pulses. Furthermore, they may couple to the magnetic field via the SOI, as will be discussed in Sec. 2.6.

The most notorious traps are caused by oxygen vacancies, present in all commonly used gate dielectrics: SiO_2, Al_2O_3 and HfO_2. With knowledge of the temperature and oxygen partial pressure during the dielectric growth, one can estimate the concentration of oxygen vacancies from a Brouwer diagram. Traps located at hetero-interfaces are called *interface traps*, those located some distance away from the interface are called *border traps* or *fixed charge traps*, while those deeper in the oxide are called *oxide traps*. Due to the presence of the surrounding lattice, fixed charge traps and oxide traps are relatively immobile, meaning they do not hop between lattice sites or shift in energy in response to an applied electric field. On the other hand, interface traps have higher mobility and may shift in location and energy.

A common interface trap is the dangling bond, which arises at the interface between amorphous oxides and crystalline Si, Ge or SiGe. Due to the different valency between oxygen and group IV compounds, the under-coordinated Si and Ge atoms behave as free radicals and can be amphoteric in nature. Thankfully, remedies exist for saturating these dangling bonds. In SiMOS, most interface-trapped charges can be neutralized by low-temperature hydrogen annealing. In this way, the Si and Ge surface becomes H terminated and the total surface trap density can be reduced to $10^{10}\,\mathrm{cm}^{-2}$ [61]. Likewise, annealing of Ge/SiGe in forming gas, a mixture of hydrogen and nitrogen, was also found to noticeably improve long-term device stability and minimize sudden shifts in both I-V curves and charge stability diagrams [64]. Combining a forming gas anneal with S passivation was also found to significantly reduce gate leakage currents and improve device stability in air [65]. Increases in conduction have also been observed when processing samples in a UV-ozone reactor, which is proposed to be due to the deliberate population of deep charge traps with carriers excited by UV photons [66].

2.6 Noise

Noise refers to any disturbance that affects the desired behavior of a system. The presence of noise limits the ability of a qubit to maintain its quantum properties, thus our ability to manipulate quantum information with high fidelity. Dislocations and point defects, as discussed in Section 2.5, are two key sources of noise, though noise can also arise from the environment or from other hardware components. Understanding the nature of noise and its mechanisms is therefore crucial for the development of scalable quantum computing technologies.

Three exponential decay constants are used to quantify coherence: (i) *relaxation time* T_1, describing the transition of a qubit from its excited state $|1\rangle$ to ground state $|0\rangle$, (ii) *pure dephasing time* T_2, also known simply as *coherence time*, describing the phase-destroying process of a qubit that puts it in an equal mixture of $|0\rangle$ and $|1\rangle$, and (iii) *inhomogeneous dephasing time* $T_2^\star$, describing the loss of phase synchronicity with either a clock or another qubit [67]. The three coherence times are related via,

$$\frac{1}{T_2^\star} = \frac{1}{2T_1} + \frac{1}{T_2}, \quad T_2 > T_2^\star. \tag{2.25}$$

The energy levels of a qubit are heavily influenced by electric fields, therefore, processes that directly alter a material's band structure and electric field, such as phonon interactions, will have the greatest bearing on T_1. Meanwhile, phase-destroying processes in qubits arise mostly from magnetic fields, which is why hyperfine noise, such as through random nuclear spins, is the main contributor to T_2 and $T_2^\star$.

2.6.1 Phonon-mediated Relaxation

Due to the similar electronegativities of Si and Ge, SiGe crystals arrange themselves symmetrically and yield a net-zero dipole moment in the crystal lattice. In such non-polar crystals, the piezoelectric effect is absent. As a result, only *inhomogeneous* lattice deformations can cause electric-field fluctuations in Si, Ge and SiGe crystals. Lattice deformations of this variety are called *deformation potential phonons*. Due to the SOI, this electric field couples to spin qubits, thereby inducing spin transitions [42]. The relaxation rate of a spin, mediated by single phonons of this type, can be expressed using Fermi's Golden Rule,

$$\frac{1}{T_1} = \frac{2\pi}{\hbar} |\langle\uparrow| H_p |\downarrow\rangle|^2 D(\Delta E), \tag{2.26}$$

where $D(\Delta E)$ is the density of phonon modes at the Zeeman level splitting ΔE, and H_p is the electron-phonon coupling hamiltonian, given by,

$$H_p = M_{\mathbf{q}j} e^{i\mathbf{qr}} (b^\dagger_{\mathbf{qj}} + b_{\mathbf{qj}}), \tag{2.27}$$

where $M_{\mathbf{q}j}$ is the electric-field to phonon coupling strength and $b^\dagger_{\mathbf{qj}}$ and $b_{\mathbf{qj}}$ are the phonon creation and annihilation operators for a given wave vector $\mathbf{q}$ and phonon branch j.

Through a change in the Zeeman energy splitting, theory predicts that the decay rate scales with $1/T_1 \propto B_0^7$ for deformation potential phonons in the low-temperature limit [68]. The scaling as a function of the QD size should follow a $1/T_1 \propto l^8$ dependence [42]. Recently, enhancements in the decay rate have been observed when ΔE is resonant with another energy transition, and the qubit spin levels hybridize with other charge or spin states [67]. This is known as a *hot spot* and can happen when tuning B_0 to values where spin-valley hybridization persists or spin-orbit admixing increases.

Despite these seemingly extreme scaling relations, and additional relaxation pathways, spin

qubit T_1 times are usually long compared to T_2 times. At higher temperatures, two-phonon processes also become important, though this is mostly circumvented by operating qubits at cryogenic temperatures. The current record for the spin lifetime in a nanostructure is 57 s, which was made possible by using a very weak magnetic field and by orientating the magnetic field along the crystalline axis with minimal SOI field [69]. The authors highlight a change in the power law scaling of the relaxation rate around 2 T from a B_0^7 dependence at high fields to B_0^5 at low fields. It is speculated that this corresponds to a crossover from a phonon-dominated to a hyperfine-dominated relaxation mechanism. Considering the record $T_2^\star$ time, which is reflective of hyperfine noise and only reaches 2.3 ms [70], the evidence is compelling to assume that current spin qubit gate fidelities are not limited by phonons [67]. In further support of this, Ref. [69] studied spins in GaAs QDs, which will have much larger gate areas than Si systems, and therefore should suffer more from l^8 scaling. Conversely, T_1 times for electron spin qubits in Si/SiGe are much lower, peaking at a few seconds [16]. Thus, it is improbable that phonons are curbing the relaxation time of qubits, especially in SiGe systems.

2.6.2 Magnetic Noise

Hyperfine noise refers to the noise caused by random fluctuations in the magnetic field. These are mainly caused by the presence of non-zero nuclear spin isotopes in the host lattice. The *Overhauser field* B_N describes the effective magnetic field arising from the collective nuclear environment. This directly couples to a spin's magnetic moment, changing its energy and therefore randomizing its Rabi frequency. Since a spin always precesses about the vector of the total magnetic field that it experiences, a fluctuating magnetic field will also cause an uncontrollable tilt in the qubit's precession axis and accumulation of unwanted phase.

The hamiltonian describing the hyperfine interaction is given by [71],

$$H_{HF} = \sum_k^N A_k \mathbf{I}_k \cdot \mathbf{S}, \tag{2.28}$$

where A_k is the coupling strength between the qubit spin operator $\mathbf{S}$ and the k nucleus spin operator $\mathbf{I}_k$. The coupling strength is weaker for lighter nuclei and for electronic wavefunctions $\psi(\mathbf{r_k})$ that have a reduced overlap with spinful nuclei. For this reason, holes, which have p-orbital symmetry and a node at their origin, have generally weaker hyperfine interactions than electrons, which have s-orbital symmetry. Also, wavefunctions enveloping a larger number of spinful nuclei will have longer inhomogeneous dephasing times, due to the averaging over more nuclear spins [67],

$$T_2^* \propto \sqrt{\frac{N}{p_I}}, \tag{2.29}$$

where N is the number of lattice sites for which $|\psi(\mathbf{r_k})|^2$ is above some threshold and p_I is the probability that a lattice nucleus has spin.

It has been shown that hyperfine interactions can lead to phonon-mediated relaxations when the level of admixing between electronic orbitals and local nuclear spins is strong [72] – another example of a hot spot. Also, when randomly fluctuating nuclear spins return to thermal equilibrium, they may transfer their energy to the lattice, driving the relaxation of spin qubits to their ground state via *spin-lattice relaxation*. Thankfully, hyperfine-induced spin dephasing and relaxation has been greatly supressed by a variety of methods, including isotopic purification (as mentioned in Sec. 2.2), nuclear polarization, and dynamical decoupling protocols [67].

2.6.3 Charge Noise

Charge noise refers to any fluctuation in the electric field of the qubit environment. This may be caused by the presence of crystalline disorder and defects, as discussed in Section 2.5, the random movement of charges, or spurious voltage noise coming from control gates. Despite single spins not directly coupling to electric noise, spin decoherence can be triggered through the SOI. The presence of micromagnets exacerbates this further by enhancing the coupling between spins and electrical drives. Through the d.c. Stark effect, charge noise can also causes a shift and broadening of a qubit's transition energy. Hence, charge noise negatively impacts all of T_1, T_2 and $T_2^\star$. Moreover, stark shifting, as well as random monatomic lattice steps, have been shown to affect local g-factors, which may therefore limit the fidelity of both single and two-qubit gates [73]. Electric field fluctuations also alter a QD's confinement potential, hence its electronic wavefunction, level spacing and how it probes the local atomic environment. Furthermore, for tunnel-coupled qubits, charge noise will cause unintentional modulation of the exchange interaction strength J (Eq. 2.3), thereby inducing two-qubit gate errors. In terms of scalability, efficient spin-cavity coupling via the charge degree of freedom relies on a quiet electric environment, making charge noise an opponent of future long-range qubit interconnects. Consequently, to the best of our knowledge, charge noise is the main limiter of overall qubit performance and has therefore garnered much research interest over the past decade.

A handful of models have been developed to describe electrical noise. All of them derive their results by considering charge traps as two-level fluctuators. As first described by Stefan Machlup in 1954 [74], the trap can be considered as only ever being in only one of two states: filled $x(t) = 1$, or empty $x(t) = 0$. This bi-stability is characteristic of a Random Telegraph Signal (RTS). The lifetime of each of these states is given by σ and τ, respectively. Assuming that the number of free charge carriers is much greater than the number of traps, then the probability of trapping charges is uncorrelated with the number of already filled traps and so the process is Markovian and takes on Poissonian statistics. Under these simplifying assumptions, the spectrum of the resulting noise signal may be acquired using the Wiener-Khinchin theorem, which states that the power spectral density is equal to the Fourier transform of the signal's autocorrelation,

$$S(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \langle x(t)x(t+s)\rangle e^{-i\omega s} ds. \tag{2.30}$$

This yields a Lorentizian noise spectrum as a result [74],

$$S(\omega) = \frac{1}{\pi}\frac{\sigma\tau}{(\sigma+\tau)^2}\frac{1/T}{\omega^2 + (1/T)^2} + \text{d.c. term}, \tag{2.31}$$

with $1/T = 1/\sigma + 1/\tau$ and $\omega = 2\pi f$. The fluctuation in carrier density, mobility, or both [75] that is caused by charge trapping translates to a fluctuating conductivity through Eq. 2.15 and therefore becomes visible in the measured current by Ohm's law.

Since a QD couples to to the Coulomb field of a charge trap (Eq. 2.32), then from this model one may also expect a $1/f^2$ dependence in the current noise spectrum of an SET/SHT, whose amplitude scales with the inverse distance from the trap $(\mathbf{r} - \mathbf{r_t})$.

$$U(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{e^2}{(\mathbf{r} - \mathbf{r_t})}. \tag{2.32}$$

In real materials, however, there likely exist multiple TLFs over a range of distances, energies and characteristic switching times. Such a correction to Eq. 2.30 was first treated by McWhorter [76], who considered an oxide trap lifetime that was dependent on its distance from the semiconductor/dielectric interface x.

$$\tau \propto \tau_0 e^x \tag{2.33}$$

In this scenario, the noise spectrum yields a $1/f$ dependence. In a later treatment, Dutta and Horn [77] considered a non-uniform distribution of activation energies $D(E)$, such that,

$$S(\omega) \propto \int f(E,\omega)D(E)dE. \tag{2.34}$$

By substituting a Debye-Lorentz model for the TLF spectrum,

$$f(E,\omega) = \frac{\tau_0 \exp(E/k_BT)}{\omega^2\tau_0^2 \exp(2E/k_BT) + 1} \tag{2.35}$$

the authors found that the frequency exponent could then take values between 1 and 2, as well as bear a temperature dependence due to the thermal activation introduced to τ – in line with experimental observations. If $D(E)$ is Taylor expanded in power of T before integration, the central result is reached:

$$S(\omega,T) \propto \frac{k_BT}{\omega} D(\tilde{E}), \tag{2.36}$$

where $\tilde{E} \equiv -k_BT\ln(\omega\tau_0)$. For the purposes of fitting experiments to theory, however, the phenomenological expression provided by Hooge [78] is most commonly used,

$$S(f) = \frac{S_0}{f^\alpha}. \tag{2.37}$$

From the works of Machulp, McWhorter, Dutta and Horn, we interpret $\alpha \sim 1$ as being linked to systems with a large number of TLFs with a sharply peaked energy distribution, and $\alpha \sim 2$ as being representative of systems with few dominating TLFs that may have non-uniform energies.

Charge Noise Experiments

There exist three main ways in which charge noise spectra can be acquired experimentally. The first, and simplest way, is to directly translate the current noise spectrum $S_I(f)$ through an SET/SHT into a charge noise spectrum by normalizing it with the lever arm α and transconductance dI_{sd}/V_G

$$S_\epsilon = \frac{\alpha S_I}{|dI_{sd}/dV_G|^2}. \tag{2.38}$$

The lever arm is a measure of the plunger gate's coupling to the QD with respect to all electrodes in the SET/SHT, and can be acquired through a Coulomb diamond measurement.

$$\alpha = \frac{C_G}{C_\Sigma}, \quad \alpha > 0. \tag{2.39}$$

The second is to implement a dynamical decoupling protocol on a qubit, such as the Carr-Purcell-Meiboom-Gill (CPMG) protocol. This produces a phase coherence parameter A_{CPMG} as a function of a wait time t_{wait} between $\pi-$pulses, which can be converted to a charge noise spectrum via a Gaussian filter function [79]:

$$S(f) \simeq -\frac{\ln A_{CPMG}}{2\pi^2 t_{wait}} \tag{2.40}$$

The third way, best suited for low frequencies, is to track the frequency detuning of a Ramsey fringe with time, relative to some detuning set-point. Fourier transforming this trace yields the energy splitting, or charge noise, spectrum.

A combination of these methods, have enabled the collection of charge noise spectra approaching nine frequency decades between 1×10^{-5} Hz and 1×10^4 Hz. When performed on a

^{28}Si/SiGe QD device, good quantitative agreement was found between the current noise spectra and Ramsey fringe methods. In this study, a transition point in the frequency exponent α was found around 1×10^{-3} Hz from 1.97 to 1.48 in the case of Ramsey fringe spectroscopy and from 1 to 2 using current noise spectra [80]. Remarkably, the measured $T_2^\star$ time faced a similar transition point, as determined by varying the duration over which an ensemble of Ramsey sequences was applied. Similar observations were made in Ref. [81]. If $T_2^\star$ were affected by white noise, which has no dependence on frequency, then its decrease with measurement time would follow an exponential dependence. However, free induction decay (FID) experiments have revealed its dependence to be Gaussian [82], suggesting that qubit dephasing arises overwhelmingly from low-frequency noise contributions. This, together with a transition point in $T_2^\star$ consistent with charge noise spectra, concludes that electric noise dominates qubit dephasing over a broad frequency range. Moreover, it testifies the robustness of simple current noise methods to accurately predict qubit coherence properties, as corroborated by Ref. [83]. In a similar study on ^{28}Si/SiGe QDs, measurements from all three methods aligned on the same charge noise spectrum, but with a constant frequency exponent of 1. While this confirms the robustness of current noise methods further, it highlights the fault in comparing noise figures of single devices – something this thesis hopes to address through statistical grid measurements.

This concern was first raised in Ref. [84], where measured charge noise spectra in neighbouring QDs showed great variability, as well as a deviation from a linear temperature dependence. This suggests that the TLF ensemble has a changing distribution, both in number and in energy, over just a few hundred nanometers. A commonly reported figure of merit is the charge noise at 1 Hz. This was found to vary from 0.59 µeV $\sqrt{\text{Hz}}$ to 1.17 µeV $\sqrt{\text{Hz}}$ in the same device. While a temporal correlation study did not reveal any correlation between the measured noise spectra, a more recent works claims the opposite [85]. Through evaluation of a simulated cross-power spectral density on spatially separated qubits, noise was found to be, infact, correlated by up to ∼10% and to decay only polynomially with distance. This could verify the non-locality measured in qubit frequencies [86] and poses concerns toward implementing future error correction schemes, which rely on qubit operations being un-correlated. With the grid architecture proposed in this thesis, the extent of non-local noise over large distances could be investigated.

In line with these works, the spin qubit community has accelerated efforts toward optimizing material stacks against charge noise. In 2016, SiMOS had reportedly superior charge noise performance with $S_\epsilon(1\,\text{Hz}) = 0.49\,\mu\text{eV}\sqrt{\text{Hz}}$ [1] as compared to identically patterned Si/SiGe which yielded $S_\epsilon(1\,\text{Hz}) = 2.0\,\mu\text{eV}\sqrt{\text{Hz}}$ [87]. However, reducing the gate oxide thickness already brought this value down to 0.84 µeV $\sqrt{\text{Hz}}$ in a later work [84], which can be attributed to a decreased trap density. Moving the QW further away from the dielectric interface in Ge/SiGe showed further improvements, with $S_\epsilon(1\,\text{Hz})$ reaching 0.2 µeV $\sqrt{\text{Hz}}$ in one measurement and averaging 0.6 µeV $\sqrt{\text{Hz}}$ [29]. This was enforced by a smaller percolation threshold that also scaled with the QW depth. The lowest values were met for $d = 55\,\text{nm}$ at $2.14\,\text{cm}^{-2}$, which also highlights the utility of the percolation threshold as a noise qualifier. Finally, by bringing the QW thickness below the Matthew-Blakeslee limit and removing an epitaxial Si cap, charge noise in Si/SiGe heterostructure has been brought down to $S_\epsilon(1\,\text{Hz}) = 0.29\,\mu\text{eV}\sqrt{\text{Hz}}$ levels [30]. A theoretical extrapolation of this value shows that a two orders of magnitude improvement in $T_2^\star$ could be expected from qubits defined in this improved heterostructure. This can be explained by a reduction in remote impurities, as well as a reduction in the number of dislocations in the QW, which is mirrored by a mobility in excess of $2 \times 10^5\,\text{cm}^2/(\text{V s})$ [88]. Worth noting is also the reduced spread in the measured noise levels over multiple devices in this study, which indicates an improved fabrication uniformity – a step in the right direction for scaling quantum processors. In the future, we foresee further improvements in charge noise performance and uniformity through careful material design.

[1] this was using a wafer grown via the float-zone process, which has not yet been tested for SiGe qubits

3
DESIGN

By developing an architecture with many identical devices, one gains statistical insight into the performance of a single device. By standardizing this architecure, and keeping designs similar across material platforms, it also becomes possible to compare the performance between devices fabricated on different heterostructures. In this chapter, we present the designs of two grids, the first suited to hole transport in Ge/SiGe, and the second to electron transport in Si/SiGe. The contribution of this thesis is the adaptation of a pre-existing Ge grid design to Si. Since many design choices overlap between the two material platforms, the design of the Ge grid is explained thoroughly. Subsequently, the Ge grid is successfully fabricated and characterized.

The grid architecture has an overlapping gate structure and is designed to be compatible with previously used cryogenic multiplexers to characterize SiMOS grids [89]. The multiplexer permits a total of 72 gates to be individually controlled, plus another 8 gates with pairwise control. To make each SET/SHT similar in design to charge sensors and QDs found in real spin qubit quantum processors, we place additional barrier gates on either side of the plunger. The barriers, placed vertically across the grid, consume a total of 36 gates, while plungers, connected via common gates placed horizontally across the grid, consume the remaining 36. This produces a total of $18 \times 36 = 648$ devices. The division by two comes from the fact that each device possesses two barrier gates. By biasing any neighbour pair column of barriers and simultaneously any single row of plungers, any individual SET/SHT in the grid may be individually operated. The final designs of the two grids are shown in Figs. 3.1 & 3.3.

3.1 SHT Design

3.1.1 Ohmics

The basic requirement for connecting an ohmic reservoir to a semiconductor is to use a material with high bulk conductivity and low contact resistance. The contact resistance, governed by the barrier height at the metal-semiconductor junction, is determined by the difference in their work functions. For hole conduction, forming an ohmic contact requires that the work function of the semiconductor ϕ_m be larger than that of the metal ϕ_s, and vice versa for electron conduction. In this way, when Fermi-levels equilibrate, the semiconductor's valence band (or conduction band in the case of electrons) bends toward the metal's Fermi-level by an amount $\phi_m - \phi_s$. This is the same direction in which band bending occurs when free charge carriers are accumulated in the semiconductor. Hence, no sizable conduction barrier exists across the metal-semiconductor junction, regardless of the bias on the metal.

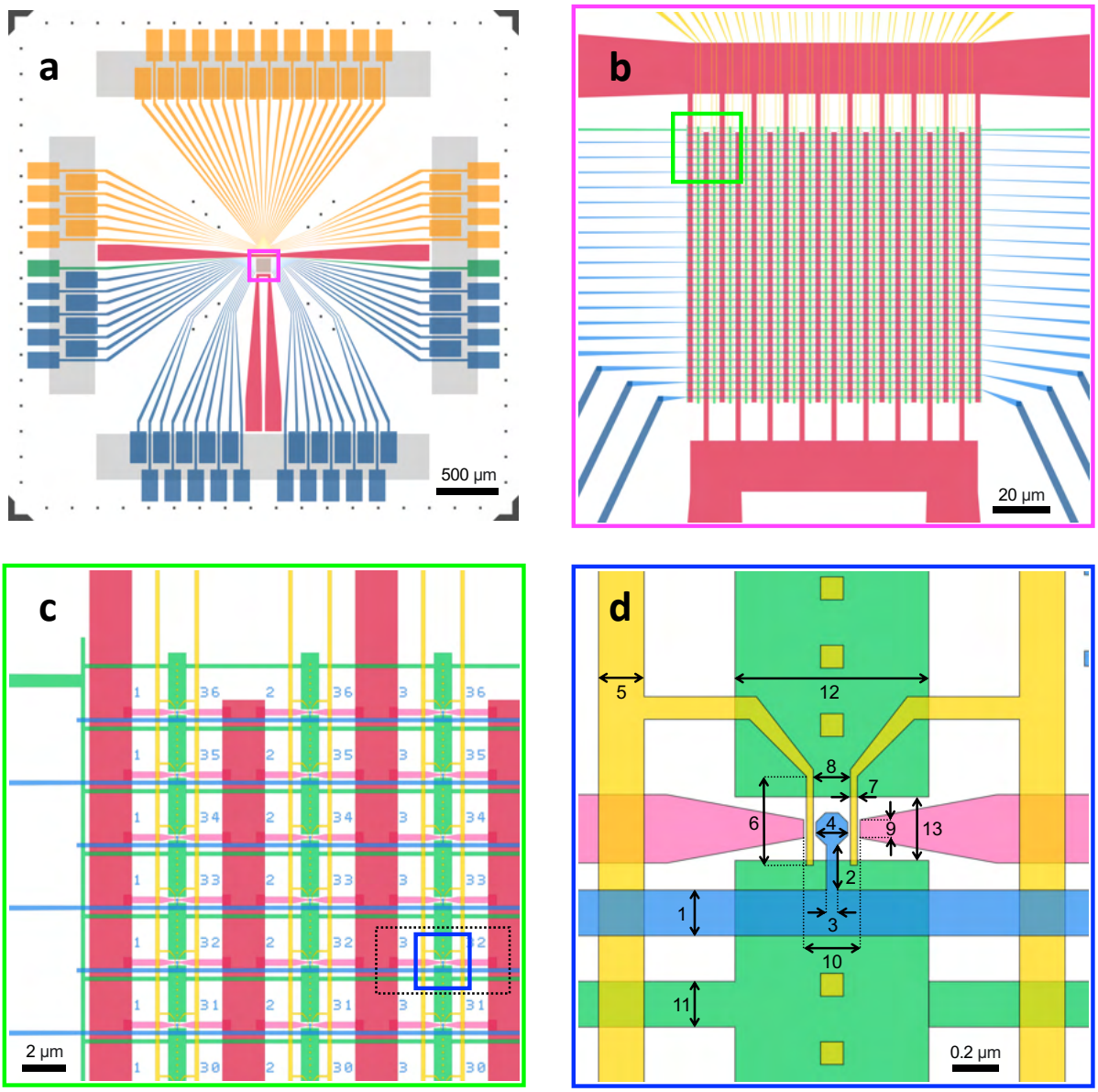


Figure 3.1: (a – d) Design of the single-hole transistor grid over four length-scales. The various colors correspond to the barrier gates (orange/yellow), plunger gates (dark/light blue), screening gate (dark/light green), diffused ohmic leads (dark/light red), SiN (grey), and alignment markers (black). Darker and lighter colours correspond to rough structures, that can be resolved by photo-lithography, and fine structures, that can only be resolved by e-beam lithography. Labeled dimensions can be found in Table. 3.1.

For hole conduction in Ge, ohmic contacts are easy to implement due to the pinning of Ge's valence band to the Fermi-level of most metals [90]. Pt metal was chosen, as this has been shown to work well in previously studied Ge/SiGe quantum devices [91, 92]. To make contact with the QW, a Rapid Thermal Processing (RTP) process is used, which works by activating diffusing of a surface-deposited metal through the heterostructure. This forces the ohmic regions to be the first layer in the fabricated flow (red). The alloying of Pt with SiGe and Ge leads to the formation of a Pt-germanosilicide and Pt-germanide, both of which exhibit metallic behaviour. Moreover, the moderate annealing temperature around 400 °C is compatible with the low thermal budget of SiGe heterostructures, limited by strain relaxation and diffusion in the QW [93]. Only Ni and Pd metals also form stable and low resistance alloys with SiGe and Ge at low temperatures

[94]. In NiSiGe, however, the Schottky barrier height is 34% higher than in PtSiGe [95], while also having inferior thermal and morphological stability [95, 96].

Since individual device selection is already made possible by manipulating plunger and barrier biases alone, the same source and drain leads can be shared by all devices. By connecting an ammeter in series with one of the leads, the current through any one of the 648 devices can be readily measured. This mode of operation is realized by using an interdigitated design. Each device lies between the fingers of any two ohmic regions, as seen in Fig. 3.1 b). The upper ohmic region (source) connects to one of the 4 shared control lines, while the lower ohmic region (drain) connects to a second pair of shared control lines. We note, however, that this design causes the polarity to switch between each finger in the grid, hence the direction of the current between each column of devices. For example, consider the first two columns of devices in Fig. 3.1 c). If the upper ohmic is positive, the first finger is positive, and so the hole current will flow from left to right. For the second column, however, the right finger is positive, and the left finger is grounded, hence, the current flows from right to left. This was done to shrink down the grid design and reduce the chances of fabrication errors, as it requires half as many electrodes.

3.1.2 Screening Gate

After the ohmic regions are diffused, the remaining gate layers can be fabricated. The order in which they are deposited influences the level of electrostatic control they have over the potential landscape in the QW. As the first gate layer, we introduce a screening gate (green), whose purpose is manifold. The primary use is to screen the electric field from the overlying plunger to the QW. This is achieved by applying zero or small biases to the screening gate, such that electric field lines become redirected from the plunger. In this way, the screening gate cuts off an otherwise continuous path of accumulated holes in the region where the gates overlap. This is important, because the plunger overlaps with the ohmic regions away from the QD, so if the accumulated regions were not cut, then there would be no way of isolating the current contribution flowing across the SHT. In the event that the applied barrier bias is sufficiently strong to itself accumulate holes in the QW, the region of overlap between the screening and barrier gate cuts this accumulated region off too. Without the screening gate, various alternate current paths would exist, as illustrated in Fig 3.2.

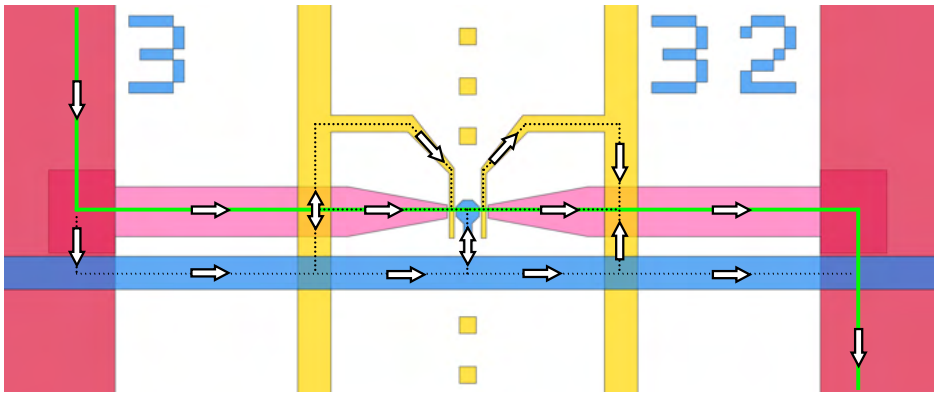


Figure 3.2: Alternative hole current paths in an SHT device in the absence of a screening gate. The green line indicates the desired current path.

The screening gate also functions as an additional degree of control over the potential landscape surrounding the QD. When biasing the screening gate negatively, it can serve to broaden the confinement potential of the QD, increasing its size and coupling to the ohmic reservoirs. The screening gate can also oppose any stray fields in the surrounding QD environment, such as

due to defects, hopefully improving device stability. As with the ohmic regions, it is controlled globally and connects to the third pair of shared lines of the cryo-multiplexer.

3.1.3 Dielectric, Plungers, and Barriers

To electrically isolate the multiple overlapping gates, as well as the diffused ohmics, Al_2O_3 is deposited in between each metal layer across the whole grid. It is chosen mainly for its compatibility with atomic layer deposition (ALD) processing, which yields a high quality and atomically smooth interface – something not possible with oxides grown via chemical vapour deposition, such as conventional SiO_2. An added benefit of Al_2O_3 is that it has a high dielectric constant, which produces a large gate-to-well capacitance, reducing the chances of leakage. This is especially important provided the thin layer of dielectric being used, which is necessary to produce sharp and highly tuneable confining potentials for the QDs.

After depositing the screening gate, there is some freedom of choice as to which gate to deposit next. Since the barriers have a smaller critical dimension, however, we choose to deposit them first (yellow). The reason for this is that their small size makes them more susceptible to misalignment. If they were deposited above the plunger, screening may occur if the gates overlap, making it difficult to regulate the coupling between the ohmics and the QD. In qubit devices, it is also desired to have a high degree of control over the tunnel couplings, therefore the barrier gates are deposited at this same level [59]. Furthermore, small gate sizes make electrostatic control less forgiving to defects. Therefore, by depositing the barrier first, there will be fewer dielectric layers and interfaces between it and the QW, over which defects can pile up. Finally, the barriers set the critical dimension of the entire grid, which poses a fabrication bottleneck. Hence, a practical reason for depositing barrier gates first is that their failure can be caught earlier on in a fabrication run, saving time.

In the past, the lift-off process during barrier fabrication had been an issue due to the incomplete removal of e-beam resist. Therefore, 'dummy structures' have been added near the fine structure of the barriers, as seen in Fig. 3.1d. These islands of metal, located below the resist surface, provide pathways for the chemical used for lift-off (AR 600-71) to reach the narrow barriers. Without them, the contact angle of the AR 600-71 with the resist is too large to directly wet the resist sidewalls at the barrier sites. It is for this reason, as well as fracture issues, that

Table 3.1: List of critical dimensions of the final SHT grid design. Original dimensions were adjusted based on preliminary fabrication runs. Labels correspond to those depicted in Fig. 3.1 d).

Label	Feature	Final Dimension (nm)	Original Dimension (nm)
1	Plunger row width	200	150
2	Plunger stem length	200	250
3	Plunger stem width	50	60
4	Plunger head width	140	148
5	Barrier column width	200	150
6	Barrier length	390	420
7	Barrier width	30	40
8	Barrier separation	163	156
9	Ohmic width	80	80
10	Ohmic separation	247	244
11	Screening row width	200	150
12	Screening width	800	460
13	Screening separation	280	188

the barrier columns that span the full grid are made wider than the fine structure near the QD location.

The last gate to be deposited is the plunger. One region of interest here is the 'neck' that leads to the 'head' of the plunger. It is made short enough such as to reduce fabrication error, but long enough to give sufficient separation between the ohmic regions (pink) and the shared plunger line (blue), while also giving space for the screening gate (green), as seen in Fig. 3.1d. A summary of the critical grid dimensions can be found in Table. 3.1. To make the grid a relevant benchmark of the material and charge transport uniformity expected in real quantum processors, the dimensions of the barriers and plungers are inspired by Refs. [6, 59]. To meet these size targets, most dimensions ended up being revised after the first fabrication run. These findings are discussed in Chapter 4. The original design of the grid's repeat unit may be found in Fig. A.1.

3.1.4 Bonding and Alignment

The gate lines leading from the grid to the edges of the coupon are known as the 'fan-out' (Fig. 3.1a), and connect to a printed circuit board (PCB) by means of aluminum wire bonding. The PCB then connects to the cryo-multiplexer through a ribbon cable. The rectangular regions at the die edges are the bonding pads, which are staggered in order to fit them within the die dimensions. An insulating SiN layer (grey) is deposited during a pre-fabrication step. It is used to elevate the bond pads, such that bonding wires do not pierce through the heterostructure and potentially down to the QW. This is opposite to the ohmic regions (red), which are meant to make electrical contact with the QW, and are therefore positioned on the inner side of the SiN. As for the outer bond pads, which may indeed pierce through to the QW, this is not of concern since the elevation on the gate lines provided by the SiN renders any continuous charge accumulation to the piercing location impossible.

Finally, the metal squares (black) placed around the die edges are used for alignment during the multiple resist exposure steps, done using electron-beam lithography. The outer markers are used for the rough structures, and the inner markers are for the fine structures. The fine structure markers are re-used by plunger and barrier layers to minimize their misalignment. A continuous line of markers at the die edges is deposited to serve as a visual aid and help maximize the precision with which slant corrections are made.

3.2 SET Design

The design of the SET grid closely resembles that of the SHT grid, with a few key differences imposed by the formation of the ohmic contacts. In the microelectronics industry, RTP on Si is known as the silicide process and commonly uses low work-function metals such as Co, Pd, Pt, Ti, Ni. The most widespread silicide is $TiSi_2$, due to its low resistivity and good thermal stability [97]. Recent studies have also shown that low contact resistances can also be achieved to SiGe when using a Ti germano-silicidation process [98, 99], making the idea of using Ti as an ohmic material in our Si/SiGe heterostructure promising. Unforunetly, however, the Schottky barrier height is usually sizeable at pristine silicide-silicon junctions, and therefore requires p-doped Si substrates. This would not work in our Si/SiGe heterostructures, because electrons are the majority charge carriers, rather than holes. This necessitates the formation of ohmic regions by donor doping, rather than metal diffusion.

There exist two ways to introduce impurity atoms (dopants) into semiconductor materials: diffusion and implantation. Since diffusion is incompatible with the thermal budget of Si/SiGe, implantation must be used. In line with previous work, P^+ ion implantation is chosen [5, 6]. Implantation works by accelerating ions through a series of magnetic and electrostatic columns

such that they gain enough kinetic energy to penetrate the surface of their wafer target. The depth profile of the implanted region is controlled via the acceleration voltage, as well as a subsequent annealing step, which serves to repair the sample from the ion bombardment and to substitute the dopant atoms in the crystalline lattice. This substitution process is called *activation*.

Due to the high-energy collisions of the P^+, a thicker resist, typically greater than 1 µm, is needed to protect the substrate. However, thick resists can not be used to draw small features. Consequently, the fine ohmic structure, shown in pink in Fig. 3.1d, would become impossible to resolve. This forces the conductive path between the interdigitated ohmics and the QD to be completed using an additional accumulation gate (lilac/magenta). The geometry of the accumulation gate is kept simple. For the most part, it traces the route of the implanted region

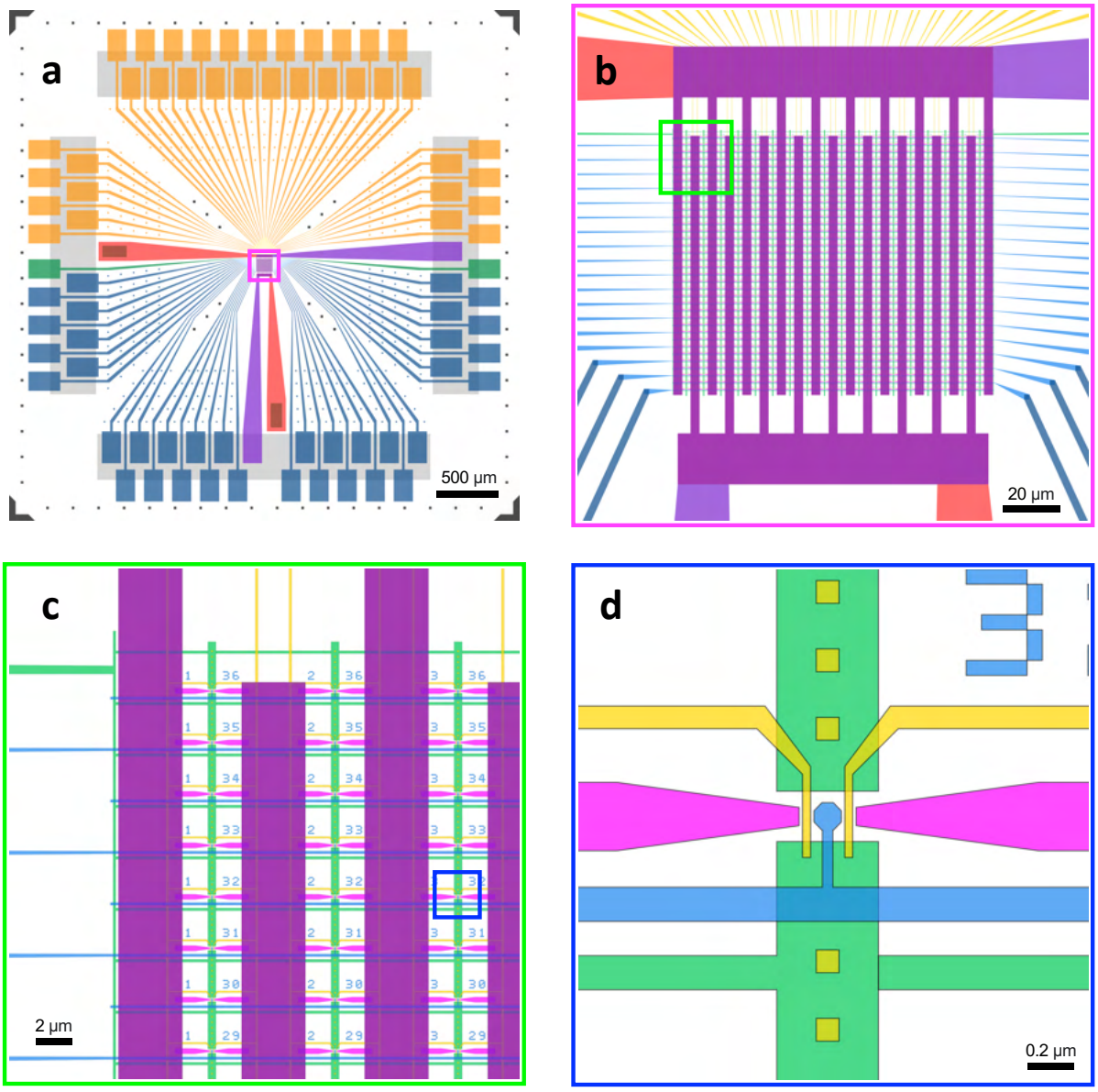


Figure 3.3: (a – d) Design of the single-electron transistor grid over four length-scales. The various colours correspond to resist layers for the barrier gates (orange/yellow), plunger gates (blue), screening gate (green), implanted ohmic leads (red), ohmic contact (brown), accumulation gate (lilac/magenta), SiN (grey) and alignment markers (black).

up until the length-scale shown in Fig. 3.3b, where the two electrodes split off to their respective bond pads. Luckily, since the SHT design used only 3 of the 4 shared control lines of the cryo-multiplexer, this leaves one remaining control line that can be used by the accumulation gate. The bond pads of the implanted regions feature additional contact metal on top (brown) for improved electrical contact.

For comparison's sake, the depth of screening, barrier and plunger gates is kept constant, with the accumulation gate deposited on top. On the downside, however, this creates a problem with the barrier placement, because where the accumulation gate overlaps with the barrier, the barrier will screen the electric field of the accumulation gate, cutting off electron accumulation in the QW. This necessitates moving the barrier column beyond the edges of the implanted regions, as seen in Fig. 3.3c. In this way, there is continuous accumulation between the QD and the implanted region. To maintain a sufficient distance between neighbouring barriers to avoid problems with fan-out, the implanted region is made wider while keeping the approximate unit cell size the same.

We point out some minor improvements in the SET grid that did not make it to the final design of the SHT grid. Firstly, more broadly spaced bond pads to ease wire bonding. Secondly, additional dummy structures between the fan-out to encourage lift-off. Thirdly, a narrower screening gate and higher aspect ratio of the accumulation gate in the fine structure, to alleviate concerns over possible tunneling given the small distance between these two gates near the QD. Finally, greater separation between the accumulation gates, barriers, and QD, to prevent gate overlap in the case of misalignment.

We also list some improvement points for future design iterations of the SET grid. For starters, it may also be worthwhile to add connectivity between the accumulation gate, in a similar way to the screening gate, so as to homogenize the accumulation potential across the grid. Also, the accumulation gate could be deposited below the barrier and plunger layers. In this way, the barrier positions from the SHT grid design can be maintained, and less voltage stress would be placed on the dielectric due to the accumulation gate's better proximity to the QW. Alternatively, keeping the barriers where they are, implantation could stop before the interdigitated region, which could be advantageous for reducing the number of defects introduced via ion bombardment near the QD. Lastly, given the larger effective mass of electrons compared to holes, one will have to reduce the plunger dimension to have the QD's lowest-lying energy levels at comparable voltages between SHT and SET devices.

3.3 Design Automation

Grid designs were generated with the help of the gdspy Python module [100]. The gdspy module offers a way to convert user-defined polygons into a GDSII file, text labels, and other information about mask layouts in hierarchical form. The module works by creating a design library `gdspy.GdsLibrary()`, which holds one or more cells `gdspy.Cell()`. It is to these cells that polygons may be added using the cell's `add` method. When doing so, the layer may be passed as an argument, which keeps track of the design's hierarchy.

Various methods exist for creating polygons, including the built-in functions `gdspy.Rectangle()`, `gdspy.Circle()` and `gdspy.Curve()`. To make flexible design code, however, one should not be constrained to specific polygon shapes. Instead, it is best to parametrize the design to a set of distances between key features in each layer. One parameter, for instance, is the barrier width, from which its vertices are then calculated. To generate the barrier's polygon, the vertices are merely connected together, which can be done using the `gdspy.boolean()` function. This is the only geometric operation used in the entire design code. Since an arbitrary number of parameters can be introduced to add more or fewer vertices, the final polygon can be modified without having to choose its shape beforehand.

The generation of vertices for each design layer is handled by user-built functions such as `get_plunger_vertices` and `get_barrier_vertices`. These functions generate the vertices for a single repeat unit of the grid, and store them in a dictionary with keys corresponding to the layer index as {`layer:[vertices]`}. Using a for-loop, which iterates over all rows and columns of the grid, these vertex functions are called a total of 648 times, each time with a different x- and y- offset to account for the position of the repeat unit in the grid. The resulting vertex dictionary from each of these function calls is appended to a larger dictionary containing all the vertices for the entire grid, called `polygons_map`. As a final step, this dictionary is passed to `gdspy.boolean()`, and finally to `gdspy.Cell.add()`. Condition statements take care of the vertex locations at the edge rows and columns, which have slightly different geometry. A similar procedure to the one described above is carried out for the fan-out, though it is a bit more cumbersome due to its lower repetitiveness. A summary of the code flow can be found in Fig. 3.4.

The full code used to generate the SHT/SET grid design may be found on GitHub under: VandersypenQutech/spin-projects/tree/master/users/cornelius. In this same repository are the relevant scripts used to process the data presented in Chapter 5.

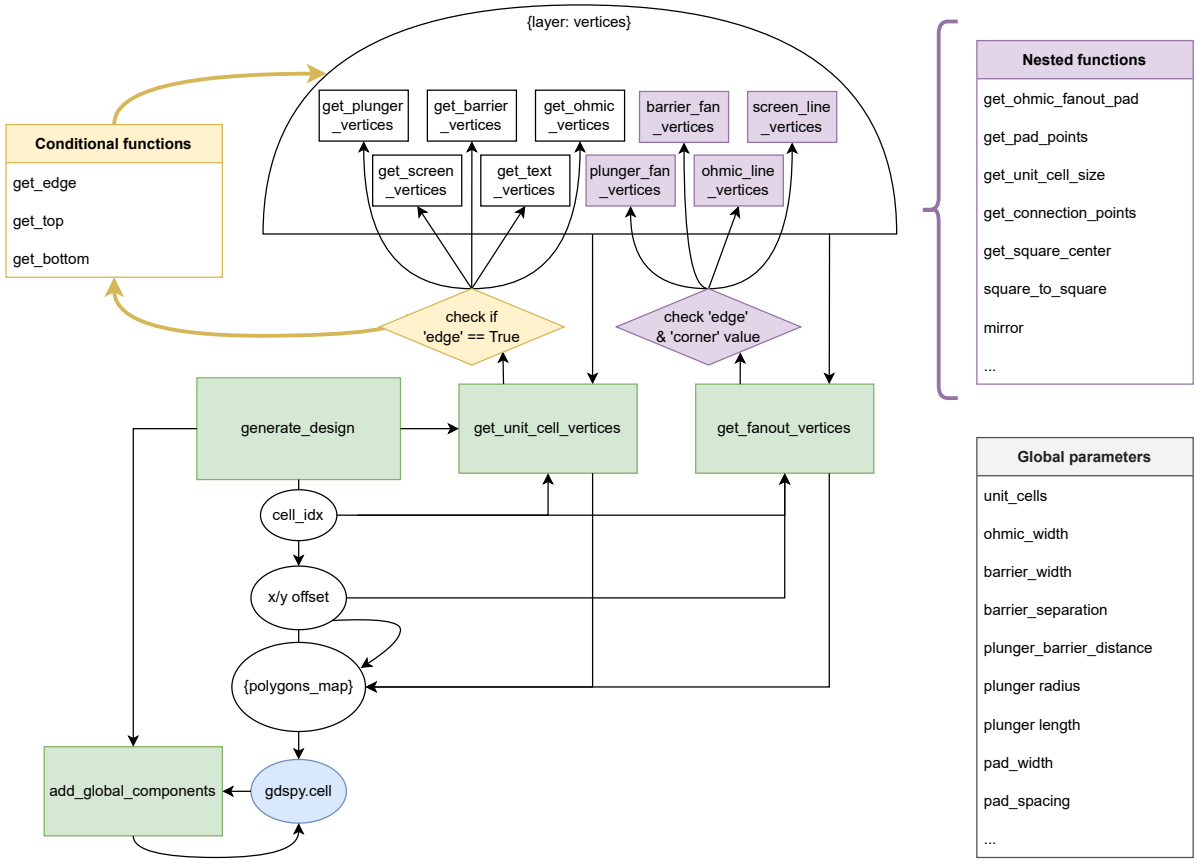


Figure 3.4: Flowchart of the code used to automatically generate .gds files containing polygons and hierarchy information for all mask layers in the SHT grid. The SET grid follows the same flow, with some adaptations. Functions are shown as squares, conditions as diamonds and parameters/variables/objects as circles. The code begins by executing `generate_design`, which retrieves grid vertices from `get_unit_cell_vertices`, fan-out vertices from `get_fanout_vertices` and SiN and marker vertices from `add_global_components`.

4

FABRICATION

Following the design presented in Chapter 3, we now walk through the fabrication of the SHT grid. Due to time constraints, the SET grid was not fabricated and is left as a project in further work (Sec. 6.2). The recipe builds on the backbone of years of process refinement in collaboration with colleagues at TNO – Netherlands Organization for Applied Scientific Research. Therefore, rather than giving details on all involved processes, attention is brought to the lessons learned between fab-runs, which ultimately lead to the successful fabrication of the SHT grid in Ge. This grid is characterized in Chapter 5.

4.1 Recipe Overview

The heterostructure growth and fabrication recipe closely follows the one found in Refs. [6, 27]. These same recipes have enabled charge sensing for the first time in a two-dimensional array of five QDs in Si/SiGe, as well as universal control of a four-qubit quantum processor in Ge/SiGe. The same Ge/SiGe heterostructure is used as in Ref. [29]. It consists of a 1.4 µm thick layer of Ge deposited on a 4 inch p-type natural Si wafer, followed by 900 nm of a reverse-graded virtual substrate of $Si_{1-x}Ge_x$, with x ranging from 1–0.8. On top of this is grown a 160 nm $Si_{0.2}Ge_{0.8}$ spacer layer, followed by a 16 nm compressively strained Ge QW, a second 55 nm thick $Si_{0.2}Ge_{0.8}$ spacer layer, and finally a thin Si cap of 1 nm. All growth is carried out using reduced pressure chemical vapor deposition (RP-CVD). Qualitatively, this heterostructure is the same as the one shown at the bottom of Fig. 2.1. After the heterostructure is grown, Ti:Pt e-beam markers and SiN are deposited in an array across the wafer, marking coupons of 20 × 20 mm, which are subsequently cut from the wafer. This concludes the so-called *pre*-fabrication stage. The process described henceforth constitutes the fabrication stage conducted during this thesis.

Each coupon contains 9 dies, as shown in Fig. 4.1, measuring 4 × 4 mm. Four of the dies are used to fabricate four SHT grids, indexed A1, A2, B1 and B2, and one die is reserved for test structures, indexed A3. In our case, die A3 holds three hall bars. The hall bar design and dimensions are provided in Fig. A.2. The ohmics of the hall bars are diffused at the same time as the ohmics of the grid, and the gate layer is deposited at the same time as the grid's screening layer. In this way, hall-bar measurements can be directly correlated with grid measurements, as the devices come from the exact same fabrication run. The remaining 4 dies, indexed B3, C1, C2, and C3, contain grids with a different fan-out geometry, which are used by our collaborator group to test new versions of cryo-multiplexers.

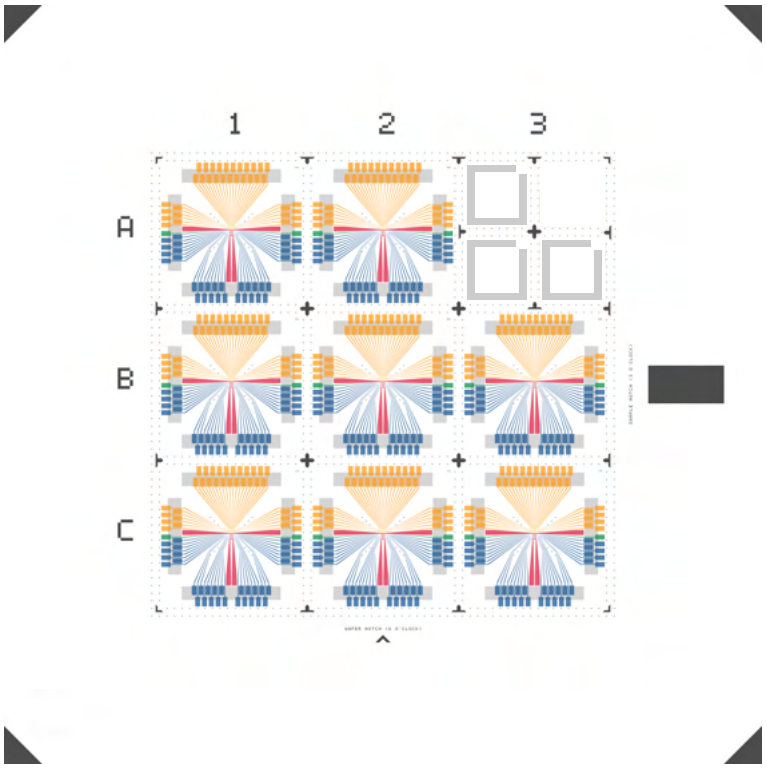


Figure 4.1: Fabrication layout of a single coupon. Dies containing the described SHT design are located at indices A1, A2, B1 and B2. Die A3 contains Hall bar structures, which designs shown in Fig. A.2. The notch on the right-hand side is used to orientate the coupon before loading into the electron-beam lithography machine for resist exposure. A sufficient border is given around the devices for ease of handling and to avoid non-uniformities of the resist close to the edges.

Directly before depositing Pt ohmics, the coupon is etched with a buffered hydrofluoric acid solution to remove any natural oxide that may have formed on the wafer surface. The diffusion of the ohmics is performed by RTP at 400 °C for 15 minutes. Al_2O_3 is deposited via Atomic Layer Deposition (ALD) at a thickness of 7 nm (70 cycles) to isolate the ohmics and at 5 nm (50 cycles) after every subsequent gate deposition step. This is done to encourage a higher gate lever arm, and similar lever arms between gates. Each gate begins with the deposition of 3 nm of Ti for adhesion, followed by 17 nm of Pd, both using electron-beam evaporation. Pd is used due to its small grain size, which permits smaller gate feature resolution and good chemical stability. Commonly used gate metal Al has been shown to oxidize and incorporate oxygen atoms below the material surface, thus potentially introducing compressive stress on underlying layers that could lead to spurious QD formation [101]. The thickness of Pd in each subsequent gate after the screening is increased by 10 nm. This is done to assist with *gate climbing*, because, with overlapping gate structures, it becomes increasingly difficult for newly deposited gate metals to conform to an uneven substrate created by the underlying gates. All gates are patterned using e-beam lithography with CSAR 6200.04 as a positive resist, chosen for its high resolution ($<$ 10 nm) and ease of lift-off [102]. Resist development is done in AR 600-546 (penthyl acetate), and lift-off in AR 600-71 (dioxolane), both followed by a cleaning step with isopropyl alcohol (IPA). To protect the wafer surface during dicing, a layer of S1813 resist is spin-coated, which is stripped off on each die afterward by immersion in acetone.

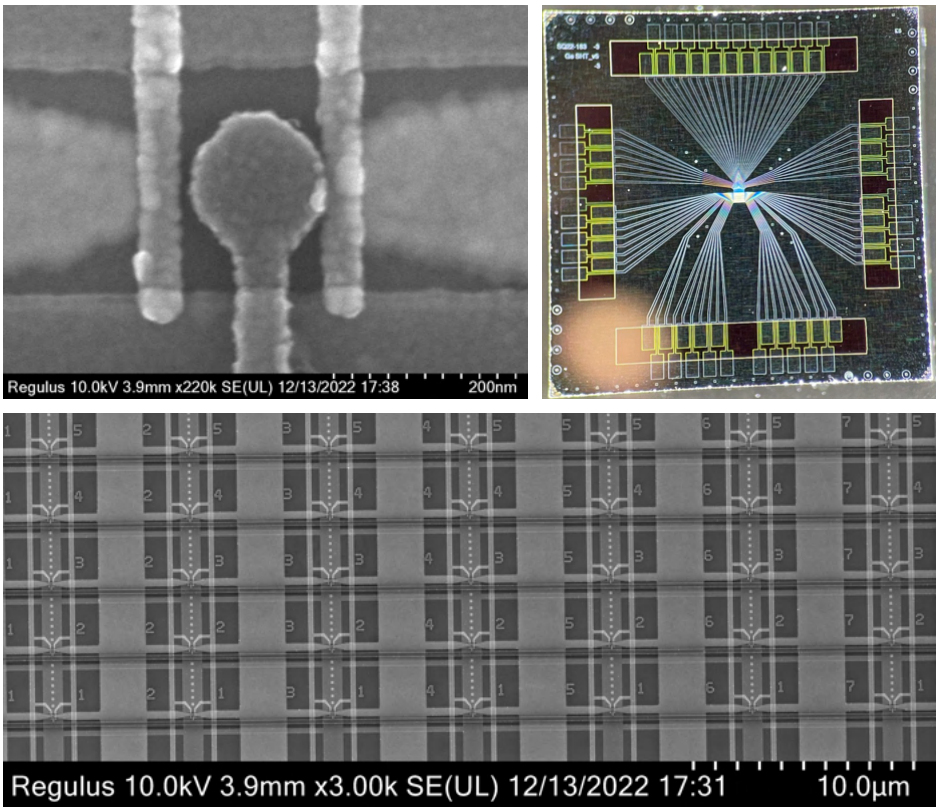


Figure 4.2: Microscope images of the final SHT grid at three length scales: single SHT device (top left), array of SHT devices (bottom), entire grid (top right).

4.2 Fabrication Errors

4.2.1 Feature Size

As alluded to in Chapter 3, an issue encountered during fabrication was that grid features were consistently larger than in designs. We set out to investigate this with the help of Atomic Force Microscopy (AFM), conducted at regular points in the fabrication process. We cross-examine these images with images made by Scanning Electron Microscopy (SEM). To illustrate our findings concisely, we focus on ohmic and screening separation as key metrics. Another quality control method used is Visible Light Microscopy (VLM), which has low enough magnification but high enough resolution to see the entire fan-out of the grid, and is therefore used to quickly and preliminarily verify that resist spin-coating, development, and lift-off steps were successful. The grid on die A2 is used as a sacrificial grid for all characterizations, such that three pristine grids could be used for measurements.

In Fig. 4.3, we present SEM images of randomly selected SHT devices on die A2. We usually capture images in Back-Scattered Electron (BSE) mode, which generates contrast based on the atomic weight of elements in the sample. This allows us to image below the sample surface while still resolving feature edges. Those regions with heavier nuclei interact with electrons more strongly, thus appearing brighter in the image as more electrons deflect toward the BSE detector, located directly above the sample. This is why the ohmic regions, made of Pt with atomic mass $\approx$195 u, appear brighter than the screening and barrier layers made of Pd, with atomic mass $\approx$106 u. Due to the interaction volume being sufficiently deep, electrons can penetrate through the thin Al_2O_3 layer, making the ohmic and screening layers still visible even after barrier lift-off.

This is useful to see the extent of overlap between the various metal layers.

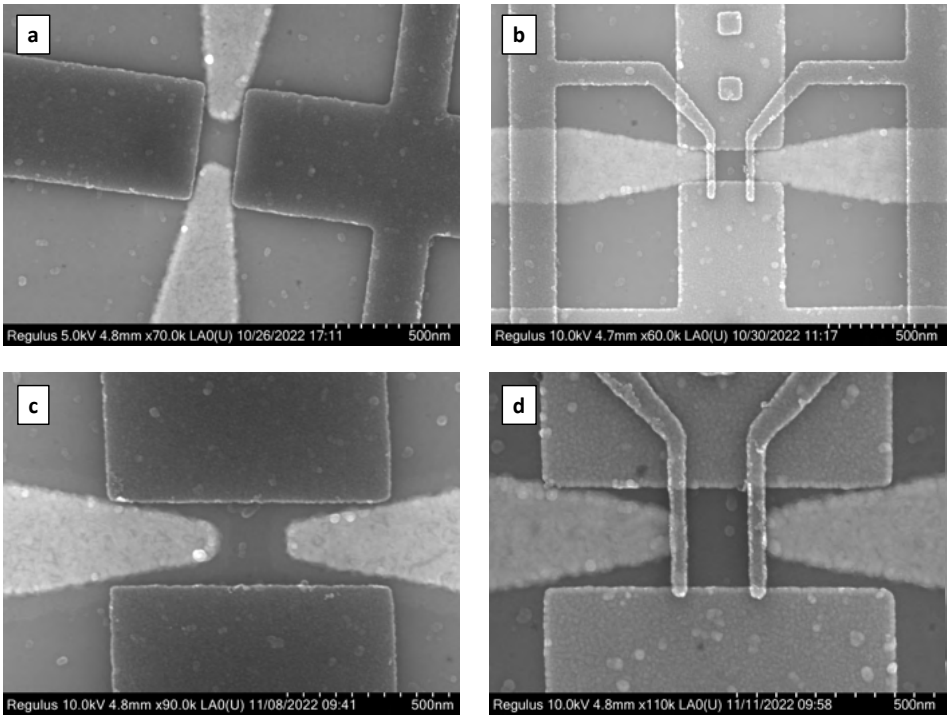


Figure 4.3: SEM images of an SHT device after lift-off of the **a/c)** screening layer and **b/d)** barrier layer during fabrication run 1 (top row) and 2 (bottom row)

As seen from the SEM images above, the extent of overlap between the various layers in the first fabrication run is larger than expected (Fig. 4.3 top row). Visually comparing the micrographs to Fig. A.1, it is clear that the screening gates are closer to one another than designed and that the barriers potentially overlap with the underlying ohmic contacts. This is problematic for SHT operation because the ohmics would screen the electric field from the barriers, while all three layers (ohmics, screen, and barriers), could screen the plunger. This limits the extent of electrostatic control over the intended QD region, thus requiring stronger gate biases to be applied. This is undesirable, as it puts additional electrostatic stress on the dielectric, provoking device instability, drift, and degradation.

We quantify the size discrepancy of features in Table 4.1. Where data is available, we provide the same dimensions as measured with AFM. There is great consistency between dimensions measured by AFM and SEM after lift-off, bringing a high level of confidence to both measurement techniques. SEM images used for dimensional characterization not presented in this Chapter may be found in Appendix B.

Raw AFM data is collected using the ScanAsyst-Air HPI probe tip. This allows for ultra-high-resolution imaging over both hard as well as soft polymeric samples, made further possible by a small nominal tip radius of just 2 nm [103]. As a result, samples can be imaged after both resist development and after lift-off with high accuracy. By studying the sample topography after each of these steps, the root cause of the sizing and other fabrication issues can be better understood. To extract feature dimensions, the raw image file is leveled using the plane subtraction method. This compensates for any piezoelectric or mechanical drift on the AFM. Then, line cuts are drawn over the image from which depth profiles are collected. These are presented in Fig. 4.4. Lastly, dimensions are measured by taking the difference between the rising and falling edge positions of the depth profiles.

Several features in Fig. 4.4 are indicative of a successful resist coating and development

Table 4.1: List of critical dimensions in SHT devices measured over various fab iterations. Original dimensions were adjusted based on preliminary fabrication runs. For barrier widths, and where multiple high-resolution images of single devices exist, averages are taken. All dimensions are in nanometers.

Feature	Target	Fab Run	Design	After Development (AFM)	After Lift-Off (AFM)	After Lift-Off (SEM)
Plunger head width	148	2	140	-	-	150 (+2)
		3	140	-	-	148 (+8)
		4	139	-	-	149 (+10)
Barrier width	40	1	40	-	-	38 (-2)
		2	30	33 (+3)	-	47 (+17)
		3	30	-	-	65 (+35)
		4	30	-	-	47 (+17)
Barrier separation	156	1	156	-	-	142 (-14)
		2	164	165 (+1)	-	154 (-10)
		3	164	-	-	127 (-37)
		4	163	-	-	153 (-10)
Ohmic separation	244	1	244	-	179 (-65)	179 (-65)
		2	248	250 (+2)	212 (-36)	216 (-32)
Screening separation	188	1	188	177 (-11)	-	149 (-39)
		2	280	267 (-13)	250 (-30)	252 (-28)
		3	279	-	-	246 (-33)
		4	280	-	-	246 (-34)

process. Firstly, we determine the average resist thickness by taking the height differences between the developed and undeveloped regions at each interface, which come to 85.3 ± 2.7 nm and 98.0 ± 4.0 nm for ohmic and screening layers, respectively. This complies with the expected thickness of the CSAR 6200.04 e-beam resist when spin-coated at 4000 rpm for 1 minute [102]. Secondly, a flat base in the depth profile indicates that exposure and development processes ran until completion. Thirdly, the sharp interfaces seen both in the AFM images and depth profiles are markers of good resist selectivity and AFM tip health, while the absence of any cracks informs us that the resist baked correctly.

It comes as no surprise, therefore, that the dimensions measured after development in Table. 4.1 are very close to the design specifications. On the other hand, the dimensions after lift-off tend to deviate significantly, which suggests that sizing issues most likely arise from poor metal evaporation, as verified by the variability and roughness seen over the depth profiles in Fig. 4.4d/h. This contradicts the common belief that sizing issues come from a proximity effect during resist exposure. The exact origin of the metal widening, however, is unknown, though a couple of hypotheses are provided. One hypothesis is that metal collisions cause fragmentation of the resist side walls during evaporation. Since the deposition rate is already low, however, the extent to which this occurs would also be small, so we deem this unlikely. Another option is that the surface diffusion of hot metal atoms causes the deposited features to broaden. However, this contribution is likely also small, because we do not apply heat to the coupon base plate, which would encourage diffusion. Moreover, if considerable self-diffusion occurred, then metal atoms would have re-arranged themselves to form a smooth surface layer, which is not seen in the depth profiles. The final option we think of is that the e-beam resist is undercut, which is undetectable by AFM. This may occur from electrons scattering more as they penetrate deeper into the resist, especially if they also meet the substrate/resist interface. Assuming the deposited metal would conform along this undercut region, then we would expect metal features to have

slanted side walls, though this is not obvious from the depth profiles either.

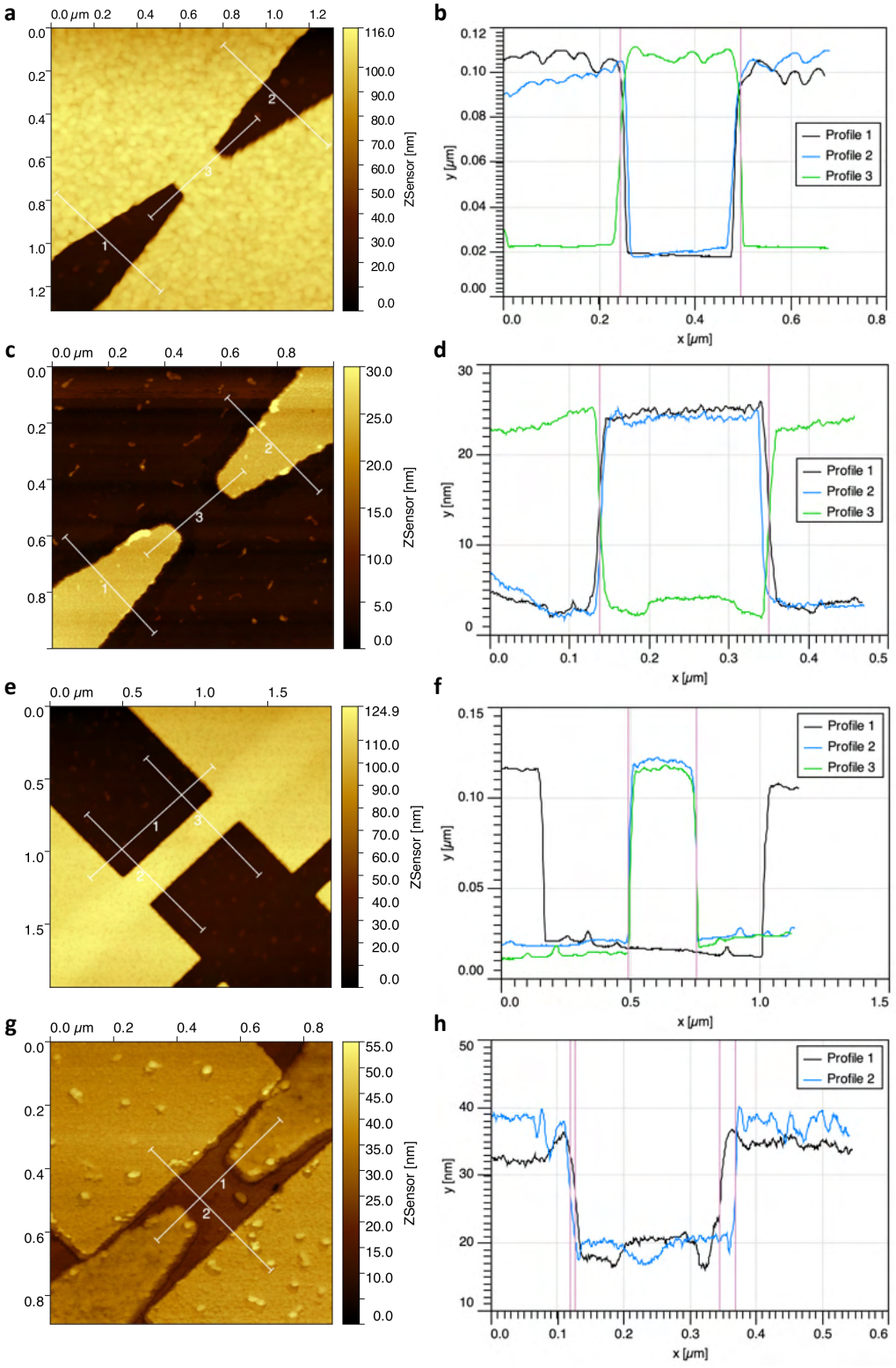


Figure 4.4: AFM images (left) taken after the development of e-beam resist and after lift-off on the **a/c)** ohmic layer and **e/g)** screening layer, alongside depth profiles of the displayed line-cuts (right). Cap widths on the line-cut endpoints indicate the width over which the depth profile is averaged. Red lines on the depth profiles indicate points between which dimensions were collected. All images are collected from fab run 2.

Due to the consistency and reproducibility with which all layers were oversized, however, we treat this error as systematic and deemed it most efficient to fix by undershooting the dimensions at the design stage. While fabrication parameters could have been tweaked, such as the e-beam exposure dose, spot size, development time, beam currents, spot sizes, and accelerating voltages, without knowledge of the root cause, this would be a cumbersome process and infeasible in the short time span of this thesis.

The first revision of key grid dimensions, highlighted in Table. 3.1, was implemented in the second fab run, and compensated well for the sizing errors we first encountered. Fig. 4.3 (bottom row) shows SEM images from this second fabrication run, and visually the SHT device looks more akin to the original design in Fig. A.1. We note that with an increased screen separation, we could afford to make the screening gate wider, which is why this dimension changed also. A wider screening gate has the benefit of more robustly cutting off unwanted hole accumulation. Upon further refinement of the fabrication process, by the fourth fab run, we were satisfied with the yield and grid dimensions to conclude the fabrication phase.

We also characterize the ohmic (before diffusion) and screening gate thicknesses in a similar fashion to the resist, and find average values of 19.9 ± 1.5 nm and 17.0 nm, which (considerably) undershoot the targets of 30 nm and 20 nm. We believe that the under-deposition of metal is unrelated to the issue of feature sizes and can be explained by an incorrect calibration of the tooling factor. The tooling factor is the ratio between the actual evaporation rate and the maximum theoretical evaporation rate. With machine wear, the tooling factor may become overstated, leading to thinner metal films being deposited.

4.2.2 Barrier Fracture

A further problem we encountered during fabrication was the fracture of barrier gates. As seen in Fig. 4.5, at least one barrier of almost all SHT devices is either contorted or entirely missing. From Fig. 4.5 inset, we can see a footprint of where the left barrier used to be, which informs us that the metal made contact with the substrate during deposition, but then broke during

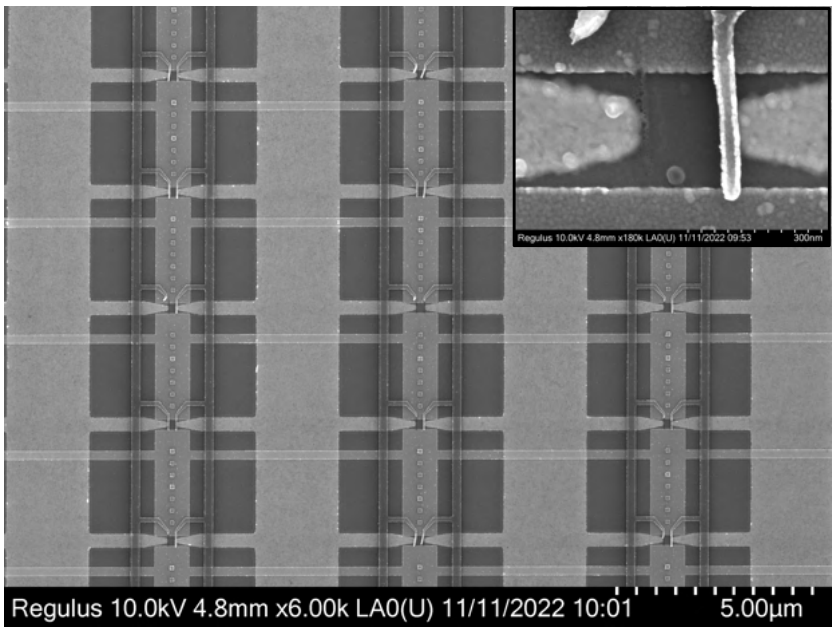


Figure 4.5: SEM image of an array of SHT devices from the second fabrication run after barrier lift-off. Numerous barriers have fractured, are missing, or have lifted up (brighter appearance). The inset shows a close-up of one device where the footprint of the original barrier position can be seen.

lift-off.

Several things could have weakened the adhesion between the barriers and the substrate. We first suspected that this was a one-off issue caused by insufficient target metal left in the e-beam evaporator crucibles. When crucibles are almost empty, the deposition rate uniformity is compromised, which can lead to residual stresses building up in the deposited metal. These stresses can cause warping when the gate metal cools down, resulting in desorption from the substrate. After changing the metal pocket between fab runs 2 and 3, however, the fracturing issue was still not fixed.

Between fab runs 2 and 3, we addressed our sonication conditions, which is a technique used to help dislodge resist during lift-off by generating high-frequency sound waves in liquid. The alternating pressure waves create cavitation bubbles on the sample surface that agitate the resist layer. We realized that this process may have been too aggressive for small gate features, like the barrier, so we reduced its power and duration. Since barrier still fractured during fab run 3, we deemed this not to be the problem either.

Upon inspecting Fig. 4.5, we found some directionality to the barrier fracture, with most barrier remnants bent leftward. This brought the idea that barriers fractured during the pipette stage of lift-off. After 2 hours in lift-off etchant, additional AR 600-71 is dispensed from a pipette onto the submerged coupon. This is meant to apply a small pressure to help flake off any residual metal film. Since most barriers fracture in the same direction, it was believed that the pipetting was too strong and had caused the deposited gate metal to rip along the etchant flow. Indeed, when removing pipetting form fab run four, the barrier fracture issue was resolved.

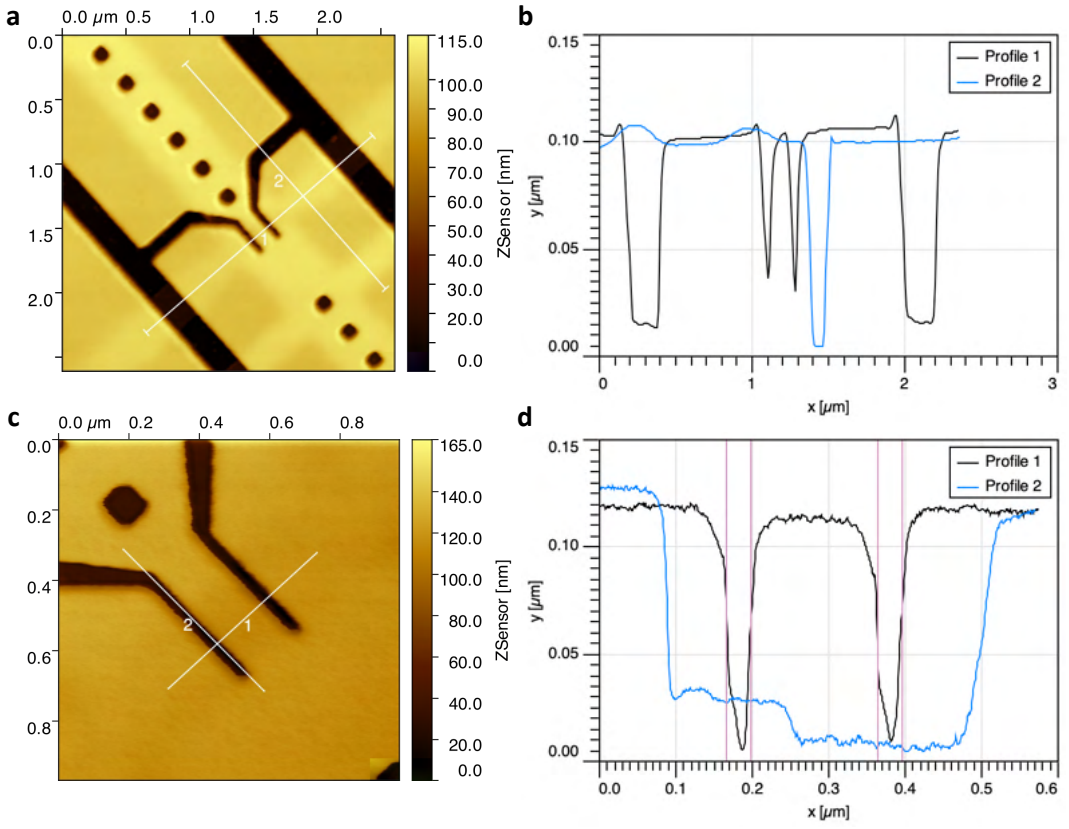


Figure 4.6: AFM images (left) taken after the development of e-beam resist alongside depth profiles of the displayed line-cuts (right). Cap widths on the line-cut endpoints indicate the width over which the depth profile is averaged. Red lines on the depth profiles indicate points between which dimensions were collected. The top row of images are from fab run 2, and the bottom row from fab run 3.

From our experience, we therefore do not recommend pipetting to aid lift-off, as the pressure that is generated is unreliable and will vary between users.

We turn to Fig. 4.6 to better understand where the poor barrier integrity stemmed from. From the first fabrication run (top row) we notice that the depth of the developed barriers is shallower than expected, which would have led to the deposited metal not making proper contact with the substrate. This could be explained, however, by the large step size of the AFM, which made it impossible to resolve the bottom of the developed barrier trench. Turning to the higher resolution scan during the third fab run (bottom row), we indeed see that the developed region meets the substrate, even with a resist thickness slightly larger than usual, at around 110 nm. In agreement with the observation in Fig. 4.5 inset, we can thus exclude the possibility that barriers broke by purely being attached to the resist side walls. The second clue lies in profile 2 of Fig. 4.6d. A step is visible around 0.25 µm, which corresponds to the ledge of the underlying screening gate. The elevation change here is about 20 nm. While gate climb is supported by making gate layers increasingly thick, this is still only 10 nm less than the intended barrier gate thickness of 30 nm. Considering the miscalibration of the tooling factor, this difference is likely even less and introduces a large stress concentration point in the deposited barrier. We deem this to be the origin of the barrier's fracture.

In fab run 2, the likelihood of fracture may have been exacerbated by the decreased overlap between the barrier and screening gate upon increasing the screening gate separation. Any potential bridging in this region will have aided the barrier bending away from the substrate and eventually fracturing further along its length.

4.2.3 Surface Particles

The final fabrication challenge we would like to highlight is contamination on the sample surface. Since the first fabrication run, we observed a large number of pillar-like particles on the sample surface, as shown in the 3D AFM map of the ohmic layer after lift-off in Fig. 4.7c. It is believed that these pillars, are residual Pt particles that transferred from the resist surface to the substrate during lift-off. These particles were also observed on bare coupons after pre-fabrication, even after cleaning with nitric acid, suggesting that a portion of them may also be Ti:Pt from the e-beam marker layer or SiN from the bond pad underlayer. This could be confirmed with energy-dispersive X-ray spectroscopy (EDX). The sharp cliffs formed at the ohmic contact edges, however, are indicative of this being mostly a lift-off issue, as this is where the resist sidewalls will have brushed against. The metal particles could not have directly been deposited underneath the resist during evaporation, because the resist's thickness and high density, as seen in Fig. 4.4, would not permit this. The high aspect ratio of the particles can be explained by the solidification direction during metal deposition. Since the substrate is much cooler than the surface to which metal atoms are added, a temperature gradient is set up out of the sample plane. As solidification proceeds, grain growth aligns with this gradient, causing the grains to take on an elongated shape. The single particles we observe may therefore be single grains of metal.

The grain heights were extracted relative to the lowest point on the sample surface. Therefore, to gauge the heights accurately, grains on the ohmic surfaces were masked out. The remaining grains were identified using a curvature threshold set to 1%. With this rudimentary approach, the vast majority of the found grains, highlighted in green in Fig. 4.7a, fall between 12–17 nm in height. Referring to the depth profile in Fig. 4.7, however, the average substrate elevation is ≈5 nm, hence, the local height of these grains is likely concentrated around 7–12 nm.

The presence of these grains is concerning, because their heights fall above the thickness of deposited Al_2O_3, meaning they could make electrical contact with the overlying gate layers. Furthermore, these grains will introduce a high degree of disorder in the Al_2O_3, likely increasing the trap density. Grains of similar height can also be seen on the sample surface in Fig. 4.4 after

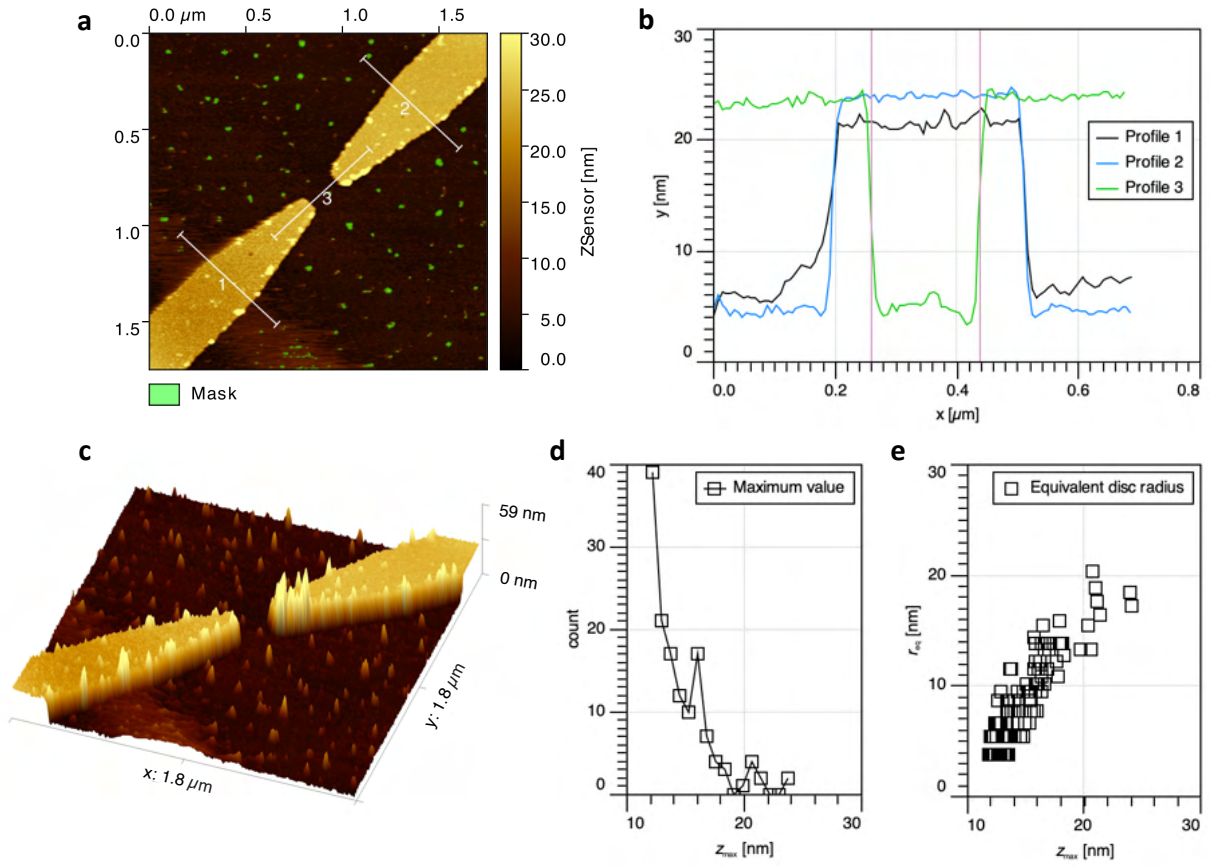


Figure 4.7: AFM-based characterization of ohmic contacts after lift-off during the first fabrication run. **a)** AFM image with particles on the substrate masked via curvature thresholding. **b)** Depth profiles along the three line-cuts shown in a). **c)** 3D mapping of the sample surface. **d/e)** Statistics on the height and correlation between height and equivalent disc radius of the particles masked in a).

the screening gate lift-off.

Fig. 4.7 e) shows the grain heights against their equivalent disc radii, which show a strong dependence as well as an abundance of smaller grains compared to larger ones. This can be explained by the coupon being held at an angle during lift-off, which makes it more difficult for larger grains to adhere to the substrate under their own weight. Moreover, sonication encourages the break up of both resist molecules and metal films, which would thus increase the concentration of small metal particles. Therefore, while sonication may not have caused barrier fracture issues, reducing sonication time and power may be of interest to reduce the number of metal grains on sample surfaces. We also note that since this work, an improved pre-fabrication process has been implemented: a sacrificial AlOx layer is deposited onto the bare wafer that temporarily protects the sample surface from SiN and Ti:Pt during lift-off. The AlOx is subsequently dissolved with a TMAH etch. With this fabrication flow, the surface particles are greatly suppressed, and we advocate for the continued use of sacrificial AlOx.

4.3 Dimensional Characterization

As a way to characterize our fabrication uniformity, we set out to measure SHT dimensions statistically. With this, we hope to learn if fabrication errors could also explain any variability seen in transport measurements. To this end, we take four high-resolution SEM images covering a total of 238 SHT devices. These are shown in Appendix B. We focus on two dimensions in particular, which should have the largest impact on charge transport: barrier width and plunger width. To carry out the characterization, several methods were attempted. The first involved using a Sobel edge detection algorithm with the OpenCV library in Python. Once edges were detected, polygons could be fitted to the edges, for instance, a rectangle for barriers and a circle for the plunger, whose dimensions could then be extracted automatically. The problem with this approach was that polygon fitting requires edge contours to be closed. Due to the limited SEM image contrast, however, the Sobel algorithm often extracted only a portion of the edges, leading to broken contours. Image filtering and morphological operations were used to try connecting the contours together but to no avail. Due to the similar grayscale values across gates in the SEM image, extracting specific gate sizes via thresholding using ImageJ software was also unsuccessful. In the end, we resorted to measuring dimensions manually by drawing lines between the barrier edges and fitting ellipses over the plungers.

We choose an elliptical fitting on the plunger over a direct measurement, as it contains richer geometrical information without much additional effort. An ellipse placed on the SEM image is fully described by the following: a major axis radius a, a minor axis radius b, the angle between the major axis and the image axes θ [1], and a centroid coordinate (x_0, y_0). Using this information, we find it most accurate to define the plunger width as the horizontal distance between the ellipse edges at the height of the centroid, as depicted in Fig. 4.8, as this is where the hole current is most likely to pass under.

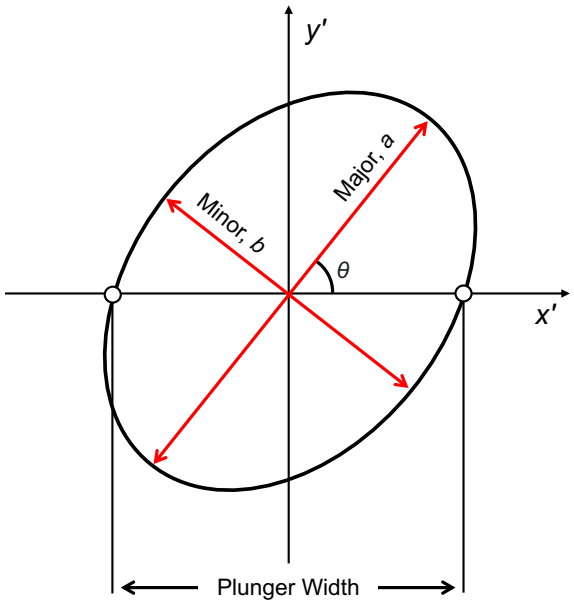


Figure 4.8: Diagram showing the elliptical fitting to plunger gates in the SHT grid, with major and minor axes indicated and corresponding radii a and b.

The general equation of an ellipse in its own frame of reference, with y aligned with the major axis and x with the minor axes, is given by,

[1] all SEM images were rotated to be aligned with the software window

$$\frac{x^2}{b} + \frac{y^2}{a} = 1. \tag{4.1}$$

In the frame of reference of the SEM image, with axes y' and x', the rotated ellipse points may be found using the coordinate transformation,

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{4.2}$$

By setting $y' = 0$ and solving the resulting system of equations, one acquires the coordinate of the x'-intercepts. The plunger width is then the difference between the x'-intercepts.

$$x' = \pm \left(\sqrt{\cos^2\left(\frac{\theta}{b}\right) + \sin^2\left(\frac{\theta}{a}\right)} \right)^{-1} \tag{4.3}$$

The measured widths are presented in Fig. 4.9. The average widths (standard deviations) of the features are – plunger; 148.6 (5.1) nm, left barrier; 52.2 (4.2) nm, right barrier; 51.5 (3.5) nm, and the plunger area is 0.01858 (0.00105) µm^2. We note the especially impressive plunger dimensions, which are on average just 0.6 nm larger than the original width target, with a relative standard deviation is just 3.4%, and 5.7% for the area. We point out the seemingly bi-modal distribution of plunger widths. We believe this is due to the SEM image resolution, which may have led to the characterization of plunger widths on either side of one pixel, which measures 8 × 8 nm. The barriers, despite design adjustments, still turn out larger than expected, though with standard deviations of only 8.0% and 6.7% of the mean for left and right barriers, respectively. We note that these deviations are comparable with the dimension of the e-beam used of 6 nm.

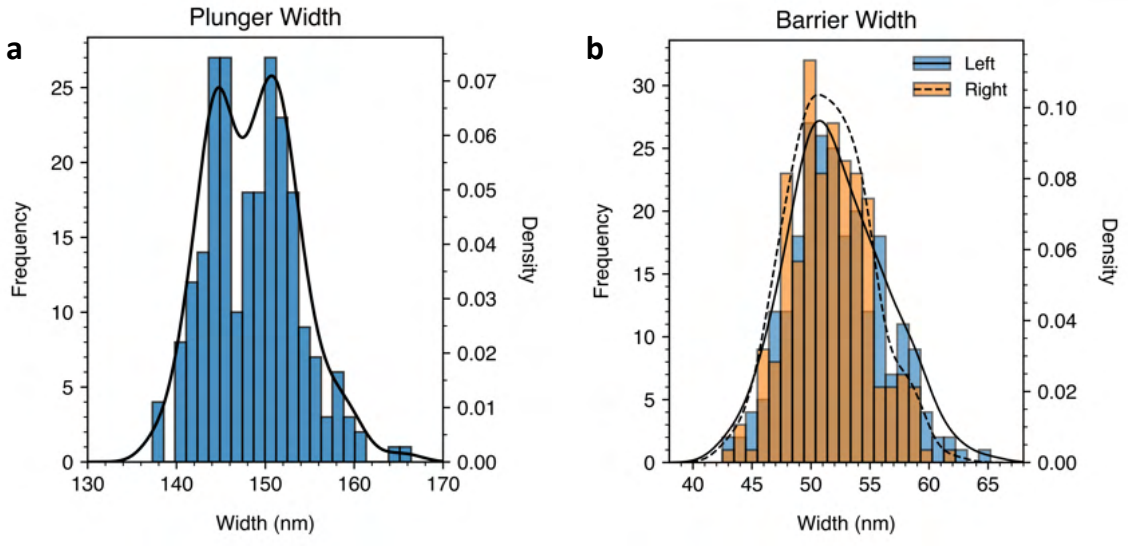


Figure 4.9: Histogram and distribution of widths for the **a)** plunger gate and **b)** barrier gates

In the ImageJ software, the coordinates of all line endpoints and ellipse centroids are saved, which allows us to reconstruct the layout of every plunger and barrier pair. To reconstruct the barrier, we simply draw vertical lines separated by the measured widths, while for the plunger, we plot its transformed coordinates explicitly using Eqs. 4.1 & 4.2. We overlap these reconstructions in Fig. 4.10 by placing all plunger centroids at the origin. The way to interpret this visualization is that those left barrier edge pairs positioned most to the right likely correspond to the right barrier edge pairs also positioned most to the right, and vice versa for the left-most barrier

edges. Since the plungers are always centered, the visualization informs us on the misalignment between the plunger and barrier layers, as well as the spread in barrier positions, relative to their plungers. From this, we deduce that there is a general left-ward misalignment of the barriers, relative to the plungers, as verified by Fig. 4.2a. This may make STH transport less sensitive to the left barrier potentials, as the plunger overlaps less with this barrier. Also, the positions of the left barriers seem to be more broadly distributed, which may be an artefact of the left barriers having a broader width distribution than the right barriers. On the other hand, the similar average widths and standard deviations between left and right barrier means they should have similar influences over the potential landscape.

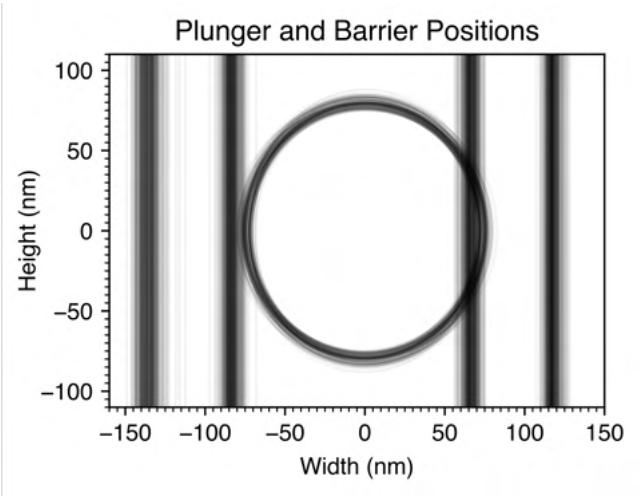


Figure 4.10: Relative positions of plunger and barrier gates as reconstructed from a dimensional characterization

To complete our dimensional characterization of the grid, we also determine the wavelength of our crosshatch pattern. In Fig. 4.11, we present an AFM scan collected over a bare wafer containing our heterostructure, with the Si cap. Using Gwyddion's roughness parameters package, we select a profile of the image that captures many neighbouring height undulations and measure their average wavelength to be 1.655 µm. We use a roughness cut-off set to 0 and a profile width of 3 px. Over the same profile, we also report an average peak to valley height of 3.760 nm, maximum peak-to-valley height of 5.420 nm, and RMS roughness of 1.201 nm. The roughness and wavelength are both lower than values reported in literature for similar SiGe thicknesses [51, 104], indicating that our heterostructure possesses few TDs and a high level of strain relaxation – a precursor to homogeneous quantum device operation. Since strain fluctuations are carried over the entire extent of the heterostructure, we can expect the crosshatch wavelength to be very similar inside and around the QW.

Unfortunately, however, comparing the crosshatch wavelength to the repeat unit size of the grid, which is about 6.25 × 2.94 µm, we learn that strain fluctuations would be undersampled by the SHT devices. Therefore, we cannot directly study the spatial dependence of SHT performance on strain fluctuations. Nevertheless, since the SHT grid was aligned with the crosshatch pattern before fabrication, the likelihood that an entire row or column of devices falls on a peak or valley of the crosshatch is high. This could help reveal the range of influence that the crosshatch could have on the performance of SHTs.

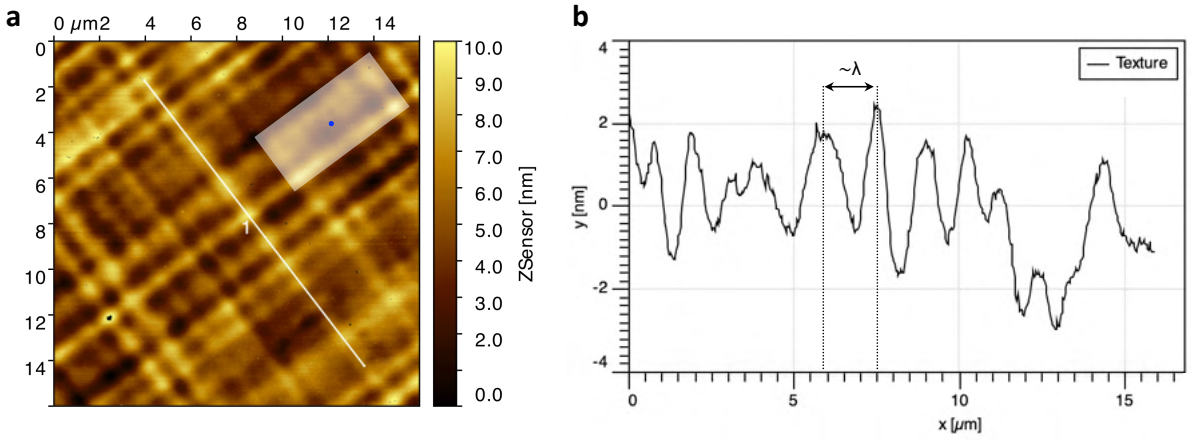


Figure 4.11: a) AFM image of a wafer coupon after pre-fabrication showing a clear crosshatch pattern. The shaded region marks the repeat unit size of an SHT device in the grid, with the blue point defining the plunger. Rows in the raw AFM image were aligned by a 'matching' method, scars were removed, and outliers caused by machine error were masked and zeroed, as seen for example near the bottom left of the image. **b)** Roughness profile along the line cut shown in a) where the spacing between roughness peaks marks the approximate crosshatch wavelength.

4.4 Connectivity and Wire Bonding

The PCB to which one die is wire bonded to has a square arrangement of bond pads, matching the layout of bond pads on the grid. The grid is glued to the PCB using a small dollop of a GE varnish. The bonding scheme for the SHT grid is shown in Fig. 4.12.

Labels on the PCB bond pads are unique and map to specific lines controlled by the cryo-multiplexer (cryo-mux). The 'vertical' pads labeled with a V (orange), meant for barriers, are divided into 9 subgroups and share 4 voltage lines. The 'horizontal' pads labeled with an H (blue), meant for plungers, are divided into 12 subgroups and share 3 voltage lines. In other words, any one vertical voltage lines controls every fourth column of barriers in the grid, while any one horizontal voltage line controls every third row of plungers. We use the convention that the pad label's first number is the subgroup and the second number is the voltage line. To address a particular SHT device, one horizontal voltage line, and two vertical voltage lines are biased, followed by the selection of the correct subgroups to which these voltages should be routed. Since the cryo-multiplexer permits only one vertical and horizontal subgroup to be active at any one time, it was imperative that barriers were bonded to the vertical lines, as these contain an even number of voltage controls within a subgroup.

The vertical subgroup labels are incremented going clockwise around the PCB, while the horizontal subgroups are incremented anti-clockwise. Since consecutive columns of barriers were also bonded clockwise, this makes mapping physical devices to their vertical control lines very straightforward: the first column of barriers is controlled by V1_1, the second column by V1_2, the 10^{th} column by V3_2, and so on. Due to the fan-out of plungers alternating sides of the grid, however, their mapping to the horizontal pads is less trivial. The first row of devices is bonded to H12_3, the second to H1_1, the third to H12_2, and so on. For example, to operate the device with row and column indices [7,7], one would have to bias V2_3/4 and H11_3.

The shared lines are located at the center points of each line of bond pads on the PCB. These are used to bond to the source and drain ohmics as well as the screening gate of the grid. The shared control of these lines is useful for checking that there is continuous connectivity of the ohmics and screening gate across the grid as well as providing an additional control point if the fan-out is anywhere broken.

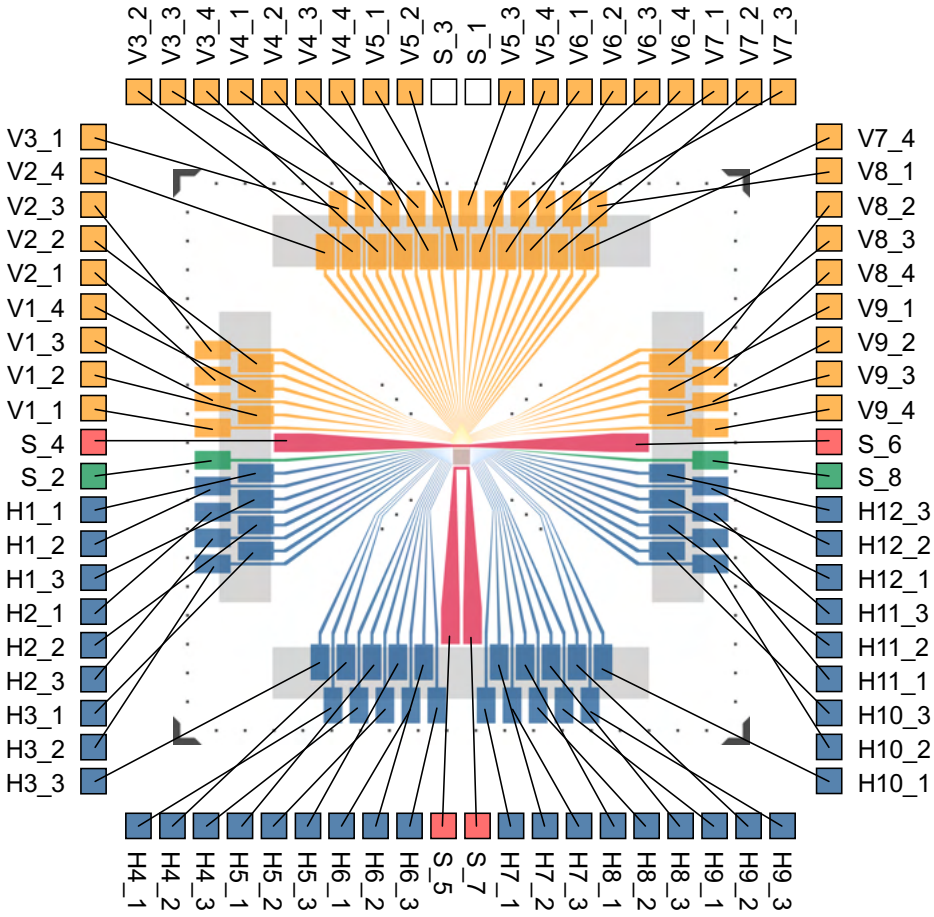


Figure 4.12: SHT grid bonding scheme. Labels on the bond pads state whether they correspond to vertical (V) or horizontal (H) lines, the device sub-group they belong to (first number) and what voltage line should be biased to control that line (second number).

The bonding procedure uses a 25 µm Al bond wire and begins with the connection of all the grid's outer bonds first, followed by the inner bonds. This prevents collisions of the wire head with existing bonds on the grid. To aid this process further, the bond height is increased to 115% of the bond distance for the outer bonds as compared to 105% for the inner bonds. To reduce the chances of failed bonds damaging the sample, bonds are always made from the PCB to the grid. To prevent electrostatic discharge (ESD) from permanently damaging devices, an ESD-safe wrist strap is used at all times. A fully-bonded grid sample is shown in Fig. 4.13.

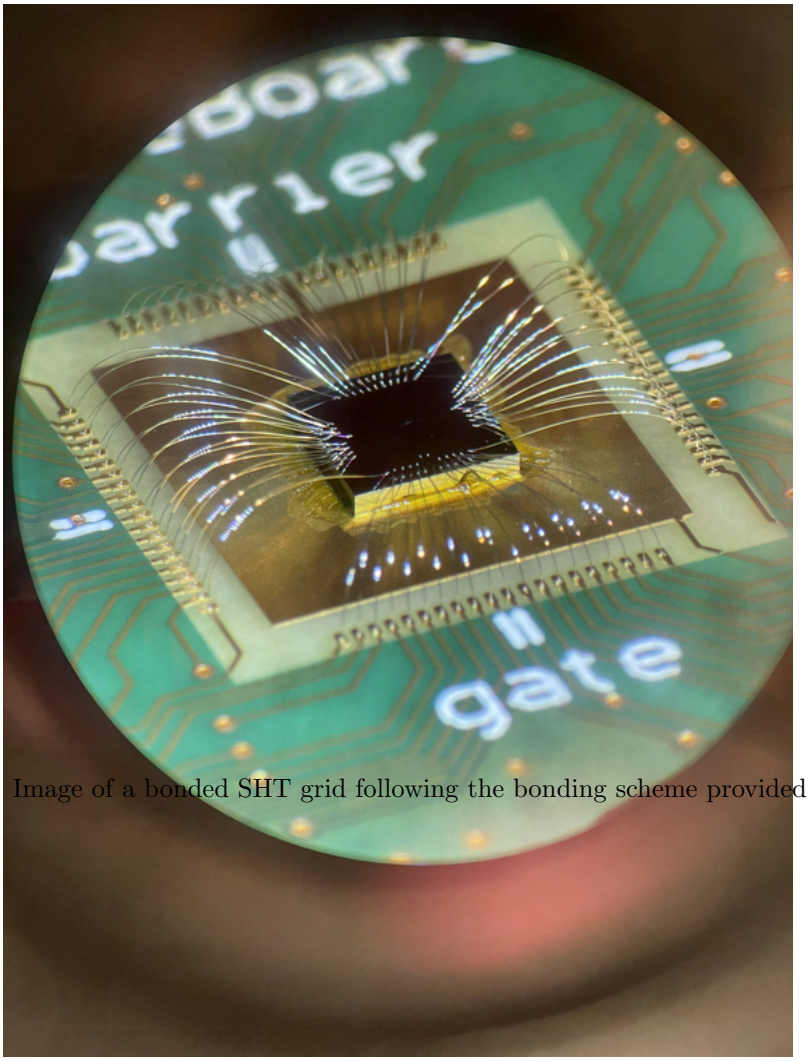


Figure 4.13: Image of a bonded SHT grid following the bonding scheme provided in Fig. 4.13.

5 MEASUREMENTS

5.1 Experimental Set-up

5.1.1 Cryogenic operation

Measurements of the grid were conducted in the attoDRY2100 – a cryogen-free cryostat operating in the temperature range 1.65 – 300 K. Unlike dilution refrigerators, which rely on the mixing of liquid helium-3 and helium-4 isotopes to generate cooling power, the attocube series works by circulating compressed helium gas through a series of expansion chambers. This allows for rapid sample cool-down and precise temperature control within 10 mK near base temperature. In addition, the cryostat has a proprietary ultra-low vibration design as well as a custom top-loading system, enabling fast turnaround times during sample exchange [105]. While qubit experiments typically take place at temperatures below 50 mK, 1.65 K is sufficient for our purposes. By using an electron in a box model and substituting 150 nm for the box length (maximum QD width), and $0.05m_e$ for Ge's effective hole mass, an approximate energy level spacing comes to $\sim$10 meV. The thermal broadening (k_BT) at 1.65 K is approximately 0.14 meV. By comparing these values, the temperature made available by the attoDRY2100 is in theory sufficiently low to enable transport through single QD levels.

The components of the attoDRY2100 include a room-temperature compressor, three heat exchangers, a regenerator, a pulse tube, a reservoir, and an orifice. The complementary pumpking kit is equipped with a turbomolecular pump that allows the sample space to reach pressures down to 10^{-7} mbar. When loading samples into the cryostat, the chamber is evacuated to 10^{-5} mbar levels before introducing helium from a rubber bladder via a needle valve. When filled, the operating pressure of the cryostat is $\sim 7 \times 10^2$ mbar.

The sample is lowered into the cryostat via a measurement insert (Fig. 5.1b). It includes 2×12 low-resistance brass wires, $2\times$ 12-pin patch cables, and break-out panels to coaxial BNC. There also exist an additional 12 manganin wires as twisted pairs. We use the patch cables to connect between the room temperature electronics and the PCB of the cryo-multiplexer and a custom ribbon cable to connect further to the grid PCB. Near the bottom of the insert is the heater stage with a calibrated temperature sensor, a base plate for sample mounting, and a breadboard-based mounting stand for room temperature adjustments [106].

5.1.2 Control Electronics and Software

The measurement set-up for the grids is shown in Fig. 5.1. Coaxial cables leaving the fridge are connected to an isolated matrix module, which is equipped with internal π-filters that suppress

interference above 10 GHz. From the matrix module, micro coaxial connectors (MCX) connect to an IVVI-rack and a Serial Peripheral Interface (SPI)-rack. Both racks are powered by isolated batteries to minimize coupling to 50 Hz signals from the power grid. The racks host a D5a module with 18-bit digital-to-analog converters (DACs) used to bias the plungers, barriers, screening gate and ohmic, a U2 module for controlling the cryo-mux PCB, an M1h module for measuring the drain current, and an S4b module as a voltage de-amplifier for the source ohmic. Information on each of these modules may be found in Refs. [107, 108]. The source voltage is de-amplified by a factor of 10^{-2}. The DC drain current signal is $\times 10^7$ amplified by the M1h and detected using a Keithley6500 Digital Multimeter (DMM), which is equipped with a 16-bit digitizer, has a 10 pA noise floor and maximum sample rate of 1 M samples/s.

All instrumentation is connected to a workstation computer that hosts a software environment built around QCoDeS, – a data acquisition and handling framework developed by the Copenhagen/Delft/Sydney/Microsoft quantum computing consortium [109]. A virtual workstation is initialized that contains the software drivers necessary to operate all modules of IVVI- and SPI-racks. A virtual grid instrument is initialized that maps between a device's physical location in the grid and its corresponding control lines according to Fig. 4.12. This information is used to bias the correct plunger and barrier voltage lines, as well as activate the correct device subgroup. Voltage lines are initialized as virtual parameters and operated through setter and getter functions by the D5a module. Software of the U2 module controls serial-input-parallel-output (SIPO) shifting registers on the cryo-mux PCB to correctly select device subgroups. To do this, a sequence of 'data' bits is passed to define which subgroup will be selected. A 'clock' signal is sent while loading each bit. A 'strobe' signal is supplied to indicate when the shift register is fully loaded and the outputs can be sent to the multiplexers.

Figure 5.1: a) Schematic of the electrical connectivity in the measurement set-up with temperature stages indicated by dashed lines. **b)** Image of the lower portion of the measurement insert. The sample is mounted on the underside of the sample PCB.

5.2 Hall Bar Measurements

Before measuring the SHT grids, Hall bar measurements were performed at 4.2 K in liquid nitrogen using a cryogenic dip-stick. We do this to benchmark the material stack's transport properties, compare them to the literature, and verify the successful growth of the heterostructure and grid fabrication. We use a standard four-wire low-frequency lock-in technique as described in Ref.[29] with a constant source-drain bias of 1 mV, fixed lock-in frequency of 17 Hz, and a negative voltage on the gate electrode to accumulate holes in the buried Ge QW.

The first turn-on curves of the hall bars are shown in Fig. 5.2. Of the three Hall bars fabricated on this coupon (HB_A, HB_B and HB_C), two turned-on while one showed gate to ohmic leakage (HB_A). Leakage is the flow of current anywhere in the device other than along the intended path. Most commonly, leakage currents form between an ohmic lead and the gate electrode, especially when the dielectrics separating them are thin. By biasing the source positively with an ammeter connected to the grounded drain electrode, ohmic conduction yields a positive current, while leakage between drain and gate will registers as a negative current.

Because the voltage difference between the drain and gate is much greater than between the source and drain, a negative leakage current usually dominates over in-plane ohmic transport, which prevents a turn-on from being measured. In Fig. 5.2 b), we see this to be the case in HB_A, while HB_B and HB_C do not show any negative current above the noise floor. The two non-leaky Hall bars show very similar turn-on thresholds in Fig. 5.2 a) around −380 mV.

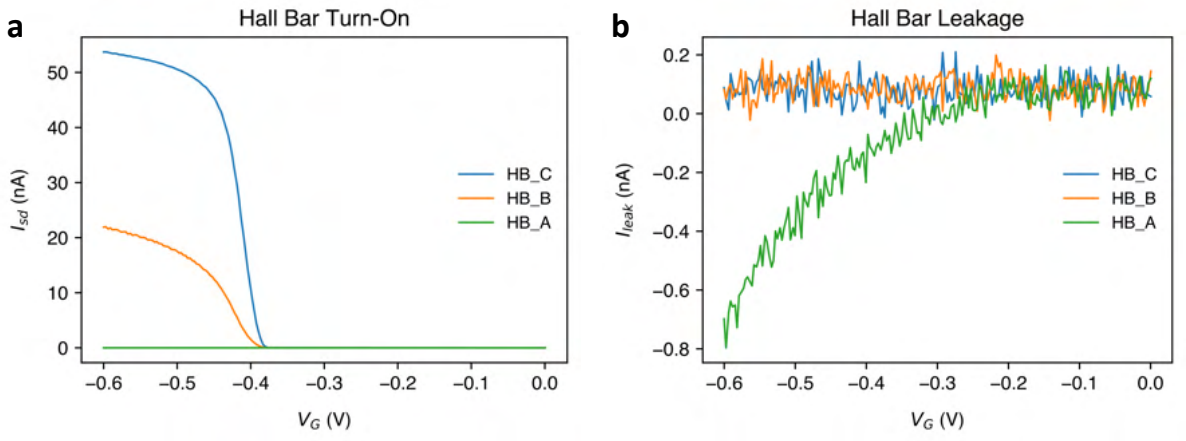


Figure 5.2: a) Source-to-drain current and **b)** leakage current measured over a sweep of the gate voltage for three hall bar devices.

We start with the determination of the hole carrier density n. To do this, we sweep B_z against R_{xy} to acquire the slope B_z/R_{xy}, which is then substituted into Eq. 2.21. This is more accurate than calculating the hole density at a single B_z value due to there often being a small offset in the Hall resistance R_{xy} at zero field ($B_z = 0$). An example fitting of this sweep is given in Fig. C.1. This fitting is repeated over multiple gate voltages V_G, each time yielding a different n. In reality, we sweep V_G while stepping B_z, because, in addition to being faster, device stability is enhanced when single voltages are not held over extended periods. The gate voltage and carrier concentration are linearly proportional and have been plotted in Fig. 5.3 b). The gate voltages applied to HB_B are larger than on HB_C, because the device experienced hysteresis when having its gate pushed to ever-higher negative voltages prior to data collection, which caused the turn-on threshold to shift away from its original value. This behaviour is discussed more extensively in Sec. 5.3.3. By extracting the slope $|n/V_G|$, we fit the gate capacitance C_G, which we find to be 152.6 nF cm^{-2} and 142.2 nF cm^{-2} for HB_B and HB_C, respectively. These values are slightly higher compared to Hall bars measured by group

members on the same heterostructure, which is reasonable given that the dielectric thickness in HB_B and HB_C is 23 nm thinner than our reference, which has $C_G \sim 105\,\mathrm{nF\,cm^{-2}}$.

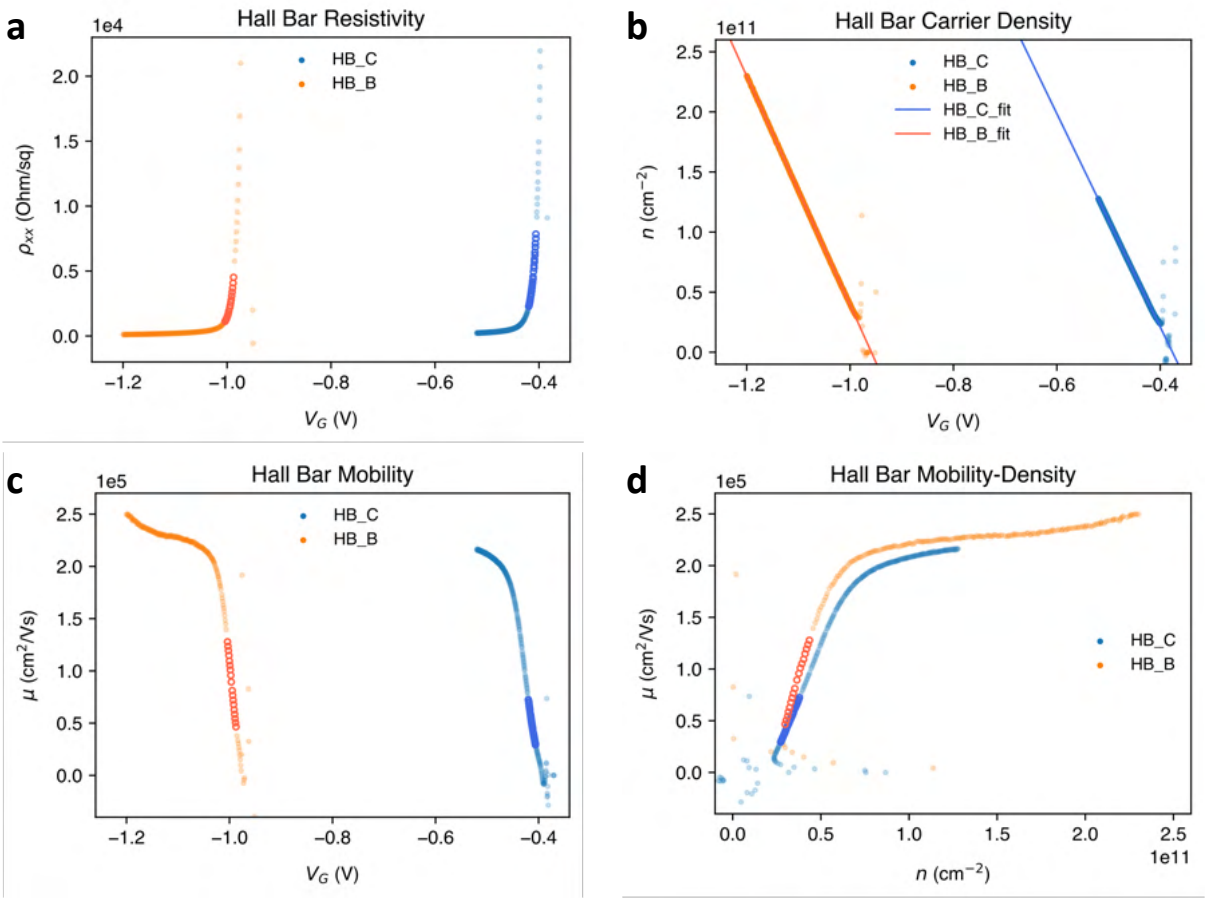


Figure 5.3: Transport data for the two (of three) non-leaky Hall bar devices. **a)** Longitudinal resistivity with gate voltage, **b)** hole carrier density with gate voltage, **c)** hole carrier mobility with gate voltage, and **d)** hole mobility as a function of the hole carrier density. Dark and open data points indicate the data region that is well-behaved in the low-density regime and used for later fitting to percolation theory.

The longitudinal resistivity ρ_{xx} is acquired simultaneously to R_{xy} at zero field, which varies with V_G due to changes in the current through the QW. This data is presented in Fig. 5.3a. Substituting ρ_{xx} into Eq. 2.16, we subsequently extract the mobility, which we plot in Fig. 5.3a as a function of the gate voltage and in Fig/ 5.3d as a function of the hole density. Just after turn-on, the mobility increases sharply due to the increased screening of remote impurities made possible by the growing number of charge carriers in the burried QW. These impurities most likely reside at the semiconductor/dielectric interface, where surface defects are most likely to form. At higher bias, however, the mobility plateaus due to transport becoming limited by the scattering from short-range impurities within or near the QW. The peak mobility we report at $n = 9.2 \times 10^{10}\,\mathrm{cm^{-2}}$ is $2.19\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$ for HB_B and $1.98\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$ for HB_C, which is very similar to previously reported values on the same heterostructure [29] and to state of the art Si/SiGe devices [30]. Qualitatively, this suggests the presence of sharp heterostructure interfaces, a generally low defect density, and QW and SiGe layer thicknesses that are reproducible by our growth process.

High mobility values are corroborated by a low percolation threshold, which we find to be $n_p = 2.27 \times 10^{10}\,\mathrm{cm^{-2}}$ for HB_B and $n_p = 2.01 \times 10^{10}\,\mathrm{cm^{-2}}$ for HB_C. The average of these

values exactly matches the threshold stated in Ref. [29] at $n_p = 2.14 \times 10^{10}\,\mathrm{cm}^{-2}$, and more than a factor of 4 lower than optimized Si/SiGe devices [88].

We extract n_p by fitting the longitudinal conductivity σ_{xx} in the low carrier density region to Eq. 2.24 from percolation theory. The percolation exponent is a critical exponent that depends on dimensionality, lattice structure, and disorder type, and has been theoretically calculated using various methods yielding values ranging anywhere from 1.1 to 1.75 [110]. For Hall bars with low peak mobilities, long-range impurity scattering plays a lesser role, and so the scaling of mobility with carrier density has even been shown to fit well using $p = 0.7$ [111]. In our case, since we see a broad mobility range dominated by remote impurities, we fix $p = 1.31$. This value is based on a numerical model for a large resistor network spanning a 2D square lattice [112] and has accurately described insulator to metal transitions in Si MOSFETs [46] as well as mobility–density data collected previously by fellow group members.

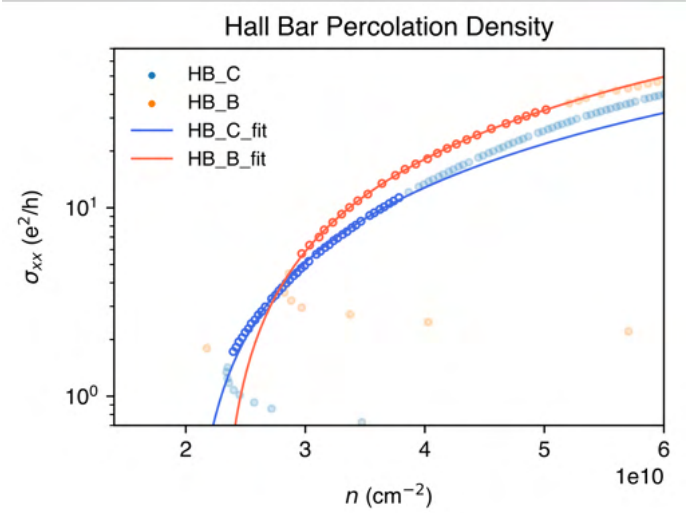


Figure 5.4: Longitudinal conductance as a function of the hole carrier density. Dark and open data markers indicate the data region that is well-behaved in the low carrier density regime and used for fitting to percolation theory (solid lines).

5.3 Grid Measurements

5.3.1 First Characterizations and Leakage

Having characterized Hall bars from the same fabrication run as the grid, we gain confidence in the electrical performance of SHTs grid. Before cooling down the sample, we carry out room temperature checks to ensure that there are no broken or shorted control lines. To do this, we sequentially apply 10 mV to all vertical and horizontal gates while measuring the drain current. In the absence of drain-to-gate leakage, the measured current should not deviate above the noise floor (in the negative direction). We cross-check this with multimeter measurements which should yield an open circuit resistance between voltage lines. For the shared lines that have doubled control over the screening gate and ohmics, we check that these lines are continuous over the grid by measuring a closed circuit resistance of 1 kΩ. Upon passing these room temperature checks, the grid is cooled down to the fridge base temperature of 1.7 K.

The protocol used for tuning the SHT grid closely follows the one found in [89], and is outlined as follows. Firstly, all gates – left barrier, right barrier, and plunger – are swept together at a negative bias to accumulate holes between the ohmics. Initially, the screening gate is grounded. By continuously measuring the drain current I_{sd}, an approximate, global voltage threshold can be found above which SHTs are in conduction. Next, while keeping the plunger voltage fixed, barrier pairs are swept back until the current is pinched off. This marks a preliminary pinch-off voltage V_{po} for the barriers. For the next part, all barriers are gated above the highest measured value of V_{po}, and turn-on threshold voltages V_{th} are determined on the plungers by sweeping V_G. At this stage, only a handful of devices located on the grid diagonal are used for tuning. After V_{th} and V_{po} are found for these few devices, appropriate voltage ranges can be estimated for sweeping the plungers and barriers over the whole grid. In Fig. 5.5, we present the first round of these full measurements.

Each square shown in subplots of Fig. 5.5 corresponds to an individual SHT device. The layout of devices in these plots, indexed by row and column, is consistent with the physical layout of devices in the grid. At each fixed barrier voltage, the plunger voltage V_G is swept through 50 points spaced equally between -1.0 V and -1.5 V. It takes approximately 1 h to carry out this sweep on all 648 devices. The full set of grid sweeps is repeated over four different barrier voltages, spaced equally between -500 mV and -750 mV. Before switching between devices, we ramp down all the gates to ground. The order of device sweeps was carried out row-wise, going from left to right, starting with device (1,1) and ending with device (36,18). The source ohmic is positively biased by 0.5 mV while the drain is grounded. The turn-on threshold of SHTs is reached once the drain current exceeds $I_{sd} \geq 100$ pA. Devices coloured in red exhibit strongly leaky behaviour with $I_{sd} < -10$ nA. Devices that neither turn-on nor leak $-10\,\text{nA} \leq I_{sd} < 100\,\text{pA}$ are shown in white.

The first thing to notice about the grid is its high degree of column-wise leakage. Since barriers are arranged column-wise, this shows a potential shortcoming of the crossbar grid architecture, which is that if any one gate is prone to leakage, then it compromises the entire row or column of devices that share that gate. We note how the measured leakage current strongly depends on the measurement conditions. By increasing the barrier voltage from -500 mV to -750 mV, the grid went from a regime of showing essentially no leakage to one in which leakage affects the majority of devices.

Initially, it was believed that the leakage could stem from a single defect point in the dielectric between drain and barrier metals, which would then dominate the measured current regardless of the device's position in the column. This proposal, however, contradicts data in Fig. 5.6, where we show the I-V curves of all devices in the grid at the four barrier voltages measured in Fig. 5.5. The large negative currents correspond to leakage.

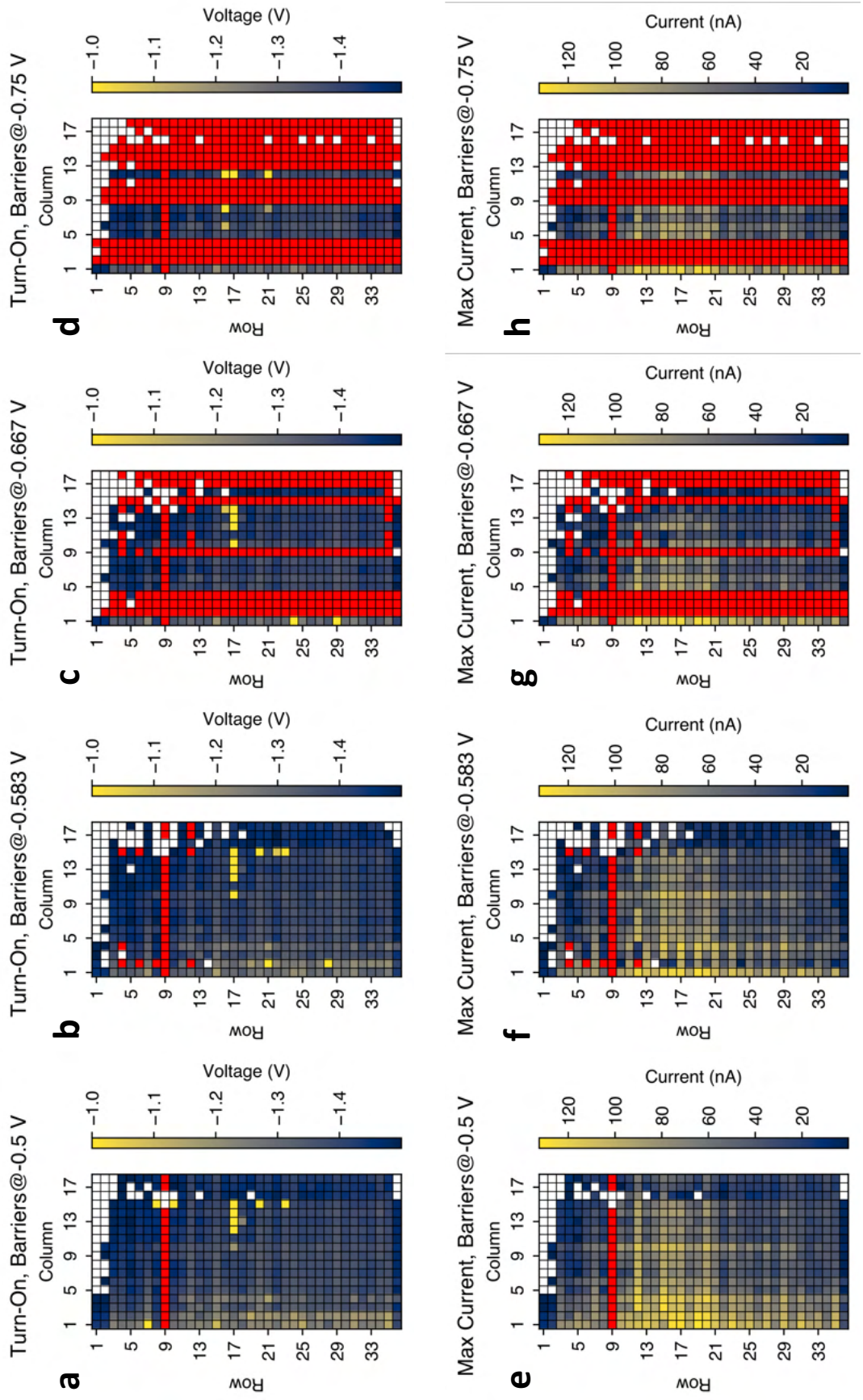


Figure 5.5: a-d) Turn-on threshold voltages and **e-h)** maximum currents of each device in the grid from a plunger sweep between -1.0 and −1.5 V at four fixed barrier voltages, equally spaced between −500 mV and −750 mV. Red squares represent leaky devices, and white squares are devices that neither turn-on nor leak. The screening gate is set to 0 V

Crucially, the onset of leakage in an overwhelming number of devices requires the plunger to also be strongly biased ($V_G < -1\,\mathrm{V}$). Therefore, leakage must happen on a device-by-device basis, and cannot come from a single point on the barrier gate, because this should yield leakage also at zero plunger bias. Also, from the fact that a plunger bias is required, we can infer that the site of leakage is likely near the QD location, where the ohmics, barriers, and plungers are all near one another. Potentially, the measured leakage stems from a tunneling current between the QW and the barriers. This mechanism requires hole accumulation and would also explain why no leakage is measured at zero plunger bias.

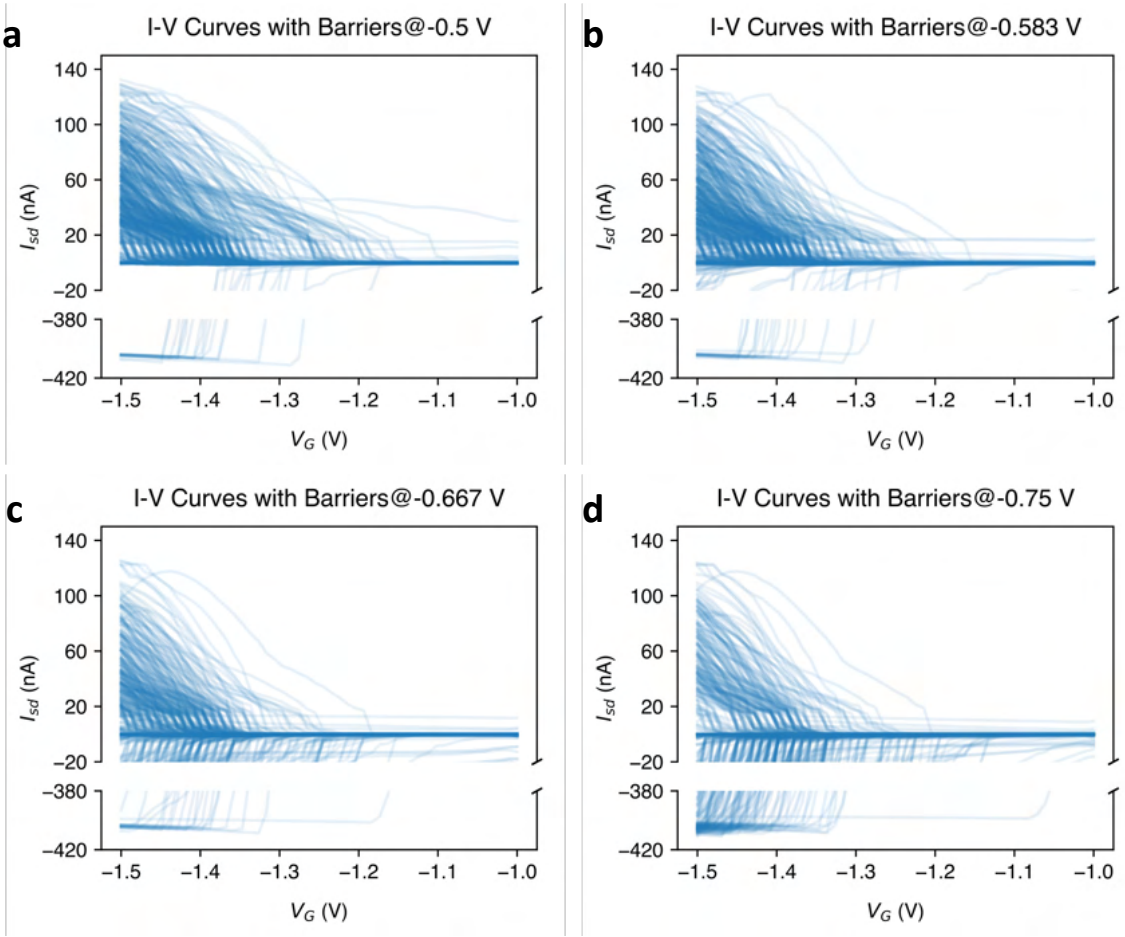


Figure 5.6: a-d I-V curves of all devices in the grid at four fixed barrier voltages, equally spaced between $-500\,\mathrm{mV}$ and $-750\,\mathrm{mV}$. The screening gate is set to $0\,\mathrm{V}$. Negative currents correspond to leakage. Saturation of negative currents to $400\,\mathrm{nA}$ is due to clipping by the amplifier of our measurement set-up.

Device-by-device leakage that is still consistent with a column-wise pattern means that the defect causing this must be distributed over the full extent of the barrier. One option is a constant barrier misalignment. Where the barrier meets ohmics in the SHT fine structure is also near where the barrier must climb over the screening gate, which could contribute to a consistently poorer dielectric quality over the column. However, due to barrier alignment being similar across all devices, as characterized in Sec. 4.3, we deem this an unlikely cause for the leakage. The other option is that barrier columns are located on the flank of the crosshatch pattern, which could lead to a less uniform dielectric than at a peak or valley.

The reason that a similar leakage pattern is not observed across grid rows, which contain plungers, may be due to the additional $5\,\mathrm{nm}$ of dielectric that separated them from the ohmics,

as compared to the barriers. For the row-wise leakage caused singularly in row 9, this may be caused by an individual defect, or potentially also from roughness introduced by the crosshatch. We believe the crosshatch to be the more likely cause, due to devices in columns 15 and 16 not showing leakage, even at $V_{barriers} = -667\,\mathrm{mV}$, which they should if the dielectric were broken directly at any point. The same argument applies to the few non-leaky devices in otherwise leaky columns. Unfortunately, however, since the crosshatch pattern was aligned with the grid, we cannot know for certain if it is the underlying cause. By extension, it is impossible to discern whether variations between rows and columns in non-leaky transport originate from individual defects or from crosshatch features.

In Sec. 4.2.3, we expressed concerns over metallic grains on the ohmic surfaces. These were especially pronounced at the edges of the ohmic contacts, which could have provoked the high leakage susceptibility of the barriers. Thankfully, however, since we require appreciable voltages to be applied to our gates to observe any leakage current, we can rule out these grains as having pierced through the entire dielectric. This on its own, however, does not preclude there being grains connecting any of the control gates themselves. We checked this by varying potentials on each of the screening, barrier, and plunger gates individually. The bias of two metals connected together is always equal to the bias of the metal with the highest potential. Since the screening gate and barriers have lower biases than the plunger, and both had a measurable effect on the current, we conclude that they must be individually changing the confining potential of the conduction channel, and therefore we roll out any gate-to-gate leakage.

Based on this first full characterization of the grid, we conclude that: barrier voltages should not exceed 500 mV to avoid leakage, and that the crosshatch pattern and individual dielectric weak spots cause leakage to occur over entire rows or columns of devices. We next investigated how the grid performance could be improved by manipulating the screening gate voltage. We first tried biasing the gate positively in the hopes that it would reduce leakage by reliving the voltage difference across parts of the dielectric near the conducting channel. As it turned out, however, a small voltage of 100 mV was sufficient to cause large shifts in the threshold voltage and the majority of devices to no longer turn on within the plunger sweep range of $-1\,\mathrm{V}$ to $-1.5\,\mathrm{V}$. This suggests a higher plunger to screening gate capacitance. We did not push devices to higher voltages to avoid dielectric breakdown, and also, as we learned from Hall bar measurements, would likely cause irreversible hysteretic behaviour.

We next tried applying a negative voltage to the screening gate. At $-500\,\mathrm{mV}$, we found notable reductions in the turn-on threshold across all devices, as well as qualitatively higher transconductances dI_{sd}/V_G, which helped in achieving higher maximum currents within the same plunger sweep range. Moreover, we carried out an additional grid sweep with barriers set to $-650\,\mathrm{mV}$ and found marked reductions in the number of leaky devices (Fig. C.3). Being the first gate layer in the stack, the screening gate's lever arm is the largest, and thus the enhanced transport can be explained by a broadening of the conduction channel. On the other hand, the reduced leakage can be explained by a reduced voltage stress on the dielectric due to the voltage difference between screening and barrier gates being reduced. These gating potentials: $-500\,\mathrm{mV}$ on the barriers, $-500\,\mathrm{mV}$ on the screen, and up to $-1.5\,\mathrm{V}$ on the plunger, we keep as a reference for future grid measurements.

The measured turn-on thresholds and maximum currents with the screening gate set to $-500\,\mathrm{mV}$ are summarized in histograms in Fig 5.7c/d. The average (standard-deviation) turn-on threshold V_{th} and maximum current I_{sd}^{max} across the grid was -1.272 (0.072) V and 137.5 (20.3) nA, respectively. We remark that this is a rather large current as compared to SHT devices measured previously in our group. However, we are not concerned about this being due to multiple devices contributing current, because then we would expect to see kinks in the I-V curve when different devices turn on, which is not observed. The high current is, therefore, a simple case of having a wide conduction channel and, thus, a low resistance for the 2DHG in the

QW. The relative deviation from the mean on the turn-on threshold is 5.7%, which remarkably matches the 5.7% deviation on the gate area found in 4.3. We find this value to be robust against the voltage range we sweep over in later, optimized, grid characterizations. Even with an average turn-on threshold at $-1.489\,\mathrm{V}$, the relative deviation is still 5.6%, and never goes outside 5.6–6.1%. We were unable to reconcile the similarity between deviations in the plunger area and V_{th} using a simple Coulomb model, and, for now, we deem it a coincidence. A proposal on how this could be further investigated is given in Chapter 6.

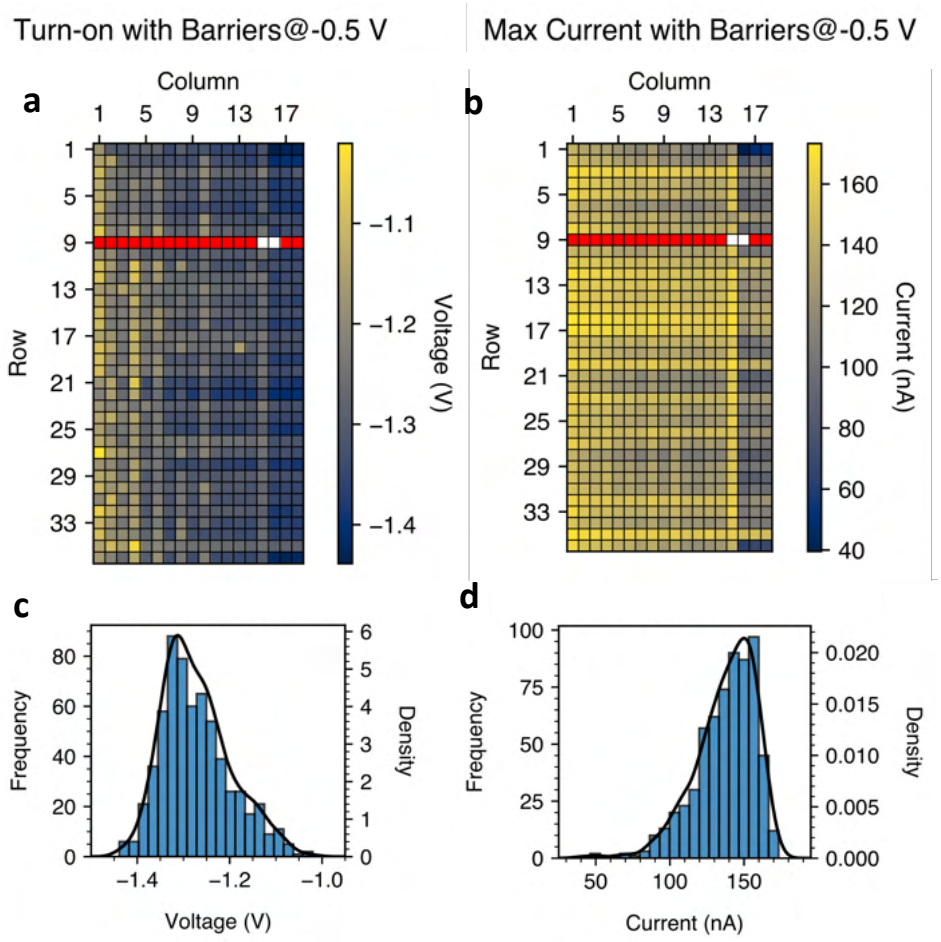


Figure 5.7: a) Turn-on threshold voltages and **b)** maximum currents of each device in the grid from a plunger sweep between -1.0 and $-1.5\,\mathrm{V}$ with barrier and screening gates set to $-500\,\mathrm{mV}$. Red squares represent leaky devices, and white squares are devices that neither turn-on nor leak. Histograms summarizing the distribution of **c)** turn-on thresholds and **d)** maximum currents.

5.3.2 Negative current before turn-on

Upon inspection of I-V curves from the first grid sweep, an interesting feature was discovered before to turn-on. Typically, a couple of hundred millivolts prior, there would be a small onset of negative current. Examples of this can be seen in select devices, shown in Fig. 5.8. In this section, we set out to investigate this behavior and find that it may be related to the extent of disorder in the grid. This makes it a useful addition to the threshold voltage as a way to statistically probe device performance.

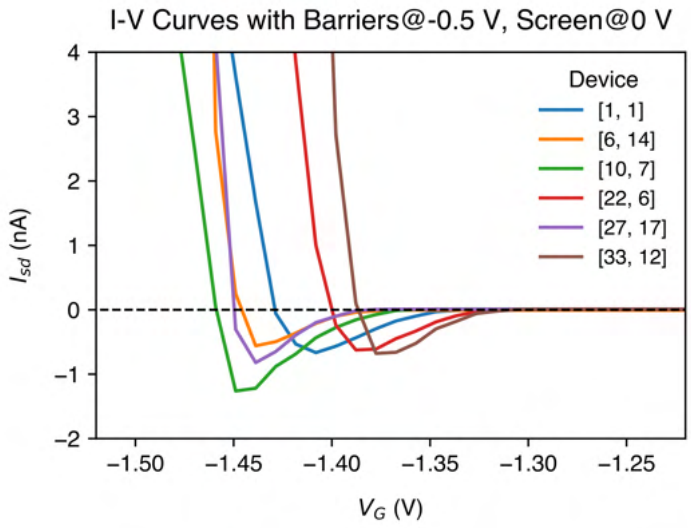


Figure 5.8: Examples of negative currents observed before SHT turn-ons in 6 devices sampled randomly across the grid.

Due to how small the negative current is, typically measuring no more than a few nanoamps, we exclude the possibility that it comes from the direct conduction of holes between drain and gate[1]. In the traditional sense, we therefore do not classify this negative current as leakage, especially as it eventually disappears with further biasing of the plunger. Moreover, it occurs exclusively before turn-on, and those devices that turn-on at higher voltages also show later negative current onsets. This points to the negative current being linked to the voltage at which holes begin accumulating inside the QW. Furthermore, since the negative current is present in all devices, it must be intrinsic to our material stack and mode of operation. Finally, the negative polarity of the current tells us that it must correspond to holes leaving the drain electrode.

In the literature on MOSFETs, a few mechanisms are recognized as contributing negative currents [113–116], though the operating conditions of SHTs dispel most of them.

- **Impact ionization:** an electron or hole with enough kinetic energy knocks an electron from a bound state in the valence band to a mobile state in the conduction band. This is also known as electron-hole pair generation.
- **Auger recombination:** electron-hole pairs recombine at an inner-shell vacancy of an atom, followed by the release of energy in the form of a second electron that is ejected from the atom.
- **Trap-assisted recombination:** electrons or holes transition between energy bands via a trap state in the bandgap formed by a defect in the crystal lattice.
- **Band-to-band tunneling:** electrons or holes tunnel between valence and conduction bands in materials separated by a narrow tunneling barrier.

For impact ionization to occur, the kinetic energy transferred to an electron would have to, assuming a perfectly elastic collision, at least overcome either silicon's or germanium's first ionization energies. These are 1.31×10^{-18} J and 1.27×10^{-18} J [117], respectively, which are far less than the thermal energy $k_B T$ supplied at 1.7 K, which is on the order of 10×10^{-23} J. With V_{sd} being in the milli-volt range, the additional kinetic energy gained by accelerating electrons in this electric field is still only on the order of 10×10^{-22} J, thereby rendering impact ionization

[1] with the general use of the word 'gate', we are referring to any of the screening, barrier or plunger gates

energetically infeasible. For similar reasons, Auger recombination is highly unlikely, as it also requires a large release of energy. Moreover, it is a three-particle process, and, therefore, requires very high carrier densities. Since we observe negative currents before turn-on, we cannot be in a regime of high carrier concentration. Band-to-band tunneling is most commonly observed in transistors with p-n junctions. Under some applied top-gate bias, a strong band bending brings valence and conduction bands on either side of the depletion region into close alignment, allowing electrons to tunnel between source and drain. Since our SHTs are chemically undoped, and the band bending at the Ge QW is mild, this leakage mechanism is also highly unlikely.

This leaves trap-assisted recombination as the last possibility. To this list, we also add capacitor charging as a fifth possible mechanism. Here, the negative current is explained by sitting at a voltage where accumulation begins in the QW, while still being below the percolation threshold. Without conduction between the ohmics, the current device detects only the movement of holes out of the drain reservoir. Using evidence collected through various measurements, we now show that trap-assisted recombination is the more likely cause for negative current over capacitor charging.

Capacitor model

In the framework of capacitor charging, the buried QW and the plunger top gate may be treated as two plates of a capacitor, with the dielectric being the $Si_{0.2}Ge_{0.8}$ spacer layer and the Al_2O_3 in between. Together with some external circuit resistance between the gate and drain, one should therefore be able to treat accumulation in the QW as a simple RC circuit, which has a characteristic charging time constant $\tau = RC$. Assuming neither the buffer layer nor oxide undergo any chemical changes that would alter their dielectric constants, the series capacitance should remain fixed, and is given by,

$$C = \left(\frac{1}{C_{spacer}} + \frac{1}{C_{oxide}} \right)^{-1}, \tag{5.1}$$

where the capacitance of each layer is given by,

$$C = \frac{\varepsilon_0 \varepsilon_r A}{d}, \tag{5.2}$$

A is the gate area, d the layer thickness, ε_0 the permittivity of free space and ε_r the relative permittivity of the medium. By Kirchoff's voltage law, the voltage drop over the external circuit V_R must be equal in magnitude, but opposite in sign, to the voltage drop over the capacitor V_C. Using the relations $V_C = Q/C$, Ohm's law $V_R = IR$, and the definition of current as $I = -dQ/dt$, we find that:

$$V_C - V_R = 0 \tag{5.3}$$

$$\Rightarrow \frac{dQ}{dt} + \frac{Q}{RC} = 0 \tag{5.4}$$

This is a first-order ordinary differential equation, which has an exponentially decaying solution

$$Q(t) = Q_0 e^{-t/RC}. \tag{5.5}$$

By differentiating both sides and solving this expression at $t = 0$, we find that

$$I(t) = I_0 e^{-t/RC} \tag{5.6}$$

where I_0 is the current through the RC circuit at $t = 0$. In our case, we begin with no free charges in the QW, so the initial current flowing into the capacitor comes entirely from the voltage drop across the external circuit V_R. We can thus re-write Eq. 5.6 as,

$$I_C = \frac{V_R}{R} e^{-t/RC}, \tag{5.7}$$

where I_C is the hole current flowing into the QW. The time taken to charge a capacitor to ≈99.3% is 5τ.

$$1 - \frac{I(t = 5RC)}{I_0} = 1 - e^{-5} = 0.99326 \tag{5.8}$$

This is true regardless of the voltage across the external circuit. With greater V_R, the greater is the number of charges to be put on the capacitor and is balanced by a higher charging current, and the charging current decays exponentially with time. By sitting at a plunger voltage below the turn-on voltage, one would therefore expect that the negative current, corresponding to the accumulation of holes and charging of this capacitor, will exponentially tend toward zero. In this way, the characteristic charging time τ can be extracted, and subsequently, the capacitance C. We set up an experiment to do exactly this and compare it to a theoretical prediction for the capacitance, which we now calculate.

Since the plunger is shared by a row of devices, we must consider a capacitor plate area equal to *all* parts of the plunger gate that neighbor a drain electrode, not only the plunger head of a single device. We do not include parts of the plunger that neighbour source electrodes, however, because their carrier injection will not contribute to the current we measure on the drain. We also do not include regions that overlap with the barrier and screening gates, nor those isolated from the ohmics by these overlapping regions, because carrier injection here would have to come from the surrounding heterostructure, which has a much lower intrinsic charge carrier concentration. Using the SHT grid design presented in 3.1, the area we calculate for the capacitor area is $A = 19.89\,\mu\text{m}^2$. The thickness of the $Si_{0.2}Ge_{0.8}$ spacer layer is 55 nm, and the combined thickness of the three Al_2O_3 layers is 17 nm. The relative electric permittivities of these materials are 15.2 [118] and 7.0 [119]. By using Eqs. 5.1 & 5.2, with $\varepsilon_0 = 8.85 \times 10^{-12}\,\text{m}^{-3}\,\text{kg}^{-1}\,\text{s}^4\,\text{A}^2$, we find $C_{spacer} = 157.4\,\text{fF}$ and $C_{oxide} = 72.5\,\text{fF}$, yielding an overall series capacitance of $C \approx 50\,\text{fF}$.

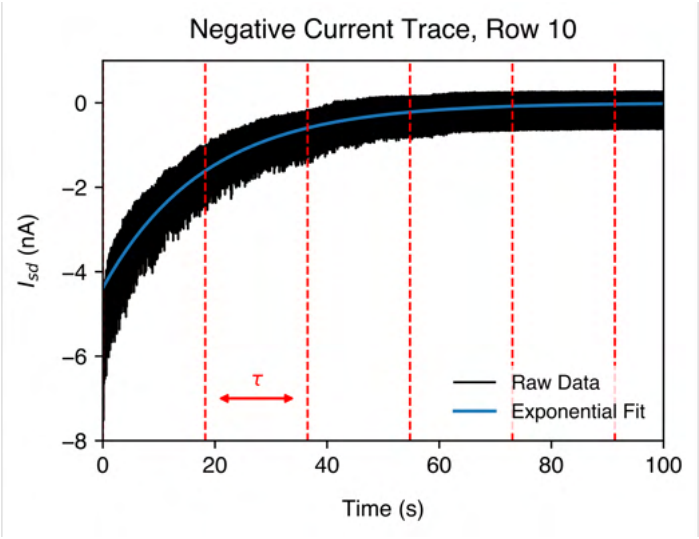


Figure 5.9: Time trace of the negative current measured before turn-on in grid row 10. The plunger bias is fixed at $V_G = -1.495\,\text{V}$, screening gate to $-0.5\,\text{V}$ and barriers to 0 V. The spacing of vertical lines (red) corresponds to the characteristic charging times $\tau = RC$, as determined by the exponential fitting function (blue). The last vertical line equals 5τ.

For the experiment portion, we wish to homogenize the accumulation of holes in all regions of the plunger line, and we do this by setting all barriers to 0 V. The device column we choose then becomes unimportant. We choose to measure row 10, because the current peak on device (10,7) was well pronounced from earlier measurements (see Fig. 5.8). We sweep the plunger of row 10 to acquire an I-V curve, from which we choose a V_G point on which to monitor the evolution of the negative current.

In Fig. 5.9, we show a time trace of the negative current in row 10, held at $V_G = -1.495\,\text{V}$. The data was acquired using the Keithley6500 digitizer at sampling rate of $2 \times 10^3\,\text{Hz}$, hence we capture some 50 Hz noise from the power grid. These are the fast oscillations seen in the current trace. We fit a function in the form of Eq. 5.6 to find $I_0 = -4.381\,\text{nA}$ and $5\tau_{exp} = 91.315\,\text{s}$. The circuit voltage across the capacitor plates we take to be the difference between the drain and plunger voltages $V_C = V_R = V_G$. The resistance through the external circuit is, therefore, $R = V_G/I_0 = 0.341\,\text{G}\Omega$. Using this value, we determine the capacitance to be $C_{exp} = 5\tau_{exp}/R = 53.518\,\text{nF}$.

The experimentally determined capacitance is considerably larger than what is predicted theoretically. Using the experimental R, we can work backward to find the theoretical charging time as $5\tau = 5RC = 0.847\,\mu\text{s}$, which is therefore much shorter. This discrepancy could arise from R being too small, which would mean that V_R is being understated. The amount V_R would have to increase by to close the gap between C_{exp} and C, however, is over 6 orders of magnitude, which is nonsensical. The same line of argument holds for the area A, which would also have to be 6 orders of magnitude greater, amounting to more than the area of the entire coupon. Even within a generous margin of error on the capacitor model, this evidence alone refutes hole injection into the QW from the ohmics as being responsible for the observed negative currents before turn-on. Therefore, we have left to explain what is causing the negative current in the first place, and why it vanishes over long timescales.

We confirm the invalidity of the capacitor charging mechanism by conducting a supplementary experiment. In Fig. 5.10, we plot the I-V curves of device (10,7) by sweeping V_G in both positive and negative directions, over various S/D biases. This also allows us to also characterize plunger pinch-offs on the device.

For $V_{sd} > 0$, we measure the usual positive conduction current due to holes traveling into the drain electrode. When flipping polarity to $V_{sd} < 0$, we measure negative conduction current, because holes must be traveling in the opposite direction, away from the drain. Regardless of polarity, with greater V_{sd}, we also see enhancements in the conduction current. Since the drain electrode is always grounded, the change in polarity as well as current enhancements are only possible if indeed the hole current passes between source and drain, and not elsewhere in the heterostructure. In this way, we confirm that any conduction current measured in this thesis is always between the ohmic contacts.

Turning to the capacitor model, pinch-off would correspond to the capacitor discharging. Here, holes return back onto the drain electrode from the QW. Since this constitutes a positive current, then we would expect there to be no negative current when sweeping V_G backward[2] and with $V_{sd} > 0$. Conversely, as seen in Fig. 5.10 d), that the negative current persists. This means that the origin of the negative current is independent of the direction of hole transport, and cares only about the stationary value of V_G. This makes the negative current a steady-state process, rather than a dynamic one, further invalidating the capacitor charging model.

[2] in the positive direction, from a point in conduction toward pinch-off

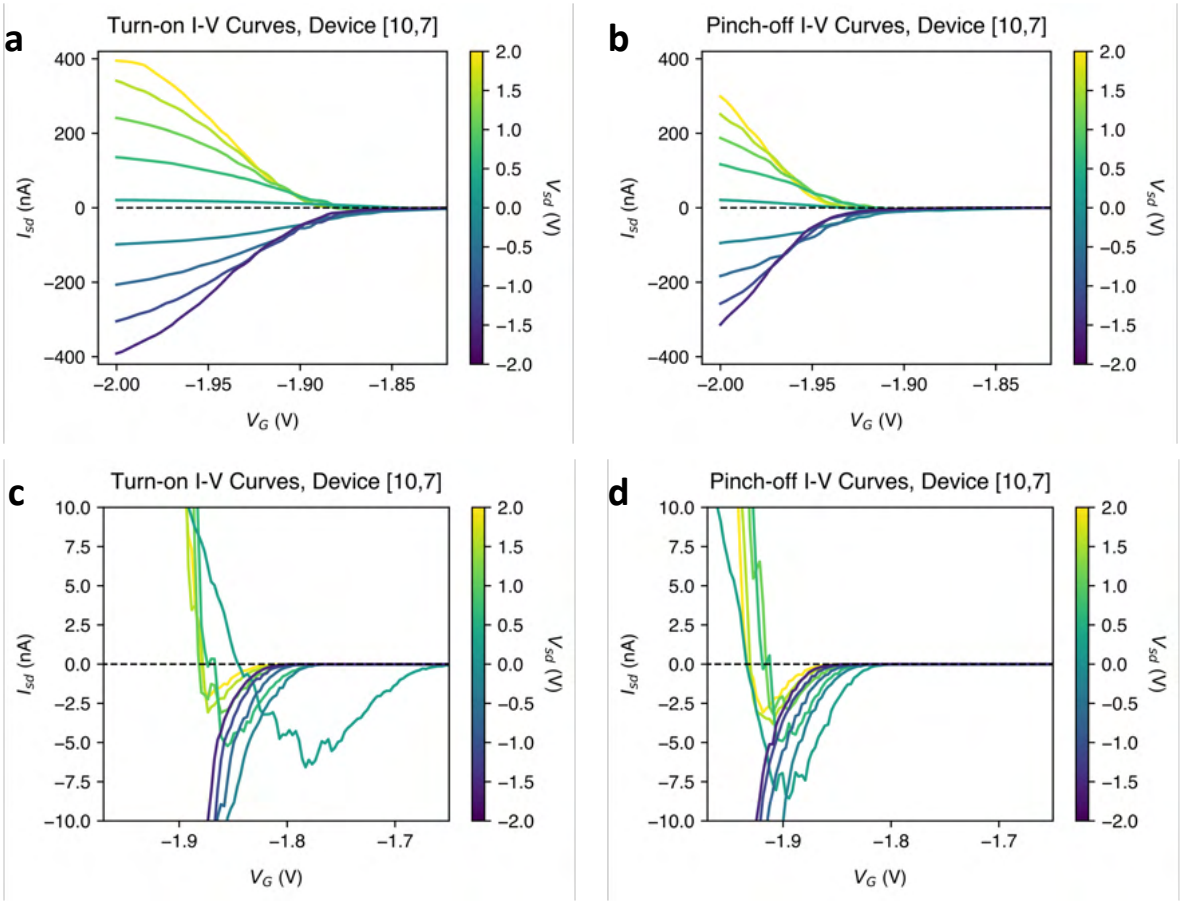


Figure 5.10: I-V curves at two different levels of magnification for **a/c)** turn-on (sweeping V_G in the negative direction), and **b/d)** pinch-off (sweeping V_G in the positive direction.

Trap-assisted recombination

Consistent with observations so far, we propose trap-assisted recombination as the cause for the negative current before turn-on. As discussed in Sec. 2.5, the oxide interface is a notoriously disordered region of thin film devices. In the presence of the surface particles imaged in Sec. 4.2.3, this disorder may be substantial. Therefore, we believe that once the plunger bias is sufficiently negative to accumulate free holes in the QW, a portion of them will tunnel onto trap states, located in or around the plunger gate dielectric. This phenomenon has been observed in both Si/SiGe and Ge/SiGe heterostructures for QWs at comparable to depths the one used in this thesis [120, 121]. Once occupying traps, the holes may then move onto the plunger gate to recombine with electrons, either via further tunneling or through drift if the trap is highly mobile. The continuous removal of charge from the QW and finite trap lifetime makes it possible for this process to reach a steady state.

This picture explains well why the onset of negative current always precedes turn-on, because both phenomena require hole accumulation. Referring to Fig. 5.10, we also notice that with higher V_{sd} bias, the magnitude of negative current diminishes. This is consistent with the proposed mechanism, because tunneling to the gate directly competes with conduction between the ohmics. By increasing V_{sd}, the driving force behind ohmic conduction increases, while the driving force for tunneling, stays the same therefore the positive current contribution becomes stronger and reduces the magnitude of the measured negative current [3]. We confirm this behavior

[3]The negative current coming from trap recombination when $V_{sd} < 0$ is masked by the negative current coming from conduction, so we cannot observe this trend here.

robustly by collecting I-V curves for all devices in the grid over a range of positive V_{sd}. This data is presented in Fig. 5.11. Therein, each data point corresponds to an averaged value across all measured devices. In all grid sweeps, at least half of the devices turn on (more than 350 devices), making these averaged values statistically significant. Error bars represent standard deviations.

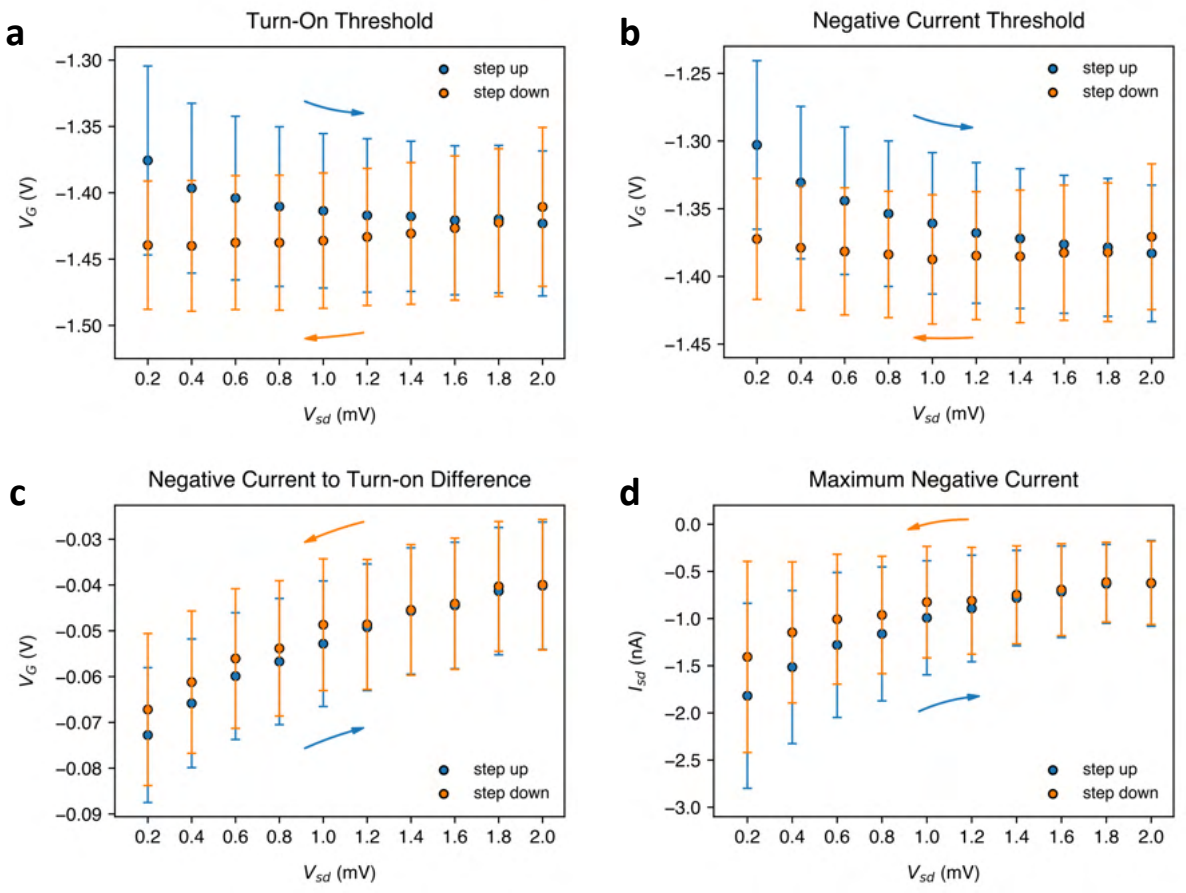


Figure 5.11: Average threshold voltages for **a)** turn-on, and **b)** negative current, of devices in the grid, measured over a range of positive source-to-drain biases. The measurement order involves first collecting I-V curves of all devices in the grid, then stepping V_{sd}, then re-measuring the whole grid. The 'step up' is carried out first, with V_{sd} incrementing positively, followed by the 'step down', with V_{sd} incrementing negatively. A resolution of 50 voltage points is used in each device sweep between -1.0 and $-1.5\,\mathrm{V}$. The threshold for negative current was set at $I_{neg} \leq -100\,\mathrm{pA}$. **c)** Difference between average turn-on and negative current thresholds. **d)** Average maximum negative current.

Evidently, as V_{sd} increases, the maximum measured negative current decreases, as shown in Fig. 5.11 d), consistent with Fig. 5.10. Furthermore, since the positive current is enhanced with greater V_{sd}, it overcomes the negative current sooner, which is why the gap between the two thresholds also shrinks, as shown in Fig. 5.11 c). Furthermore, due to both ohmic conduction and tunneling relying on hole accumulation, the trend in maxium negative currents mirrors that of V_{th} in Figs. 5.11a/b.

In support of the trap-assisted recombination mechanism, we introduce the SRH model, invented by William Shockley, William Thornton Read, and Robert N. Hall in 1952 [122]. The central equation of the SRH model, a derivation of which may be found in Appendix D, is the steady-state recombination rate of electrons and holes over a trap level located in a semiconductor

bandgap:

$$R = \frac{C_n C_p (np - n_i^2)}{C_n(n + n_1) + C_p(p + p_1)}. \tag{5.9}$$

Here, C_n (C_p) is the probability per unit time that an electron (hole) in the conduction (valence) band will be captured for the case that all traps are empty, n (p) is the density of electrons (holes) in the conduction (valence) band, n_1 (p_1) is the electron (hole) density at the conduction (valence) band if the Fermi level were to fall at the trap level and n_i is the intrinsic carrier concentration. The SRH model informs us of three ways in which the electron-hole recombination rate can increase: (i) the probabilities of conduction electrons or valence holes to be captured by traps increases (C_n or C_p increase), (ii) the density of charge carriers at the band edges increases (n or p increase), or (iii) the energy difference between the trap level and the band edges decreases (n_1 or p_1 decrease). Most relevant to the operation of SHT devices during a rapid I-V sweep will be p, as this is directly influenced by the plunger voltage. By increasing V_G, the availability of charge carriers p to participate in trap-assisted recombination, which qualitatively explains why the negative current we measure increases for larger V_G before turn-on. As for n, as it is not the density of conduction electrons in the semiconductor that holes recombine with, but rather we propose it is with the electrons on the gate metal, n is very large compared to p, and therefore unlikely to affect the recombination statistics much.

The decay of negative current with time can also be explained by hole trapping. We recall that fixed-charge and oxide traps have low mobilities and long lifetimes. When donor-type traps of this type are filled, they screen the negative electric field of the plunger gate from the QW, leading to a reduction in the number of accumulated free charges. This would lead to a smaller measured negative current through a reduction in p. The long time scales over which the negative current plateaus, not possibly explained by the capacitor model, therefore makes sense within a recombination model, because the capture of holes by oxide traps is a slower process compared to hole accumulation. If trap-assisted recombination is indeed responsible for these negative current, as we propose, then tunneling as a cause for the leakage measured in Figs. 5.5 $ 5.6 also becomes highly probable.

5.3.3 Device drift

Device drift refers to any process that changes the performance of a device through an unintended change to its electrical properties over time. So far, we have omitted the mention of device drift from our measurement analyses. However, we believe its effect is significant and responsible for: the shift in ρ_{xx} and n with V_G in Fig. 5.3a/b, the slow current decay in Fig. 5.9, the difference between V_{th} and V_{po} in Fig. 5.10 and the V_{sd} trends and hysteresis in Fig. 5.11a/b. The way we observe drift in SHT devices is through shifts in threshold voltages and currents. Following the first grid characterizations, we attempted to measure pinch-offs of individual barrier gates, as well as isolate Coulomb oscillations in preparation for charge noise measurements, however, device drift made this an impossible task. Hence, we dedicate the remainder of this thesis to this topic. We believe that device drift shares origins with the trapping mechanisms mentioned above, and therefore we find it an indispensable tool to better understand material disorder. In this section, we use the statistical power of grid measurements to bring new insights into the causes of drift and how it could be mitigated. These results are also precursory to Sec. 5.3.4, where we introduce new ways to quantify device stability.

We begin our discussion by highlighting another observation on V_{th} and I_{sd}^{max} from grid sweeps. As seen in Figs. 5.5 & 5.7, there is a clear pattern in moving from left to right along grid rows. To confirm whether this is due to fabrication error, leading to a tendency for left-hand-sided devices to turn on sooner, or a case of drift, we conduct a grid sweep in two directions. In the first case, we sequentially sweep devices in the grid as normal, from left to right along

rows and down, from device (1,1) to (36,18). In the second case, measured immediately after, we sweep the grid backward – from right to left along rows and up, from device (36,18) to (1,1). If the pattern is due to fabrication error, then its orientation should not change with respect to the sweep direction.

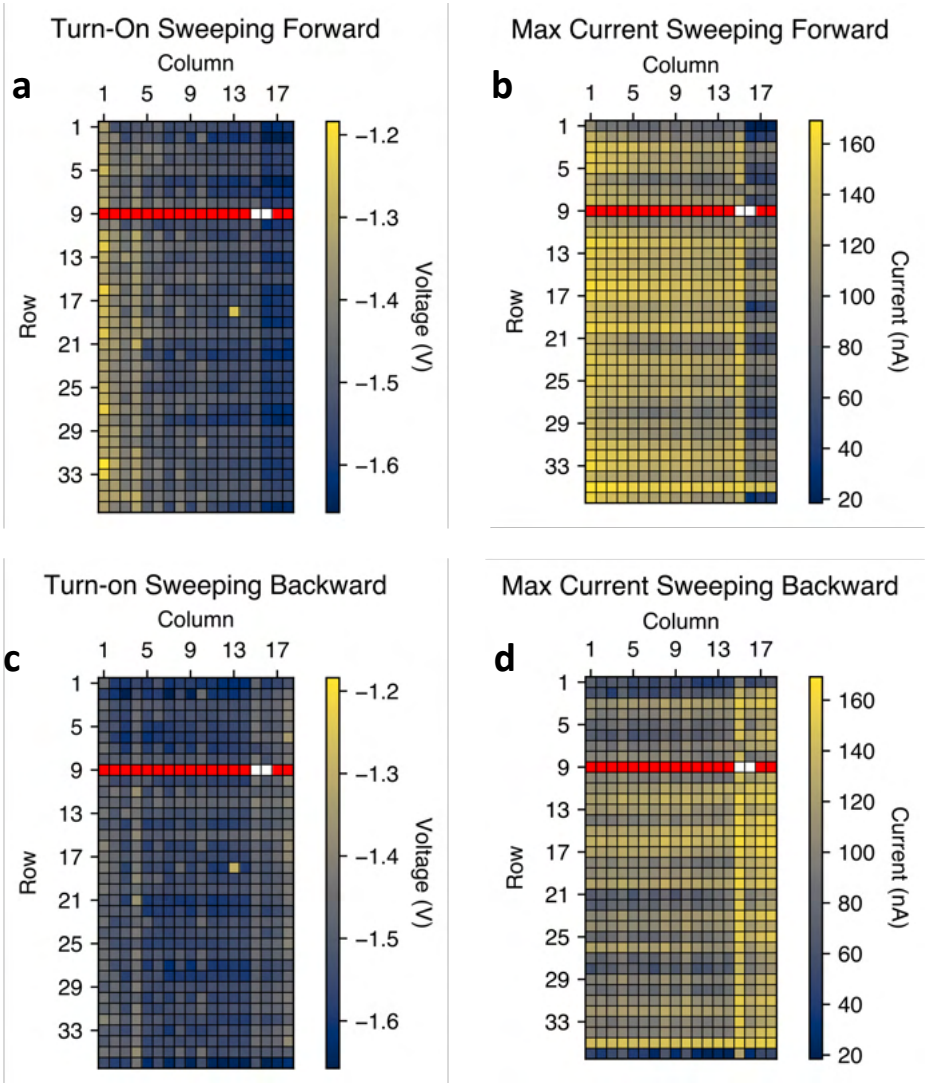


Figure 5.12: a/c) Turn-on threshold voltages and **b/d)** maximum currents of each device in the grid from a plunger sweep over 100 points between -1.0 and $-1.7\,\mathrm{V}$, and $V_{barrier} = V_{screen} = -500\,\mathrm{mV}$. Top figure panels correspond to devices being swept sequentially from left to right in a row and down, and the bottom figure panels for the sweep done right to left in a row and up. Red squares represent leaky devices, and white squares are devices that neither turn on nor leak.

As is clear from Fig. 5.12, the pattern does indeed flip orientation, both in the V_{th} and I_{sd}^{max} and must therefore be caused by drift. Since the plunger is shared across a row, those devices measured last in a row will experience prolonged voltage exposure. We believe that this causes the number of holes that populate deep traps in the oxide and at dielectric interfaces to increase. In general, these holes do not need to reach the gate dielectric via the QW as we propose in trap-assisted recombination, but could come directly from the surface of the ohmics or from the surrounding heterostructure. This leads to a screening of the electric field from the plunger for all $V_G < 0$, making it increasingly difficult to accumulate holes in the QW. As a result, V_{th} increases for devices measured further along a row. Since the plunger is always swept to the same value, in this case $-1.7\,\mathrm{V}$, these devices are therefore less likely to reach current

saturation, which explains the similar reduction in I_{sd}^{max} along the row. With this information, we can now also confidently attribute the skewness of histograms presented in Fig. 5.7 c/d to device drift.

We also observe that this trend decays along the row. This means that, at some point in the row sweep, the number of traps that are captured and released equilibrates. Nonetheless, since V_G is always returned to 0 V when switching devices, but we never fully recover the V_{th} of the first devices measured in a row, there must exist a fixed number of trapped charges that are never released over the time scales of individual device measurements.

We verify the theory of oxide charge trapping with two further measurements. First, we decouple drift from the grid by repeatedly measuring a single device. This is shown in Fig. 5.13a. As expected, we see a similar reduction in V_{th} and I_{sd}. Moreover, we see a convergence of the I-V curves with increasing sweep number, consistent with the row-wise decay over the grid. This confirms that the shared plunger is affecting all devices in the row as if they were each individually being operated.

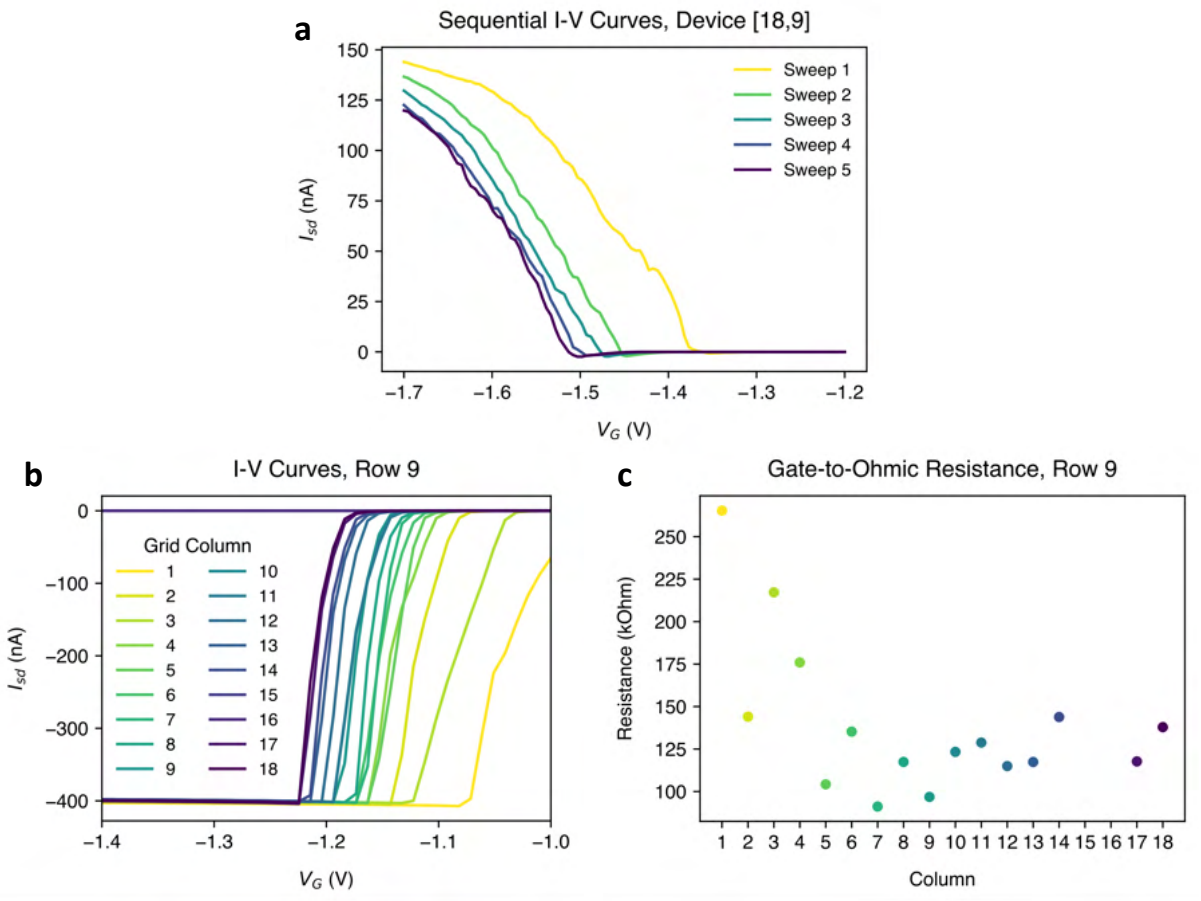


Figure 5.13: a) Sequential I-V curves of a single device in the grid over 100 plunger voltage points between −1.2 V and −1.7 V, and with barriers and screening gates both set to −500 mV. **b)** I-V curves of leaky devices in row 9 from a plunger sweep over 50 points between -1.0 and −1.5 V. **c)** Gate-to-drain resistances extracted by a linear fitting to I-V curves in b.

Next, we retrieve the I-V curves over the leaky devices in row 9. This may appear counter-productive at first, because leakage is generally undesirable, but since leakage directly probes transport across the dielectric, we can use it to our advantage and understand how the chemistry of our gate oxide changes as a row measurement progresses. We present the collection of device sweeps from row 9 in Fig. 5.13b. Since leakage, much in the same way as transport in the QW,

requires a sufficiently strong field, it is unsurprising to see the same trend in the leakage threshold as with the turn-on threshold of non-leaky devices. Over longer voltage exposures, holes become increasingly trapped in the oxide, thereby screening the plunger field from the ohmics. This is why for devices swept further along a row, the threshold for leakage also increases, and shows a somewhat convergent behaviour.

Since the leakage current between drain and gate is ohmic with respect to V_G, we can extract its resistance by fitting a linear slope to the I-V curve. This is presented in Fig. 5.13c. Interestingly, the drain-to-gate resistance generally decreases for devices measured further along a row, which are the same devices that show later leakage thresholds. We make sense of this by realizing that a portion of the charges injected into the dielectric must be highly mobile. Therefore, once the leakage threshold is reached, these mobile charges can contribute additional free carriers for conduction. This is further evidence of device drift being caused by the filling of charge traps in the oxide.

We further analyze grid behaviour under both forward and backward sweeps by plotting correlations between the turn-on threshold, and both maximum positive and negative currents. From Fig. 5.14a, we can see that I_{sd} reaches a maximum around 140 nA for most devices, but those with V_{th} beyond ~ -1.5 V show decaying maximum currents. This is expected, because with the plunger sweep only going up to -1.7 V, these devices are unable to reach current saturation. Hence, this gap of ~ 0.2 V puts an upper bound on the voltage range required for most devices to reach their maximum currents.

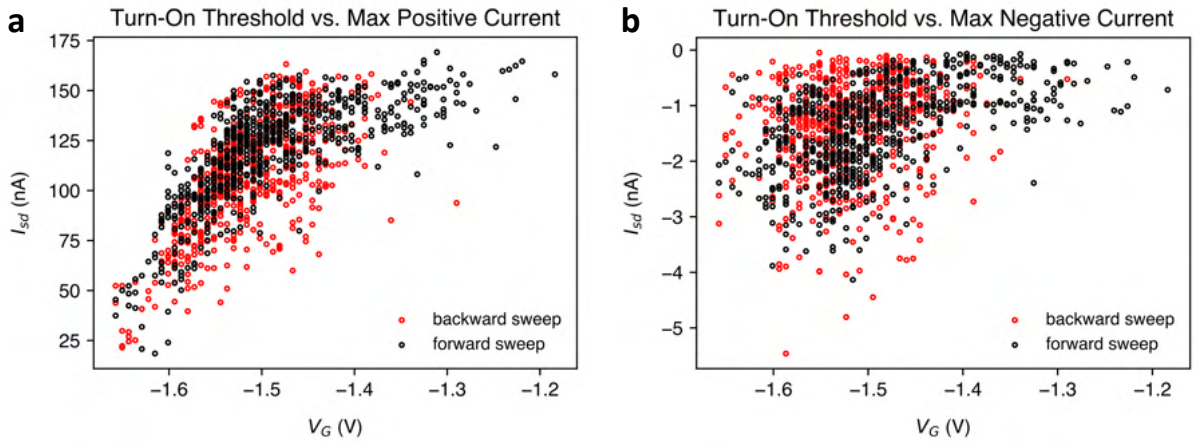


Figure 5.14: Scatter plots showing the correlation between the **a)** turn-on threshold and maximum *positive* current, **b)** turn-on threshold and maximum *negative* current, over grid sweeps carried out in the forward and backward directions.

Table 5.1: List of Pearson correlation coefficients between the turn-on threshold and maximum positive and negative currents over both forward and backward grid sweeps. Corresponds to data in Fig. 5.14

	Turn-on / max. current	Turn-on / min. current
Forward sweep	0.725	0.514
Backward sweep	0.649	0.203

From Fig. 5.14b, we learn that there is a weak correlation between the turn-on threshold and maximum negative current. Pearson correlation coefficients are given in Table. 5.1. This dependence could arise from there being a high concentration of donor-type interface traps located very close to the semiconductor valence band. The typically shorter lifetimes of interface

traps make them more likely to contribute to trap-assisted recombination than fixed charge and oxide traps. Interface traps also readily shift in energy under applied external electric fields. Hence, holes that accumulate in the QW at higher V_G gain a greater availability of interface traps to tunnel onto, which could contribute to the higher measured negative currents at higher V_G. We find no correlation between the maximum positive and negative currents (Fig. C.2), likely because the large spread in positive currents overshadows the comparatively smaller background of negative currents.

Comparing the overall scatter of data for forward and backward grid sweeps, we point out the absence of data in the low V_G region under the backward grid sweep (red markers). This means that even over the hour-long timescales that separate the repeated sweeps on individual devices, essentially, none are able to meet their original V_{th} from the forward sweep. This suggests that the lifetime of certain traps is even longer than the minute timescales gauged by the shifts in V_{th} across individual grid rows. We believe the reduction in Pearson correlation coefficients between forward and backward sweeps is an artifact caused precisely by this lack of data at low V_G.

Mitigating drift

At first glance, drift poses a great obstacle to using the grid architecture as a way to characterize device uniformity, due to device behaviour being strongly influenced by its voltage history. However, we find two ways to circumvent this by applying a sequence of preparatory voltage biases. This is the first time such schemes are being reported on a multi-device scale. The inspiration behind these schemes is based on two principles: (i) donor traps unfilling rapidly if the gate polarity is abruptly flipped, and (ii) devices reaching a near-equilibrium trap occupation after sufficient voltage exposure. We use these two principles to introduce 'shock' and 'hold' schemes for I-V measurements.

Shock:

1. Apply +1 V on the plunger for 2 s
2. Step the plunger to 0 V
3. Start measuring and sweep the plunger from 0 V to −1.7 V
4. Step the plunger back to 0 V
5. Select the next device and repeat steps 1–4

By shocking each device with a positive bias before measurement, we effectively erase its voltage history by forcing the release of trapped holes from within the dielectric. Then, we choose to always sweep from 0 V to help gradually re-introduce the electric field into each device. This reduces the chances of re-filling highly localized traps, which could compromise device transport further along the row, by giving time for charges to slowly redistribute.

Hold:

1. Apply −1.7 V on the plunger, −450 mV on all barriers and −250 mV on the screen[4], for 5 s
2. Step the plunger to −1 V
3. Start measuring and sweep the plunger from −1 V to −1.7 V
4. Step the plunger back to −1 V and select the next device
5. Repeat steps 2–3 for each device in the row
6. Select the next row and repeat steps 1–5

[4]having measured the grid more times since its initial characterization, we found these new and improved voltage points to operate at

By temporarily holding all devices in a row at an equal voltage, we encourage a homogeneous filling of charge traps across all devices. We choose a holding voltage high enough to cover the full energy range of traps that hypothetically could be filled during individual device sweeps. In this way, we artificially accelerate the drift of all devices equally before measurement. We disable the gate-zeroing by the cryo-mux when switching devices and choose a narrow sweep range on each device. This prevents undoing the induced drift and minimizes any dynamic response from the already-filled traps between measurements.

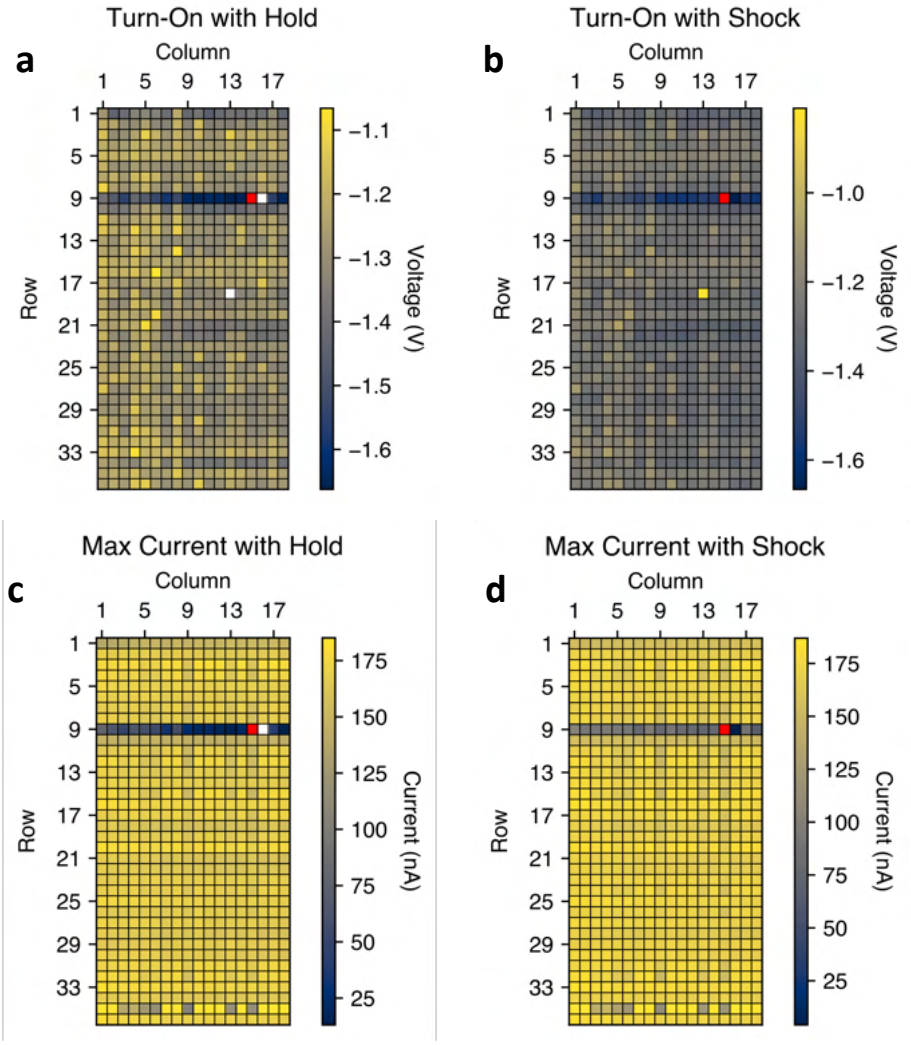


Figure 5.15: a/b) Turn-on threshold voltages and **c/d)** maximum currents of each device in the grid when measured using 'hold' (left) and 'shock' (right) schemes. $V_{barrier} = -450\,\text{mV}$ and $V_{screen} = -200\,\text{mV}$. Red squares represent leaky devices, and white squares are devices that neither turn on nor leak.

In Fig. 5.15, we show the results from the grid sweeps in the forward direction employing both shock and hold schemes. Undoubtedly, both schemes are excellent ways to counteract device drift; any trend in V_{th} or I_{sd}^{max} across grid rows appears to be almost, if not entirely, eliminated. Recovery (de-trapping) times in p-type Si/SiGe capacitors pulsed with 0.7 V have been reported at tens of seconds [123]. Our ability to recover V_{th} in a fraction of this time indicates that traps in our system may have smaller capture cross sections and thus be more responsive to applied electric fields. With these schemes, device variations are also better exposed, appear to be random, and not associated with any particular row or column, suggesting that local device

disorder could be dominant over variations caused by the crosshatch pattern.

With optimized gate voltages of $V_{barrier} = -450\,\text{mV}$ and $V_{screen} = -200\,\text{mV}$, were also able to bring the previously leaky row 9 into conduction, with the exception of a single device. We are thus more confident about leakage in this row being caused by a single defect on device (9,15). With lowered V_{screen} and $V_{barrier}$, the driving force for drain-to-gate leakage is reduced, allowing positive currents to dominate in the remaining devices in this row. Interestingly, however, we measure larger overall negative currents in the grid (see Fig. C.4). We owe this to the lowered V_{screen} and $V_{barrier}$ making the percolation threshold more difficult to reach, which allows tunneling over traps to persist for longer before holes are accumulated across the entire conduction channel. Overall, the successful operation of 647 of the 648 devices is an impressive feat and showcases the feasibility of fabricating a large grid of devices with high yield in an academic cleanroom.

We supplement these visualizations with histograms, where we directly compare the distributions of V_{th} and I_{sd}^{max} with the forward grid sweep from Fig. 5.12a/b. Due to the removal of threshold decay along rows, we are able to alleviate the V_{th} distribution of any skewness, which is especially pronounced in the reference histogram between $-1.4\,\text{V}$ and $-1.2\,\text{V}$. The distributions in V_{th} using both 'hold' and 'shock' schemes are much more uniform. Likewise, the 'resetting' of devices through the voltage shock enabled much earlier V_{th} due to reduced screening, which allowed devices to reach current saturation much sooner. This is reflected by the sharp and uniform I_{sd}^{max} distribution in Fig. 5.16b, which we deem to be the most accurate representation of our device uniformity. Since we collected 'hold' measurements directly after 'shock' measurements, 'hold' distributions enjoy a similar average V_{th} and narrow I_{sd}^{max}, because the voltage history of each row is mostly erased when the plunger is biased to $+1\,\text{V}$ when measuring the last device in each row. Very recently, using stress voltages of varying magnitudes and durations has been shown to enable hysteretic tuning of devices in Si, and bring V_{th} and V_{po} close to their initial values [124]. Therein, the authors also hypothesize that this behaviour is due to a cyclic de-trapping of charges in or close to the dielectric.

The average (standard-deviation) V_{th} we report for the reference, hold and shock measurements are: -1.489 (0.083) V, -1.275 (0.078) V and -1.248 (0.074) V. Likewise, for I_{sd}^{max}: 119.631 (26.212) nA, 163.808 (21.604) nA and 168.865 (16.673) nA. Albeit the presence of dielectric disorder, as evident through the spread in V_{th}, the especially low standard deviation in I_{sd}^{max} is a testament to the grid's high dimensional uniformity and homogeneous disorder landscape inside the QW.

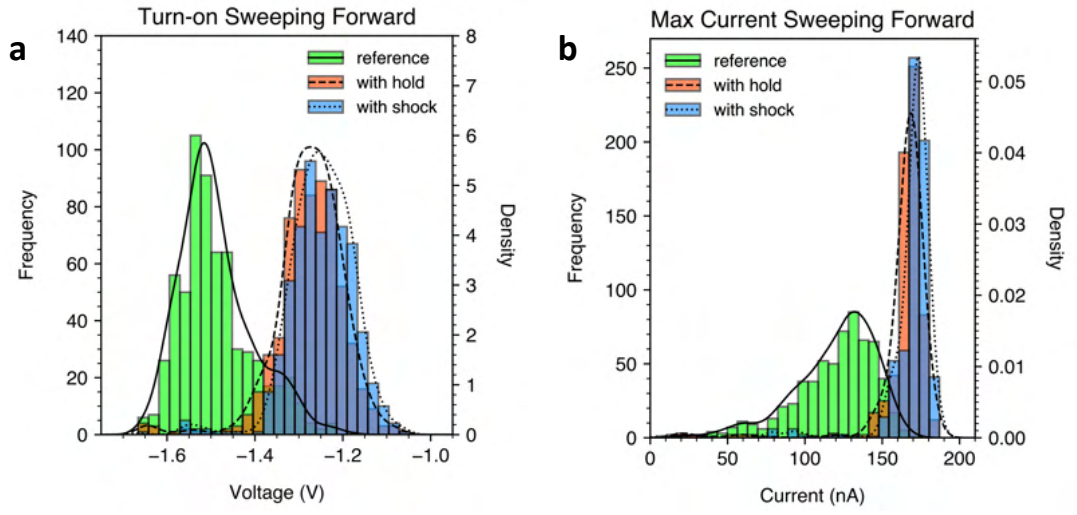


Figure 5.16: Histograms comparing the distribution of **a)** turn-on thresholds and **b)** maximum currents when sweeping grids using 'shock' and 'hold' schemes.

5.3.4 Current Stability and Trap Kinetics

As mentioned in Sec. 2.6.3, charge noise can be measured through fluctuations in SHT current, caused largely by a changing occupation of trap states. From this, one can gain insight into the overall level of disorder. However, such measurements rely on the average current being stationary – something we have been unable to achieve in any of the devices on the grid. Nevertheless, since the rate of device drift will inherently be linked to the density of traps, and the likelihood of populating them, we believe measuring drift can serve as a good alternative to measuring charge noise. As we have learned so far, drift can affect device performance on second to hour timescales. In this section, we quantify this behaviour.

The experiment we design investigates the device stability as a function of time and proceeds as follows. First, a reference I-V curve is collected with a high voltage resolution of 600 points from 0 V to −1.7 V with a delay of 0.01 s between points. Then, the plunger is held at its maximum value of −1.7 V, and the current is sampled 1500 times, with a delay of 0.1 s between samples. Considering a mild, but constant delay in the apparatus, this amounts to a total acquisition time of 211.5 ± 1 s. Next, the current at each point in this trace is mapped onto its corresponding voltage from the I-V curve. In this way, we can plot the effective voltage that the device is experiencing in time, in other words, the amount of voltage that is screened. Since I-V curves are not always monotonically increasing, we introduce a maximum step size during this mapping protocol to minimize the number of abrupt voltage jumps.[5] We choose a small subset of 5 devices to measure, located on the grid diagonal. We do this to avoid re-using plunger and barrier lines between measurements to minimize data bias. In addition, we use the 'shock' scheme on each measurement, which will become important for comparing mappings between grid samples. We repeat this over plunger sweeps up to −1.2 V, −2.2 V and −3.0 V. Pushing the plunger to a higher bias resulted in leakage, and any lower would result in most devices not turning on. To maintain a similar level of voltage resolution, we scaled the number of voltage points to 800 for the higher voltage sweeps. To keep the duration of measurements similar, we also scaled the voltage delay time accordingly to 0.0075 s such that all I-V curves were collected over exactly 6 seconds.

In Fig. 5.17, we show an example of the voltage mapping for devices with $V_G^{hold} = -1.7\,\mathrm{V}$. The current trace appears to be monotonically decreasing, with a few exceptions, which we believe are small ensembles of charges un-trapping, relieving the QW of some screening, and thus temporarily enhancing conduction. Perhaps the most striking result, however, is that after sufficient time ∼200 s, the current completely vanishes in all devices. This would indicate that we have filled a sufficient number of traps to screen the electric field of the plunger by enough to cut off accumulation in the QW. To show that this statement is not unreasonable, we apply the Coulomb model to a very simple scenario.

We consider a thin slice of dielectric, whose width is 1 nm and whose length is the ohmic separation ∼200 nm. Using a moderate guess of the trap density at $1 \times 10^{13}\,\mathrm{cm}^{-2}$ [125], the total number of traps in this slice area would then be 20. Let us imagine that we fill all these traps with a hole, such that they each contribute a positive charge $q = 1.602 \times 10^{-19}\,\mathrm{C}$. Next, we use the principle of superposition to solve for the electrostatic potential these holes would induce onto a single reference point inside the QW. We choose this point to be midway between

[5] This was initially done to help map voltages in the negative current region where there always exist two voltages that share the same current. However, the step size still had to be kept large enough to enable mapping to high voltages, where the current dropped the fastest. This resulted in the mapping becoming stuck in a local minimum near the maximum negative current point. An adaptive step sized based on the voltage gradient would have resolved this but is not something we had time to implement. We find it unlikely, however, that this extra mapped range would have changed the outcome of our later fittings by much, due to how narrow it is.

the ohmics. To do this, we solve,

$$U = \sum_{r} \frac{1}{4\pi\epsilon_0\epsilon_r}\frac{q}{r} \tag{5.10}$$

where r runs over all the distances between the trap and the reference point. Having chosen the slice to be thin, we collapse the problem to 2D, and express Eq. 5.10 as,

$$U = \sum_{y=55}^{72}\sum_{x=-100}^{100} \frac{1}{4\pi\epsilon_0\epsilon_r}\frac{q}{\sqrt{x^2+y^2}} \tag{5.11}$$

where x is the position of the trap along the dielectric slice and y is the vertical distance from the QW. Since we have 20 traps, we step x in increments of 10 nm, whereas for y, we step it in increments of 0.1 nm as this is the approximate thickness of each dielectric layer deposited via ALD. For simplicity, we do not impose a boundary condition at the semiconductor/dielectric interface to avoid having to solve the potential at multiple points along r. We use $\varepsilon_r = 15.2$ of $Si_{0.2}Ge_{0.8}$ to calculate a conservative guess for the potential[6]. Carrying out the sum yields 4.15 V, which would indeed be large enough to pinch off accumulation in the QW, even if only a fraction of the traps are filled. We warn, however, that this value is very sensitive to the trap density. On the other hand, this dependence is true regardless of the method used to calculate the potential, which means that by fabricating cleaner dielectrics, there is great scope to improve the stability of devices affected by drift due to screening by charge traps.

We can use this finding, combined with the decay time in the current trace, to make sense of Fig. 5.17c, where we show a series of I-V curves collected at different sweep rates on a single device. As the sweep rate decreases, V_{th} increases. We understand this as traps having more time to become populated at each voltage point when a sweep is conducted slower, which leads to a progressive screening of the plunger from the QW. For a delay time of 1 s (yellow), the full sweep takes >200 s and no turn-on is measured, which likely corresponds to a sweep rate that approximately matches the rate at which traps are populated. Beyond V_{th} at the remaining sweep rates, we note a similar qualitative transconductance (slope). This is rather unexpected as we imagine that slower sweeps would permit more trapping and therefore produce slower increases in I_{sd}. The absence of this may be due to a widening of the conduction channel with higher V_G, which would balance the screening effect. This would also explain why there is no obvious trend in the measured negative currents (Fig. 5.18c inset).

We now turn to the mappings in Fig. 5.17d, where we observe a striking dependence on the effective voltage with time. The shape of these mappings do not exactly follow the shape of the I-V curves, which reveals that the rate at which traps are populated is not linear. To better quantify this behaviour, we return to the SRH model. Therein, trapping kinetics are modeled as a first-order chemical reaction, which may be expressed as [122],

$$\frac{dp_t}{dt} = C_p(p_0 - p_t)p_i, \tag{5.12}$$

where p_t is the density of filled hole traps, p_0 is the density of available traps, and p_i is the number of injected holes in the oxide. The solution of this first-order equation is an exponential function of the form,

$$p_t(t) = p_0\left(1 - e^{(-t/\tau)}\right) \tag{5.13}$$

[6]There exist many other ways this problem could have been solved, such as by working in 3D and considering a discretized rectangular volume of traps, or a cylindrical one and working in cylindrical coordinates. Better yet, one could use an FE model to accurately include our device geometry and different electrical permittivities. In general, traps may also take on fractional charges, be modeled using a dipolar interaction [126], and their density will reduce deeper in the oxide. We found this method to be the simplest while still giving a good result.

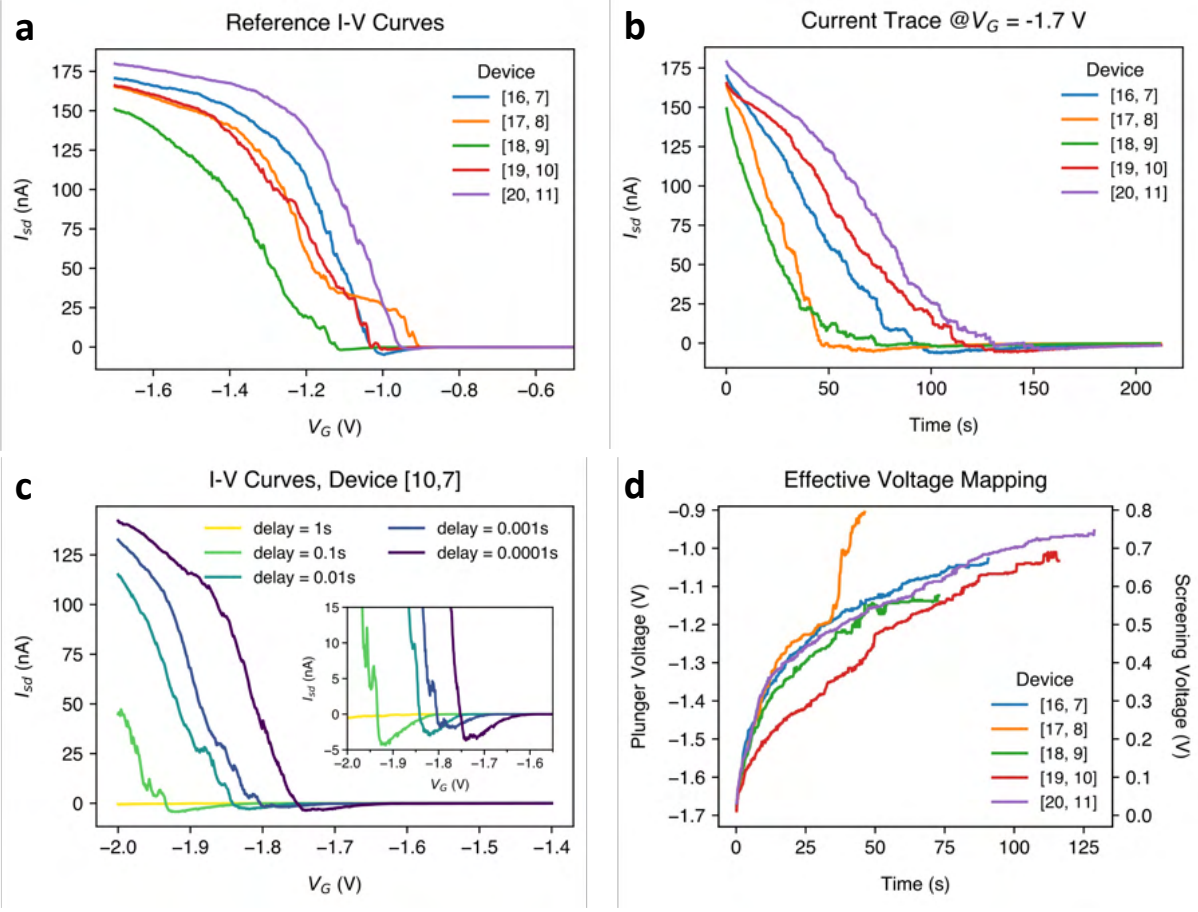


Figure 5.17: a) High resolution I-V curves collected on 5 devices over the grid diagonal with barriers set to −450 mV and screening gate to −200 mV. **b)** Current traces of the devices from a) when held at a constant plunger voltage of −1.7 V. **c)** A mapping between the current in b) and voltage in a). **d)** A series of I-V curves acquired over different plunger sweep rates on a single device with a constant voltage resolution of 600 points. The delay corresponds to the time between each voltage step.

with $\tau = (C_p p_i)^{-1}$. This simple model has been successfully employed to describe the charging of dielectrics in early works, measured via a shift in flat-band voltage [127], and also extended to include de-trapping and traps over various energies [128]. However, there is a major flaw with this model, which is that C_p is assumed constant. In reality, second-order effects such as Coulomb repulsion between traps [129], cause a reduction in their capture probability and therefore deviations of experimental data from theory [130]. We encountered the same issue when fitting our data, especially with the envelope of our voltage mapping. Therefore, we turned to a logarithmic decay model, as proposed by Wolters and Schoot [131].

To account for the decreased capture probability as the space-charge density in the dielectric increases, one can assume that a filled trap inactivates a fixed volume h of the dielectric from further trapping. The remaining free volume in which trapping can occur then becomes $V - h$. If trapping is an independent stochastic process, then the factor by which C_p decreases follows a power law $(1 - h/V)^{p_t}$ [131]. Using the approximation for $h \ll V$,

$$\left(1 - \frac{h}{V}\right)^{p_t} \approx \exp\left(-p_t \frac{h}{V}\right), \tag{5.14}$$

the revised rate equation then becomes,

$$\frac{dp_t}{dt} = C_p(p_0 - p_t)p_i \exp\left(-p_t\frac{h}{V}\right). \tag{5.15}$$

Under the further assumption that $p_t \ll p_0$, such that p_t in the first factor vanishes, the solution to Eq. 5.15 can be found easily via separation of variables:

$$\begin{aligned} \int_0^{p_t} \exp\left(p_t\frac{h}{V}\right) dp_t &= \int_0^t C_p p_0 p_i \, dt \\ \frac{V}{h}\left(\exp\left(p_t\frac{h}{V}\right) - 1\right) &= C_p p_0 p_i t \\ \Rightarrow p_t &= \frac{V}{h}\ln\left(\frac{h}{V}C_p p_0 p_i t + 1\right) \end{aligned} \tag{5.16}$$

Comparing this result with Eq. 5.12, accounting for a changing capture probability has resulted in an overall slower capture rate. Moreover, this model predicts that if the interaction of traps with their environment were to increase ($\uparrow V/h$), there should be a proportional reduction in the overall capture rate. This derivation is not yet rigorous, however, as it still assumes that the number of carriers in the dielectric p_i is constant. In general, the *fluence* Q of holes to the dielectric depends on their flux J, which in turn depends on the electric field E at the QW.

$$Q = \int_0^t J_0 \exp\left(\frac{E}{E_0}\right) dt' \tag{5.17}$$

With increased screening as trapping progresses, the quantity E will decrease, thereby further reducing the rate of p_t. Mapping between p_t and our effective voltage therefore becomes a highly non-trivial problem.[7] To the best of our knowledge, no models exist to physically motivate such data fitting, so we take the liberty to introduce our own empirical model, while still maintaining a logarithmic dependence:

$$V_{eff} = a\ln\left(\frac{(t+b)}{(t+b)+1}\right) + c. \tag{5.18}$$

In the above, a, b and c are constants to be fit. The greatest deviation from Eq. 5.16 comes from the modified argument of the logarithmic function. We do this to bound our fitting function to physically reasonable limits. Renormalization in the denominator ensures that the fitted voltage never exceeds 0 V, even for $t \to \infty$, as this would correspond to holes trapping against the electrostatic potential gradient. The b parameter applies a rigid shift on the fitting, which is necessary for we do not initialize our current trace from $V_G^{hold} \to \infty$. The c parameter is introduced as a way to approximate the effective voltage at infinite time, which gives a picture of the total number of filled traps over the energy interval between 0 V and the V_G^{hold} during the current trace collection. This makes the c parameters particularly powerful, especially for devices that drift rapidly, as it unlocks energy information below V_{th}, which one could not gauge from a regular I-V or current trace measurement. The a parameter, similar to the coefficient in Eq. 5.16, still reflects the influence of already-filled traps on their environment. We set parameter bounds of $p_{min} = [0, 0, -3]$ and $p_{max} = [40, 40, 0]$ for $p = [a, b, c]$. Overall, we find excellent agreement between this fitting function and our data (Fig. C.8), as well as robustness over different fitting ranges. This highlights the adequacy of a logarithmic process, consistent with trap-filling kinetics, for describing the voltage mappings we construct.

[7] From HB measurements, we know that the carrier concentration in the QW n, and thus the likely number of holes in the dielectric p, will scale linearly with V_G (Fig. 5.3). By the Coulomb model, p_t would also cause a linear screening of V_G. We attempted to substitute these relations into Eq. 5.15 to solve for V_G directly, however, this yielded non-elementary functions as solutions.

In Fig. 5.18, we show the decay times for the current to reach the turn-on threshold for various V_{Ghold} alongside fitted effective voltage mappings using 5.18. We note that we have compiled data from two different grid samples. The first, called 'reference', is the same grid that has been discussed so far, while 'UV treated' is a grid from a different die that we exposed to UV light for 30 minutes. Grid sweeps of this sample maybe be found in Fig. C.5. Given that UV treatment has demonstrated the ability to decrease sample resistivity [132] and our group has observed reductions in HB background currents after implementing UV treatment, we believed it could enhance device stability and decide to introduce it here as point of comparison.

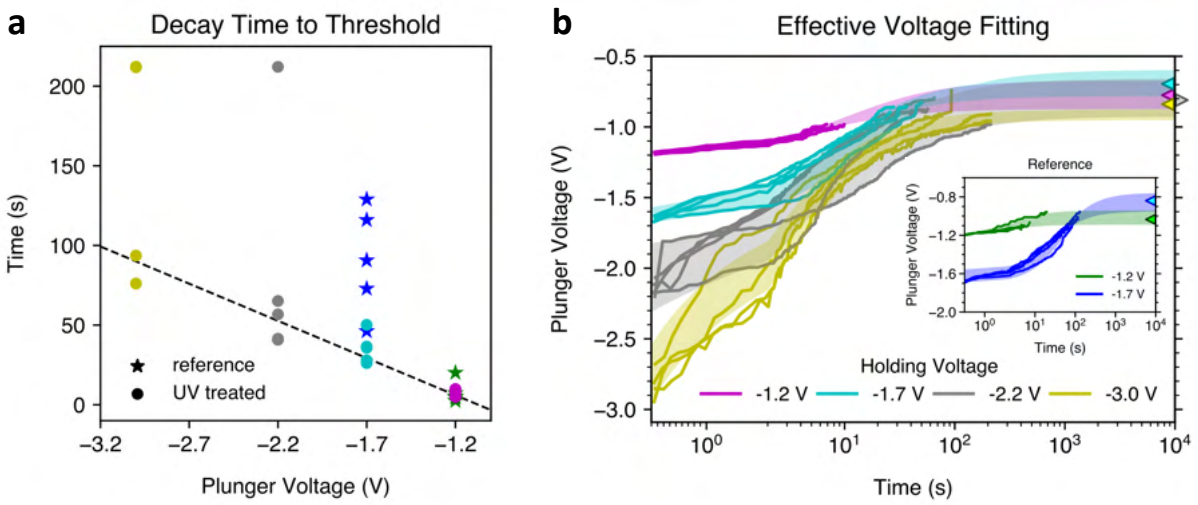


Figure 5.18: a) Time taken for devices to reach V_{th} over a current trace with plunger voltages held at $-1.2\,\mathrm{V}$, $-1.7\,\mathrm{V}$, $-2.2\,\mathrm{V}$ and $-3.0\,\mathrm{V}$. **b)** Maps of the effective voltage induced by the plunger. Shaded regions show the envelope of the fitted functions using Eq. 5.18. Triangular markers are positioned at the average c value for each holding voltage. We use a logarithmic time axis for visual clarity.

In both samples, we observe longer decay times for higher holding voltages. We believe this to be due to a greater number of traps that must be filled to screen the plunger until V_{th}. Moreover, donor-type trapping rates decrease exponentially with their energy from the valence band edge, and can range from pico-second to years [133]. This suggests that during I-V sweeps, we already fill most shallow traps and it is mainly traps near the band-gap center that we are populating when holding V_G, and this becomes slower for higher V_G^{hold}. In addition, the progressively larger number of filled traps will likely restrict the trapping rate further due to more abundant Coulomb interactions. Excluding outliers, we give a visual aid to show an approximately linear trend in the decay times on the UV-treated sample. In the measurements at $V_G^{hold} = -3.0\,\mathrm{V}$ and $V_G^{hold} = -2.2\,\mathrm{V}$, three devices never reach pinch-off within the allocated 211.5 ± 1 seconds, while the remainder all pinch off below 100 seconds. On the other hand, the I-V curves of all these devices are remarkably similar (Fig. C.7). This shows that quantifying device drift in this way bears more information on the level of dielectric disorder than plain I-V curves, and may help locate devices that are in a particularly disordered environment.

We also notice that the devices that perform best were consistent across different V_G^{hold}, indicating that the local density of traps may have a minimal effect on how they are distributed in energy. In other words, if a device has few traps located low in energy, then it is likely that the same device also has few traps located high in energy. This isn't to say, however, that the energy distribution itself is uniform. We can learn of its shape by closer inspecting our fitted parameter values, which are summarized in Table. 5.2.

We firstly focus on $V_G^{hold} \in \{-1.7\,\mathrm{V}, -2.2\,\mathrm{V}, -3.0\,\mathrm{V}\}$ from the UV treated sample. We caution that our analysis goes under the assumption that the amount of charges that are trapped

Table 5.2: Average decay times for device currents to reach their turn-on thresholds and averages of fitting parameters of Eq. 5.18 used to fit the effective voltage maps in Fig. 5.18. Numbers in brackets indicate standard deviations.

Holding Voltage (V)	Decay Time (s)	a	b	c
−1.2 Ref.	7.83 (6.26)	0.516 (0.316)	2.397 (1.308)	-1.034 (0.055)
−1.7 Ref.	90.80 (29.66)	29.299 (13.163)	35.300 (12.356)	-0.840 (0.085)
−1.2 UV	7.61 (1.82)	3.580 (1.257)	7.728 (1.836)	-0.775 (0.072)
−1.7 UV	35.15 (8.48)	8.810 (4.710)	8.577 (3.970)	-0.695 (0.069)
−2.2 UV	83.11 (65.09)	7.969 (4.524)	6.206 (5.306)	-0.780 (0.089)
−3.0 UV	148.300 (63.90)	5.202 (0.846)	2.137 (0.328)	-0.912 (0.027)

during the collection of I-V curves is negligible compared to the dynamics that unravel from when the current trace begins. For this reason, we choose to exclude data at $V_G^{hold} = -1.2\,\mathrm{V}$ from our analysis, because its decay times are on similar timescales to the I-V acquisition time. This explains why data for $V_G^{hold} = -1.2$ behaves as an anomaly, despite its low data spread. With this, the first observation is that c decreases with more negative V_G^{hold}, although the difference between c and V_G^{hold} grows. The average differences are 1.005 V, 1.420 V and 2.088 V. This would suggest that the number of traps that become available for filling increases with greater plunger bias, in agreement with our argument for longer decay times. Parameter b also decreases, which accommodates the deeper starting point in voltage of our mappings with more negative V_G^{hold}. Parameter a is the only parameter that influences the shape of the fitting function. It decreases for more negative V_G^{hold}, which we interpret as an acceleration in the initial trapping rate. The exact cause of this change is unknown, but we speculate it could be from a combination of an increased hole density in the QW, thus increasing p_i, as well as higher hole energies. Higher hole energies have been shown to increase their capture probability [134], and this may be occurring in our system when the accelerating voltage during tunneling between the QW and dielectric is increased by higher V_G^{hold}.

We also highlight that those devices with longer current decay times, and correspondingly greater a and c parameters, were the same devices to measure higher I_{sd}^{max} (see Fig. C.7). Since the saturation current is limited by mobility, we believe that our measured c parameter could thus be reflective of the number of remote impurity scatterers, and stem overwhelmingly from charge traps. Also, the margin by which c differs between devices is often much greater than their differences in I_{sd}. Hence, we believe our voltage mapping technique could serve as a supplement to accurately probe remote disorder. A key advantage here is that the voltage mapping takes place on a device-by-device basis, at the nanoscale, and requires no modification to the experimental set-up. While Hall bar measurements provide excellent disorder metrics μ and n_p, these are average values collected over large gate areas that are forgiving to sporadic nanoscale defects, and do not account for device drift. By combining conventional charge transport measurements with dynamic device behaviour, one can potentially gain a more holistic view on material disorder.

We now compare the reference and UV-treated samples, both held at −1.7 V. Evidently, the reference measurements have much larger average a and b parameters, although a more negative c parameter. This is indicative of a decelerated trapping process, as well as fewer total traps to be filled. This goes against our original intentions with UV treating the grid, where we hoped to reduce the trap density. This corroborates the measured decay times, which tend to be shorter for the UV treated sample than the reference, indicating a worsened device stability. We see this behaviour also in full grid sweeps (Fig. C.5) where the row-wise trend in V_{th} and I_{sd}^{max} decays much quicker in the UV treated sample compared to the reference, which would

indeed correspond to a faster trapping rate. We hypothesize that the better current stability in reference devices could stem from an excessive UV dose, which created an oxidizing environment that formed additional traps rather than neutralizing existing ones. On the one hand, this could be seen as advantageous, especially if using the 'hold' scheme, because then a homogeneous trap filling across devices rows can be reached more readily after UV treatment.

Lastly, we address the generally large spread in decay times between devices, which carries over to our fitting parameters, especially a and b. These often have standard deviations in excess of 50% of the mean. Conversely, parameter c is more resilient against device variability and has standard deviations within around 10%. This is one downfall of our chosen fitting function, which can produce similar y values by moving a and b in the same direction. A solution would be to fix b depending on V_G^{hold} and make only a and c free parameters. With measurement repetition and a greater sample of devices, however, we expect the standard deviations of a and b to go down.

6

CONCLUSIONS

6.1 Summary

In this thesis, we set out to investigate statistically the impact of material disorder on the performance of quantum devices in SiGe heterostructures. This work is of particular interest for scaling spin-based quantum computers due to the growing number of qubits in these materials. With larger processors, the probability of any one qubit landing on an atomic defect increases. Furthermore, the presence of a crosshatch pattern on SiGe, which we measure to have a wavelength of 1.655 µm and maximum peak-to-valley height of 5.420 nm, poses a further challenge for processors that exceed this size.

To probe disorder over these various length-scales, we fabricate the first grid of 648 single-hole transistor (SHT) devices on a Ge/SiGe heterostructure using a crossbar architecture. We encounter two main challenges during its fabrication, namely, a systematic over-sizing in critical dimensions, as well as frequent fracture of barrier gates. We resolved these issues by revising our design dimensions and eliminating etchant flow via pipettes during lift-off. We also raise concerns over pillar-shaped particles on sample surfaces, however later confirm that these did not inhibit the operation of our SHTs.

Using SEM images, we characterize the width uniformity of the plunger, left barrier, and right barrier, and find narrow distributions of 148.6± 5.1 nm, 52.2± 4.2 nm and 51.5± 3.5 nm. Under optimized conditions, we are able to turn-on 647/648 of our SHTs, and measure voltage thresholds and maximum currents at -1.248 ± 0.074 V and 168.865 ± 16.673 nA. Overall, the excellent device yield and great dimensional uniformity highlights the robustness of our fabrication flow. Performance variability appears to be random across the grid, suggesting that local disorder may be dominant over disorder caused by the crosshatch under our specific operating conditions.

During I-V sweeps, we find reproducible negative current behaviour prior to turn-on in all SHTs. We theorize this as being due to either capacitor charging as holes hop from the drain reservoir into the buried Ge quantum well (QW), or from holes tunneling between the QW and the plunger top gate. As a capacitor model could not explain the steady state nature of the negative current, nor its persistence during pinch-off sweeps, we claim trap-assisted recombination as the most probable cause. This is corroborated but a high susceptibility for leakage that we measure in the grid, as well as severe current drifts and hysteresis. Since the latter is typically attributed to charge trapping at the semiconductor/dielectric interface and in gate oxides, trap-assisted recombination reconciles these two observed behaviours well. Moreover, we measure reduced drain-to-ohmic resistance on devices with increased drift, which further points

toward an increased carrier density in the dielectric.

Device drift manifested itself as a row-wise decay in turn-on thresholds V_{th} and maximum currents $I_{sd_{max}}$ in the grid, which we verified by switching the grid measurement direction. This pointed to drift occurring on second-to minute timescales, however, even with certain device measurements being separated by more than 2 hours, original V_{th} and $I_{sd_{max}}$ could not be reached. This indicates a much slower drift process, and, if caused by oxide charge trapping, corresponds to trap lifetimes exceeding hours. The severity of drift prevented isolation of Coulomb oscillations and the measurement of charge noise across the grid – an initial goal of this thesis. Nevertheless, two schemes are proposed to overcome drift, which we call 'shock' and 'hold'. When implemented, both yield superior device uniformity and alleviate data skew.

We quantify device drift by introducing a voltage mapping technique and interpret its results using kinetic theories of charge trapping. Remarkably, we find that SHT currents vanish completely after $\sim$200 s when holding their plungers at $V_{G_{hold}} = -1.7$ V. In general, we also find that devices held at higher V_{Ghold} have longer current decay times, albeit an accelerated initial trapping rate. Through a fitting, we interpret this as an increased trap density at higher energies. With the aim of improving device stability, we compare UV-treated devices with untreated ones but unexpectedly find inferior stability, likely due to sample oxidation.

In conclusion, we confirm the utility of a crossbar architecture for statistically characterizing material properties and fabrication yield, but find device stability a major issue that would require attention for the reliable operation of future quantum devices in Ge/SiGe heterostructures.

6.2 Further Work

To tackle device drift, a more thorough investigation into charge-trapping is needed. Three methods could assist with this: deep-level transient spectroscopy (DTLS), capacitance-voltage (CV) measurements, and charge pumping. DTLS works by studying the transient gate capacitance over different voltage pulses, while CV measurements use an AC gate voltage to study the response time of traps through a changing capacitance. Similarly, charge pumping measures the current produced by charges (de-)trapping as a function of an AC gate voltage. Any one of these methods is able to extract trap lifetimes as well as trap densities over a range of energies, which could help verify the observations made in this thesis. DTLS can also yield capture cross-sections. With this information, the location of traps, and their chemistry, can be better understood, opening the door towards a fabrication process that is informed by our specific disorder type. To improve our existing voltage mapping technique, however, a positive voltage shock on the plunger could also be used prior to applying the holding voltage. In this way, the effect of shallow charge trapping, which will have partially occured over the duration of the I-V curve acquisition, can be exposed.

Several improvements could be made to our existing fabrication flow to reduce the trap density. For example, thermal treatment could be used to minimize residual stresses and atomic disorder in our heterostructure through diffusion, thus providing fewer sites for charges to trap. Annealing in forming gas could also help to passivate existing traps. In addition, the number of interfaces at which traps may form can be reduced simply by removing our Si capping layer [30]. Alternatively, the Si could be replaced by a Ge capping layer, which could act as a boundary for tunneling between the QW and the dielectric [120], or the dielectric itself could also be improved. Using HfO_2 or HfO_2/Al_2O_3 has been shown to yield a lower interface trap density due to a 'self-cleaning' process [135]. Since traps lie in the material band gap, one may also consider reducing the Si content in our SiGe spacer layer to reduce the bandgap energy. This will have the added benefit of reducing the extent of strain relaxation due to better lattice matching, thus reducing the dislocation density and roughness of the crosshatch pattern. Bulk leakage

currents have also been correlated with an increased threading dislocation density [136], which could induce noise around the QW environment. Changing SiGe compositions could therefore be worth considering in future work, but being careful to balance this change with the trade-off in band-offset, which has implications on the extent of heavy-hole light-hole splitting.

With these improvements, we hope that future experiments will be able to operate in a regime that is minimally affected by device instability. Since charge trapping is also a thermally activated process, we recommend carrying out experiments at lower temperatures than 1.7 K, such as in a dilution refrigerator. By isolating Coulomb oscillations, we hope that the grid architecture can be useful in also statistically characterizing charge noise. During these measurements, we believe it could be interesting to sweep the grid in different ways, including top-to-bottom and through random device sampling. By minimizing the repeated use of shared plunger lines, we believe a less biased statistic on the performance of an individual device can be collected. Moreover, we recommend future grid fabrication runs to align the crosshatch pattern with the grid diagonal, such as to de-couple its contribution from point defects that may lay on shared plunger or barrier lines, thereby dominating device performance across rows and columns. In this way, and under improved fabrication processes, we hope to learn if the crosshatch may still limit device performance.

To make sense of device variability, one may engage in finite-element simulations. To model the impact of dimensional variations and heterostructure thicknesses on turn-on voltage thresholds, the NextNano software is very well suited. Using its Schrödinger-Poisson solver and compatibility with Python, simulations can be run over hundreds of device dimensions easily. NextNano can also provide energy-level spacings of quantum dots. Understanding deviations in level spacing can be useful when collecting charge noise measurements across many devices that may succumb to size variations. NextNano also allows for the modeling of the band structure in the presence of nearby charge defects. Combining these two sources of error would be an excellent way to validate experimental results against theory. In addition, collecting SEM images after measurements can help reveal the root for anomolous behaviour, such as the leakage on device (9,15), which we hypothesize is due to a discontinuity in the dielectric.

Finally, one can use the design proposed in this thesis for Si SETs to benchmark device performance between Ge and Si systems. In this way, we hope to bring statistical confidence into the metrics that favour either of these systems and to bring relevance to the performance of these devices at length-scales that meet the needs of future quantum processors.

REFERENCES

1. Field, J. V. & Wright, M. T. Gears from the Byzantines: a portable sundial with calendrical gearing. *Annals of Science* **42,** 87–138 (1985).
2. Intel. *The story of the Intel® 4004* `https://www.intel.com/content/www/us/en/history/museum-story-of-intel-4004.html#:~:text=In%5C%201971%5C%2C%5C%20the%5C%20Intel%5C%C2%5C%AE,technology%5C%20held%5C%20560%5C%20million%5C%20transistors.` (2022).
3. Apple. *Introducing M1 Pro and M1 Max: the most powerful chips Apple has ever built* `https://www.apple.com/newsroom/2021/10/introducing-m1-pro-and-m1-max-the-most-powerful-chips-apple-has-ever-built/` (2022).
4. Arute, F. *et al.* Quantum supremacy using a programmable superconducting processor. *Nature* **574,** 505–510 (2019).
5. Zwerver, A. *et al.* Qubits made by advanced semiconductor manufacturing. *Nature Electronics* **5,** 184–190 (2022).
6. Lawrie, W. *et al.* Quantum dot arrays in silicon and germanium. *Applied Physics Letters* **116,** 080501 (2020).
7. Yang, C. H. *et al.* Operation of a silicon quantum processor unit cell above one kelvin. *Nature* **580,** 350–354 (2020).
8. Petit, L. *et al.* Universal quantum logic in hot silicon qubits. *Nature* **580,** 355–359 (2020).
9. Krinner, S. *et al.* Engineering cryogenic setups for 100-qubit scale superconducting circuit systems. *EPJ Quantum Technology* **6,** 2 (2019).
10. Joshi, S. & Moazeni, S. Scaling up Superconducting Quantum Computers with Cryogenic RF-photonics. *arXiv preprint arXiv:2210.15756* (2022).
11. Veldhorst, M., Eenink, H., Yang, C.-H. & Dzurak, A. S. Silicon CMOS architecture for a spin-based quantum computer. *Nature communications* **8,** 1–8 (2017).
12. Patra, B. *et al.* Cryo-CMOS circuits and systems for quantum computing applications. *IEEE Journal of Solid-State Circuits* **53,** 309–321 (2017).
13. Geck, L., Kruth, A., Bluhm, H., van Waasen, S. & Heinen, S. Control electronics for semiconductor spin qubits. *Quantum science and technology* **5,** 015004 (2019).
14. Xue, X. *et al.* CMOS-based cryogenic control of silicon quantum circuits. *Nature* **593,** 205–210 (2021).
15. Pauka, S. *et al.* A cryogenic CMOS chip for generating control signals for multiple qubits. *Nature Electronics* **4,** 64–70 (2021).
16. Stano, P. & Loss, D. Review of performance metrics of spin qubits in gated semiconducting nanostructures. *Nature Reviews Physics* **4,** 672–688 (2022).

17. Gerbert, P. & Ruess, F. *The Next Decade in Quantum Computing—and How to Play* `https://www.bcg.com/publications/2018/next-decade-quantum-computing-how-play` (2022).

18. Raussendorf, R. & Harrington, J. Fault-tolerant quantum computation with high threshold in two dimensions. *Physical review letters* **98,** 190504 (2007).

19. Wang, D. S., Fowler, A. G. & Hollenberg, L. C. Surface code quantum computing with error rates over 1%. *Physical Review A* **83,** 020302 (2011).

20. Xue, X. *et al.* Quantum logic with spin qubits crossing the surface code threshold. *Nature* **601,** 343–347 (2022).

21. Noiri, A. *et al.* Fast universal quantum gate above the fault-tolerance threshold in silicon. *Nature* **601,** 338–342 (2022).

22. Vandersypen, L. *et al.* Interfacing spin qubits in quantum dots and donors—hot, dense, and coherent. *npj Quantum Information* **3,** 1–10 (2017).

23. Borjans, F., Croot, X., Mi, X., Gullans, M. & Petta, J. Resonant microwave-mediated interactions between distant electron spins. *Nature* **577,** 195–198 (2020).

24. Stockklauser, A. *et al.* Strong coupling cavity QED with gate-defined double quantum dots enabled by a high impedance resonator. *Physical Review X* **7,** 011030 (2017).

25. Seidler, I. *et al.* Conveyor-mode single-electron shuttling in Si/SiGe for a scalable quantum computing architecture. *npj Quantum Information* **8,** 1–7 (2022).

26. Philips, S. G. *et al.* Universal control of a six-qubit quantum processor in silicon. *arXiv preprint arXiv:2202.09252* (2022).

27. Hendrickx, N. W. *et al.* A four-qubit germanium quantum processor. *Nature* **591,** 580–585 (2021).

28. Hendrickx, N. W. & Fuhrer, A. A spin qubit hiding from the noise. *Nature Nanotechnology* **17,** 1040–1041 (2022).

29. Lodari, M. *et al.* Low percolation density and charge noise with holes in germanium. *Materials for Quantum Technology* **1,** 011002 (2021).

30. Paquelet Wuetz, B. *et al.* Reducing charge noise in quantum dots by using thin silicon quantum wells. *Nature Communications* **14,** 1385 (2023).

31. Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Physical Review A* **57,** 120 (1998).

32. Elzerman, J. *et al.* Single-shot read-out of an individual electron spin in a quantum dot. *nature* **430,** 431–435 (2004).

33. Petta, J. R. *et al.* Coherent manipulation of coupled electron spins in semiconductor quantum dots. *Science* **309,** 2180–2184 (2005).

34. Itoh, K. M. & Watanabe, H. Isotope engineering of silicon and diamond for quantum computing and sensing applications. *MRS communications* **4,** 143–157 (2014).

35. Nadj-Perge, S., Frolov, S., Bakkers, E. & Kouwenhoven, L. P. Spin–orbit qubit in a semiconductor nanowire. *Nature* **468,** 1084–1087 (2010).

36. Hu, Y., Kuemmeth, F., Lieber, C. M. & Marcus, C. M. Hole spin relaxation in Ge–Si core–shell nanowire qubits. *Nature nanotechnology* **7,** 47–50 (2012).

37. Bosco, S., Hetényi, B. & Loss, D. Hole spin qubits in Si FinFETs with fully tunable spin-orbit coupling and sweet spots for charge noise. *PRX Quantum* **2,** 010348 (2021).

38. Ha, W. *et al.* A flexible design platform for Si/SiGe exchange-only qubits with low disorder. *Nano Letters* **22,** 1443–1448 (2021).

39. Losert, M. P. *et al.* Practical Strategies for Enhancing the Valley Splitting in Si/SiGe Quantum Wells. *arXiv preprint arXiv:2303.02499* (2023).

40. Paquelet Wuetz, B. *et al.* Atomic fluctuations lifting the energy degeneracy in Si/SiGe quantum dots. *Nature Communications* **13,** 7730 (2022).

41. Hanson, R. *et al.* Single-shot readout of electron spin states in a quantum dot using spin-dependent tunnel rates. *Physical review letters* **94,** 196802 (2005).

42. Hanson, R., Kouwenhoven, L. P., Petta, J. R., Tarucha, S. & Vandersypen, L. M. Spins in few-electron quantum dots. *Reviews of modern physics* **79,** 1217 (2007).

43. Fuhrer, A. & Fasth, C. *Coulomb blockade in Quantum Dots* Web. Lecture notes from FFFN35 – Physics of low-dimensional structures and quantum devices. Lund University. Apr. 2007.

44. Van der Zant, H. *Lecture 9 – Charging effects* Web. Lecture notes from AP3261 – Mesoscopic Physics. Delft University of Technology. Sept. 2021.

45. Drude, P. Zur elektronentheorie der metalle. *Annalen der physik* **306,** 566–613 (1900).

46. Tracy, L. *et al.* Observation of percolation-induced two-dimensional metal-insulator transition in a Si MOSFET. *Physical Review B* **79,** 235307 (2009).

47. Kim, J.-S., Tyryshkin, A. M. & Lyon, S. A. Annealing shallow Si/SiO2 interface traps in electron-beam irradiated high-mobility metal-oxide-silicon transistors. *Applied Physics Letters* **110,** 123505 (2017).

48. Li, P. & Zhang, Z. Standing wave effect and fractal structure in dislocation evolution. *Scientific reports* **7,** 4062 (2017).

49. Chen, H. *et al.* Crosshatching on a SiGe film grown on a Si (001) substrate studied by Raman mapping and atomic force microscopy. *Physical Review B* **65,** 233303 (2002).

50. Sawano, K. *et al.* Observation of strain field fluctuation in SiGe-relaxed buffer layers and its influence on overgrown structures. *Materials science in semiconductor processing* **8,** 177–180 (2005).

51. Sawano, K., Usami, N., Arimoto, K., Nakagawa, K. & Shiraki, Y. Strain field and related roughness formation in SiGe relaxed buffer layers. *Thin solid films* **508,** 117–119 (2006).

52. Sakaki, H., Noda, T., Hirakawa, K., Tanaka, M. & Matsusue, T. Interface roughness scattering in GaAs/AlAs quantum wells. *Applied physics letters* **51,** 1934–1936 (1987).

53. Ueno, T., Irisawa, T., Shiraki, Y., Uedono, A. & Tanigawa, S. Low temperature buffer growth for modulation doped SiGe/Ge/SiGe heterostructures with high hole mobility. *Thin Solid Films* **369,** 320–323 (2000).

54. Holỳ, V., Li, J., Bauer, G., Schäffler, F. & Herzog, H.-J. Diffuse x-ray scattering from misfit dislocations in SiGe epitaxial layers with graded Ge content. *Journal of applied physics* **78,** 5013–5021 (1995).

55. Chahine, G. A. *et al.* Imaging of strain and lattice orientation by quick scanning X-ray microscopy combined with three-dimensional reciprocal space mapping. *Journal of Applied Crystallography* **47,** 762–769 (2014).

56. Evans, P. *et al.* Nanoscale Distortions of Si Quantum Wells in Si/SiGe Quantum-Electronic Heterostructures. *Advanced Materials* **24,** 5217–5221 (2012).

57. Zoellner, M. H. *et al.* Imaging structure and composition homogeneity of 300 mm SiGe virtual substrates for advanced CMOS applications by scanning X-ray diffraction microscopy. *ACS applied materials & interfaces* **7,** 9031–9037 (2015).

58. Corley-Wiciak, C. *et al.* Nanoscale Mapping of the 3D Strain Tensor in a Germanium Quantum Well Hosting a Functional Spin Qubit Device. *ACS Applied Materials & Interfaces* (2023).

59. Borsoi, F. *et al.* Shared control of a 16 semiconductor quantum dot crossbar array. *arXiv preprint arXiv:2209.06609* (2022).

60. Deelman, P. W., Edge, L. F. & Jackson, C. A. Metamorphic materials for quantum computing. *MRS Bulletin* **41,** 224–230 (2016).

61. Sze, S. M., Li, Y. & Ng, K. K. *Physics of semiconductor devices* (John wiley & sons, 2021).

62. Antonova, I., Kag an, M., Polyakov, V., Golik, L. & Kolodzey, J. Effect of interface states on population of quantum wells in SiGe/Si structures. *physica status solidi (c)* **2,** 1924–1928 (2005).

63. Queisser, H. J. & Haller, E. E. Defects in semiconductors: some fatal, some vital. *Science* **281,** 945–950 (1998).

64. Hardy, W. J. *et al.* Gate-Defined Quantum Dots in Ge/SiGe Quantum Wells as a Platform for Spin Qubits. *ECS Transactions* **92,** 17 (2019).

65. Sardashti, K. *et al.* Sulfur passivation for the formation of Si-terminated Al2O3/SiGe (0 0 1) interfaces. *Applied Surface Science* **366,** 455–463 (2016).

66. Spruijtenburg, P. C., Amitonov, S. V., van der Wiel, W. G. & Zwanenburg, F. A. A fabrication guide for planar silicon quantum dot heterostructures. *Nanotechnology* **29,** 143001 (2018).

67. Burkard, G., Ladd, T. D., Nichol, J. M., Pan, A. & Petta, J. R. Semiconductor spin qubits. *arXiv preprint arXiv:2112.08863* (2021).

68. Khaetskii, A. V. & Nazarov, Y. V. Spin-flip transitions between Zeeman sublevels in semiconductor quantum dots. *Physical Review B* **64,** 125316 (2001).

69. Camenzind, L. C. *et al.* Hyperfine-phonon spin relaxation in a single-electron GaAs quantum dot. *Nature communications* **9,** 3454 (2018).

70. Laucht, A. *et al.* A dressed spin qubit in silicon. *Nature nanotechnology* **12,** 61–66 (2017).

71. Abragam, A. *The principles of nuclear magnetism* **32** (Oxford university press, 1961).

72. Erlingsson, S. I. & Nazarov, Y. V. Hyperfine-mediated transitions between a Zeeman split doublet in GaAs quantum dots: The role of the internal field. *Physical Review B* **66,** 155327 (2002).

73. Ferdous, R. *et al.* Interface-induced spin-orbit interaction in silicon quantum dots and prospects for scalability. *Physical Review B* **97,** 241401 (2018).

74. Machlup, S. Noise in semiconductors: spectrum of a two-parameter random signal. *Journal of Applied Physics* **25,** 341–343 (1954).

75. Hooge, F., Kleinpenning, T. & Vandamme, L. K. Experimental studies on 1/f noise. *Reports on progress in Physics* **44,** 479 (1981).

76. McWhorter, A. L. *et al.* 1/f noise and related surface effects in germanium. (1955).

77. Dutta, P. & Horn, P. Low-frequency fluctuations in solids: 1 f noise. *Reviews of Modern physics* **53,** 497 (1981).

78. Hooge, F. N. 1/f noise is no surface effect. *Physics letters A* **29,** 139–140 (1969).

79. Yoneda, J. *et al.* A quantum-dot spin qubit with coherence limited by charge noise and fidelity higher than 99.9%. *Nature nanotechnology* **13,** 102–106 (2018).

80. Struck, T. *et al.* Low-frequency spin qubit energy splitting noise in highly purified 28Si/SiGe. *npj Quantum Information* **6,** 40 (2020).

81. Kranz, L. *et al.* Exploiting a Single-Crystal Environment to Minimize the Charge Noise on Qubits in Silicon. *Advanced Materials* **32,** 2003361 (2020).

82. Dial, O. *et al.* Charge noise spectroscopy using coherent exchange oscillations in a singlet-triplet qubit. *Physical review letters* **110,** 146804 (2013).

83. Connors, E. J., Nelson, J., Edge, L. F. & Nichol, J. M. Charge-noise spectroscopy of Si/SiGe quantum dots via dynamically-decoupled exchange oscillations. *Nature communications* **13,** 940 (2022).

84. Connors, E. J., Nelson, J., Qiao, H., Edge, L. F. & Nichol, J. M. Low-frequency charge noise in Si/SiGe quantum dots. *Physical Review B* **100,** 165305 (2019).

85. Rojas-Arias, J. S. *et al.* Spatial noise correlations beyond nearest-neighbor in ^{28}Si/SiGe spin qubits. *arXiv preprint arXiv:2302.11717* (2023).

86. Boter, J. M. *et al.* Spatial noise correlations in a Si/SiGe two-qubit device from Bell state coherences. *Physical Review B* **101,** 235133 (2020).

87. Freeman, B. M., Schoenfield, J. S. & Jiang, H. Comparison of low frequency charge noise in identically patterned Si/SiO2 and Si/SiGe quantum dots. *Applied Physics Letters* **108,** 253108 (2016).

88. Degli Esposti, D. *et al.* Wafer-scale low-disorder 2DEG in 28Si/SiGe without an epitaxial Si cap. *Applied Physics Letters* **120,** 184003 (2022).

89. Bavdaz, P. *et al.* A quantum dot crossbar with sublinear scaling of interconnects at cryogenic temperature. *npj Quantum Information* **8,** 86 (2022).

90. Dimoulas, A., Tsipas, P., Sotiropoulos, A. & Evangelou, E. Fermi-level pinning and charge neutrality level in germanium. *Applied physics letters* **89,** 252110 (2006).

91. Sammak, A. *et al.* Shallow and undoped germanium quantum wells: a playground for spin and hybrid quantum technology. *Advanced Functional Materials* **29,** 1807613 (2019).

92. Tosato, A. *et al.* Hard superconducting gap in a high-mobility semiconductor. *arXiv preprint arXiv:2206.00569* (2022).

93. Uberuaga, B. P., Leskovar, M., Smith, A. P., Jónsson, H. & Olmstead, M. Diffusion of Ge below the Si (100) surface: Theory and experiment. *Physical review letters* **84,** 2441 (2000).

94. Kittl, J. *et al.* Silicides and germanides for nano-CMOS applications. *Materials Science and Engineering: B* **154,** 144–154 (2008).

95. Lee, R. T. P., Chi, D. Z. & Yeo, Y.-C. Platinum Germanosilicide as Source/Drain Contacts in P-Channel Fin Field-Effect Transistors (FinFETs). *IEEE transactions on electron devices* **56,** 1458–1465 (2009).

96. Demeurisse, C. *et al.* Thermal stability of NiPt-and Pt-silicide contacts on SiGe source/drain. *Microelectronic Engineering* **84,** 2547–2551 (2007).

97. Gambino, J. & Colgan, E. Silicides and ohmic contacts. *Materials chemistry and physics* **52,** 99–146 (1998).

98. Yu, H. *et al.* TiSi (Ge) Contacts Formed at Low Temperature Achieving Around 2×10-9 Ohm.cm2 Contact Resistivities to p-SiGe. *IEEE Transactions on Electron Devices* **64,** 500–506 (2017).

99. Yu, H. *et al. Titanium (germano-) silicides featuring 10- 9 Ω· cm 2 contact resistivity and improved compatibility to advanced CMOS technology* in *2018 18th International Workshop on Junction Technology (IWJT)* (2018), 1–5.

100. Gabrielli, L. H. *Python Package Index - gdspy* Revision 9d1917ec. `https://gdspy.readthedocs.io/en/stable/` (2022).

101. Brauns, M., Amitonov, S. V., Spruijtenburg, P.-C. & Zwanenburg, F. A. Palladium gates for reproducible quantum dots in silicon. *Scientific reports* **8,** 1–8 (2018).

102. *E-Beam Resist AR-P 6200 series (CSAR 62)* AR-P 6200.04,AR-P 6200.09,AR-P 6200.13, AR-P 6200.18. AllResist (July 2019).

103. *SCANASYST-AIR-HPI* AR-P 6200.04,AR-P 6200.09,AR-P 6200.13, AR-P 6200.18. Bruker (Apr. 2023). `https://www.brukerafmprobes.com/p-3967-scanasyst-air-hpi.aspx`.

104. Sawano, K., Koh, S., Shiraki, Y., Usami, N. & Nakagawa, K. In-plane strain fluctuation in strained-Si/SiGe heterostructures. *Applied physics letters* **83,** 4339–4341 (2003).

105. Attocube systems AG. *attoDRY, Closed-cycle Cryostats* Web. Brochure. 2021.

106. Attocube systems AG. *Features – Basic Measurement Insert option for attoDRY1000/2100* `https://www.attocube.com/en/products/cryostats/features/basic-measurement-insert` (2023).

107. *QT designed instrumentation* `https://qtwork.tudelft.nl/~schouten/ivvi/doc-mod/index-mod.htm` (2023).

108. *SPI Rack Modules* `https://qtwork.tudelft.nl/~mtiggelman/modules.html` (2023).

109. *QCoDeS* 0.38.0.dev540+master.g4d1dacb3a. `https://qcodes.github.io/Qcodes/` (2022).

110. Shklovskii, B. I. & Efros, A. L. *Electronic properties of doped semiconductors* (Springer Science & Business Media, 2013).

111. Mi, X. *et al.* Magnetotransport studies of mobility limiting mechanisms in undoped Si/SiGe heterostructures. *Physical Review B* **92,** 035304 (2015).

112. Fogelholm, R. The conductivity of large percolation network samples. *Journal of Physics C: Solid State Physics* **13,** L571 (1980).

113. Brar, B. *et al. Impact ionization in high performance AlGaN/GaN HEMTs* in *Proceedings. IEEE Lester Eastman Conference on High Performance Devices* (2002), 487–491.

114. Richter, A., Glunz, S. W., Werner, F., Schmidt, J. & Cuevas, A. Improved quantitative description of Auger recombination in crystalline silicon. *Physical review B* **86,** 165202 (2012).

115. Ielmini, D., Spinelli, A. S. & Lacaita, A. L. Experimental evidence for recombination-assisted leakage in thin oxides. *Applied Physics Letters* **76,** 1719–1721 (2000).

116. Chen, J., Chan, T., Chen, I., Ko, P. & Hu, C. Subbreakdown drain leakage current in MOSFET. *IEEE Electron Device Letters* **8,** 515–517 (1987).

117. James, A. M. & Lord, M. P. *Macmillan's chemical and physical data* (Macmillan, 1992).

118. Oehme, M. *et al.* Two-dimensional hole gases in SiGeSn alloys. *Semiconductor Science and Technology* **37,** 055009 (2022).

119. Groner, M., Elam, J. & Fabreguette, F. Electrical characterization of thin Al2O3 films grown by atomic layer deposition on silicon and various metal substrates. *Thin solid films* **413,** 186–197 (2002).

120. Su, Y.-H., Chuang, Y., Liu, C.-Y., Li, J.-Y. & Lu, T.-M. Effects of surface tunneling of two-dimensional hole gases in undoped Ge/GeSi heterostructures. *Physical Review Materials* **1,** 044601 (2017).

121. Laroche, D. *et al.* Scattering mechanisms in shallow undoped Si/SiGe quantum wells. *AIP Advances* **5,** 107106 (2015).

122. Shockley, W. & Read Jr, W. Statistics of the recombinations of holes and electrons. *Physical review* **87,** 835 (1952).

123. Schwartz, P. & Sturm, J. Microsecond carrier lifetimes in strained silicon-germanium alloys grown by rapid thermal chemical vapor deposition. *Applied physics letters* **57,** 2004–2006 (1990).

124. Meyer, M. *et al.* Electrical control of uniformity in quantum dot devices. *arXiv preprint arXiv:2211.13493* (2022).

125. Han, J.-H., Takenaka, M. & Takagi, S. Analysis of interface trap density of plasma post-nitrided Al2O3/SiGe MOS interface with high Ge content using high-temperature conductance method. *Journal of Applied Physics* **120,** 125707 (2016).

126. Shehata, M. *et al.* Modelling semiconductor spin qubits and their charge noise environment for quantum gate fidelity estimation. *arXiv preprint arXiv:2210.04539* (2022).

127. Nicollian, E., Berglund, C., Schmidt, P. & Andrews, J. Electrochemical charging of thermal SiO2 films by injected electron currents. *Journal of Applied Physics* **42,** 5654–5664 (1971).

128. Chen, G. & Xu, Z. Charge trapping and detrapping in polymeric materials. *Journal of Applied Physics* **106,** 123707 (2009).

129. Schulz, M. & Pensl, G. *Insulating Films on Semiconductors: Proceedings of the Second International Conference, INFOS 81, Erlangen, Fed. Rep. of Germany, April 27–29, 1981* (Springer Science & Business Media, 2013).

130. Williams, C. Kinetics of trapping, detrapping, and trap generation. *Journal of electronic materials* **21,** 711–720 (1992).

131. Wolters, D. & Van Der Schoot, J. Kinetics of charge trapping in dielectrics. *Journal of applied physics* **58,** 831–837 (1985).

132. Zwanenburg, F. A. *et al.* Silicon quantum electronics. *Reviews of modern physics* **85,** 961 (2013).

133. Brammertz, G. *et al.* Characteristic trapping lifetime and capacitance-voltage measurements of GaAs metal-oxide-semiconductor structures. *Applied physics letters* **91,** 133510 (2007).

134. DiMaria, D. & Cartier, E. Mechanism for stress-induced leakage currents in thin silicon dioxide films. *Journal of Applied physics* **78,** 3883–3894 (1995).

135. Lu, B., Lv, H., Zhang, Y., Zhang, Y. & Liu, C. Comparison of HfAlO, HfO2/Al2O3, and HfO2 on n-type GaAs using atomic layer deposition. *Superlattices and Microstructures* **99,** 54–57 (2016).

136. Giovane, L. M., Luan, H.-C., Agarwal, A. M. & Kimerling, L. C. Correlation between leakage current density and threading dislocation density in SiGe pin diodes grown on relaxed graded buffer layers. *Applied Physics Letters* **78,** 541–543 (2001).

A

DEVICE DESIGNS

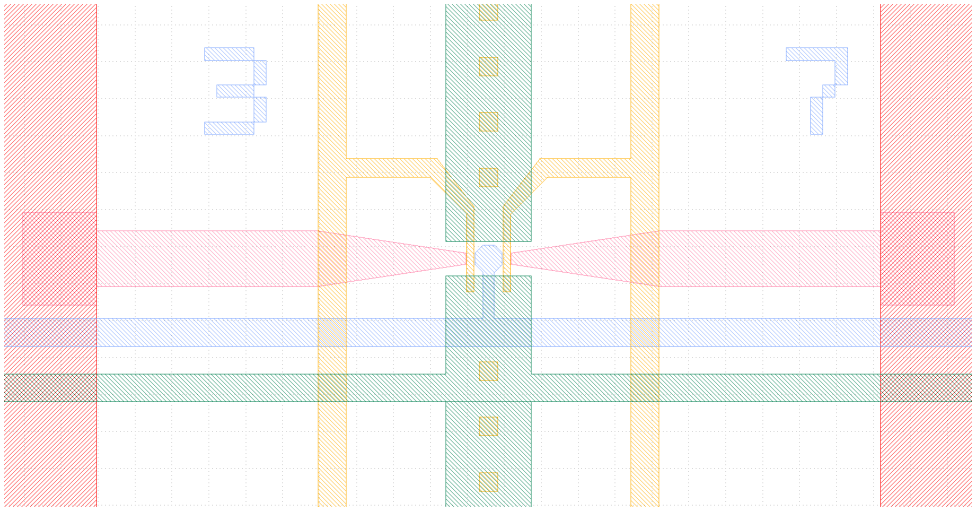
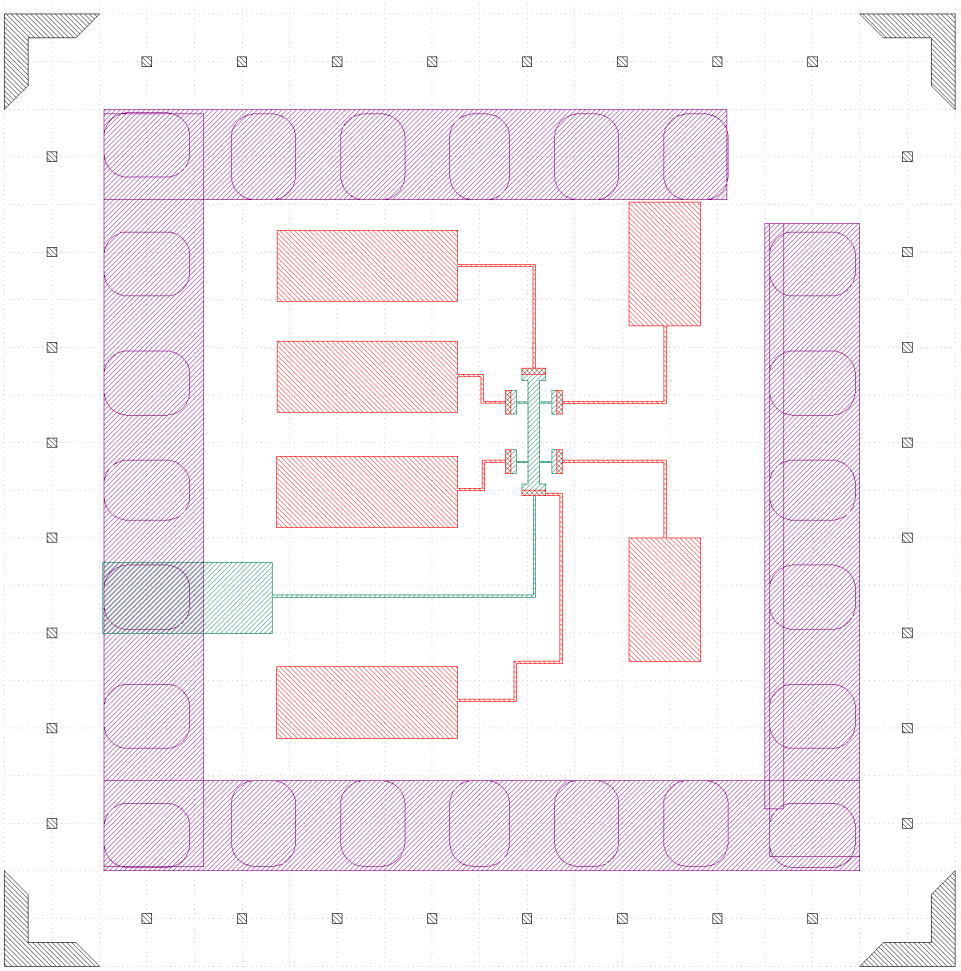


Figure A.1: Original SHT repeat unit design whose dimensions are provided in Fig. 3.1.



.

Figure A.2: Hall bar design, with top a top gate in green, ohmic leads in red, SiN in purple and markers in black. The HB length and width is 243 µm and 25 µm. The separation between longitudinal and transverse ohmic contacts is 125 µm and 95 µm

B

FABRICATION MICROGRAPHS

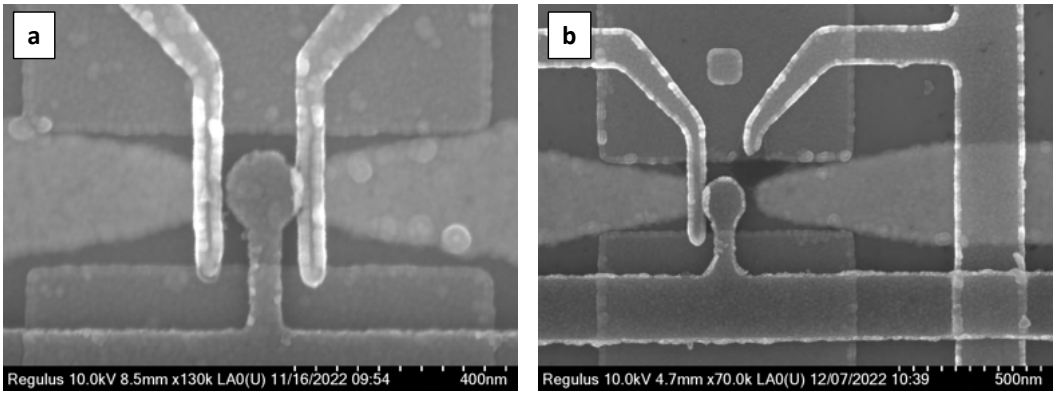


Figure B.1: SEM images after plunger lift-off during **a)** fabrication run 2 and **b)** fabrication run 3.

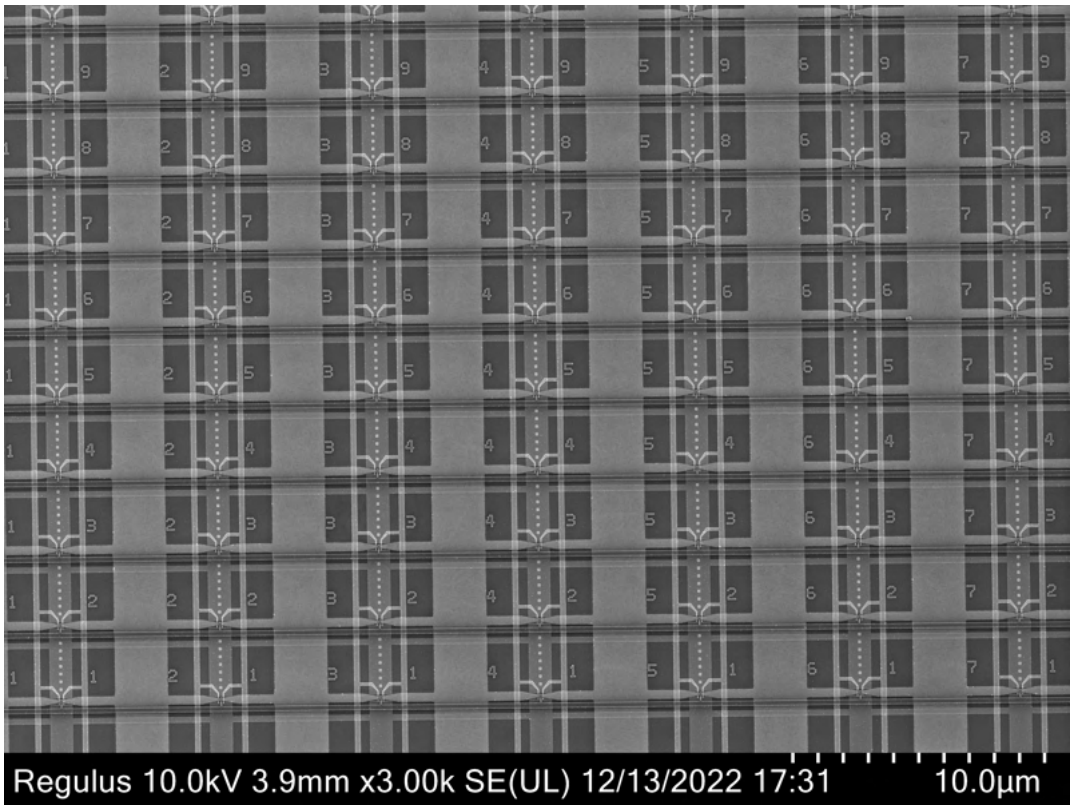


Figure B.2: SEM image 1/4 used in the dimensional characterization of plunger and barrier width

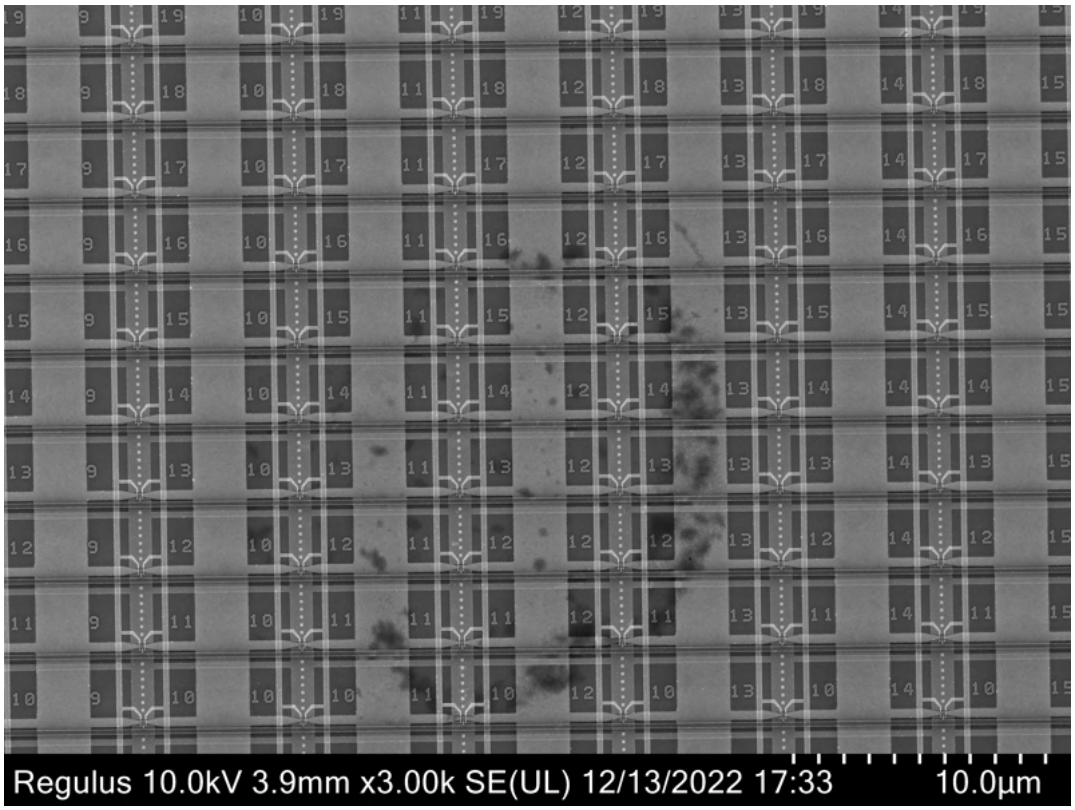


Figure B.3: SEM image 2/4 used in the dimensional characterization of plunger and barrier width

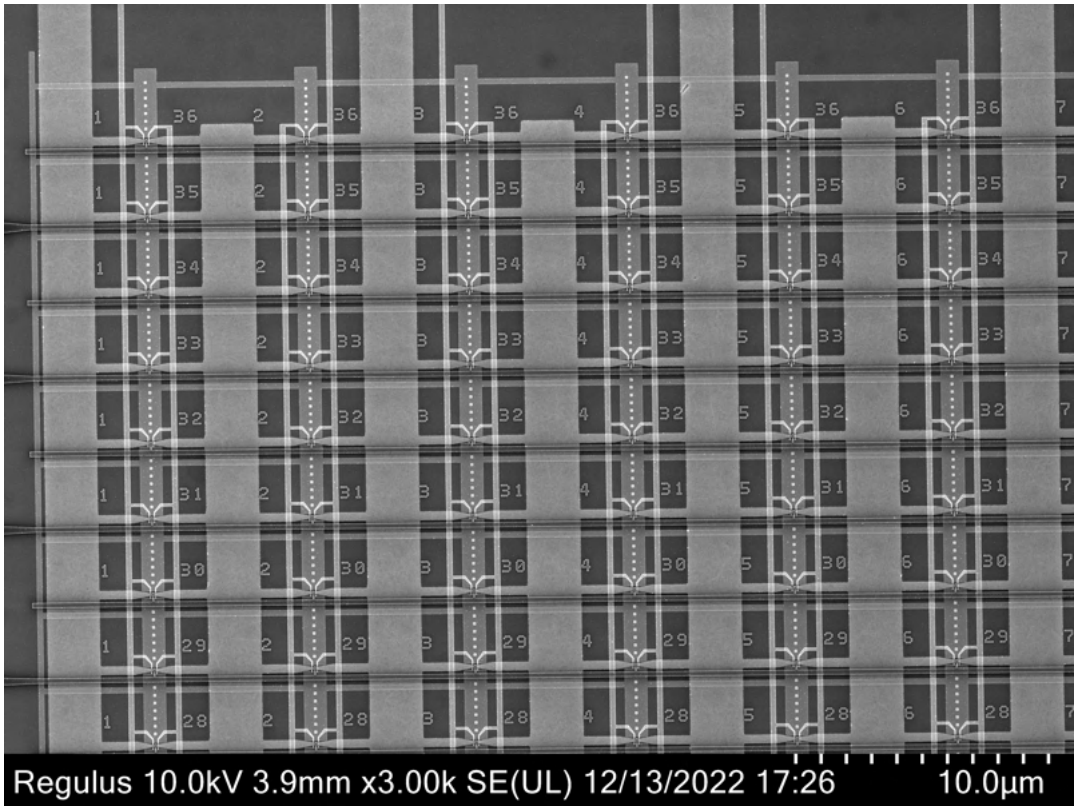


Figure B.4: SEM image 3/4 used in the dimensional characterization of plunger and barrier width

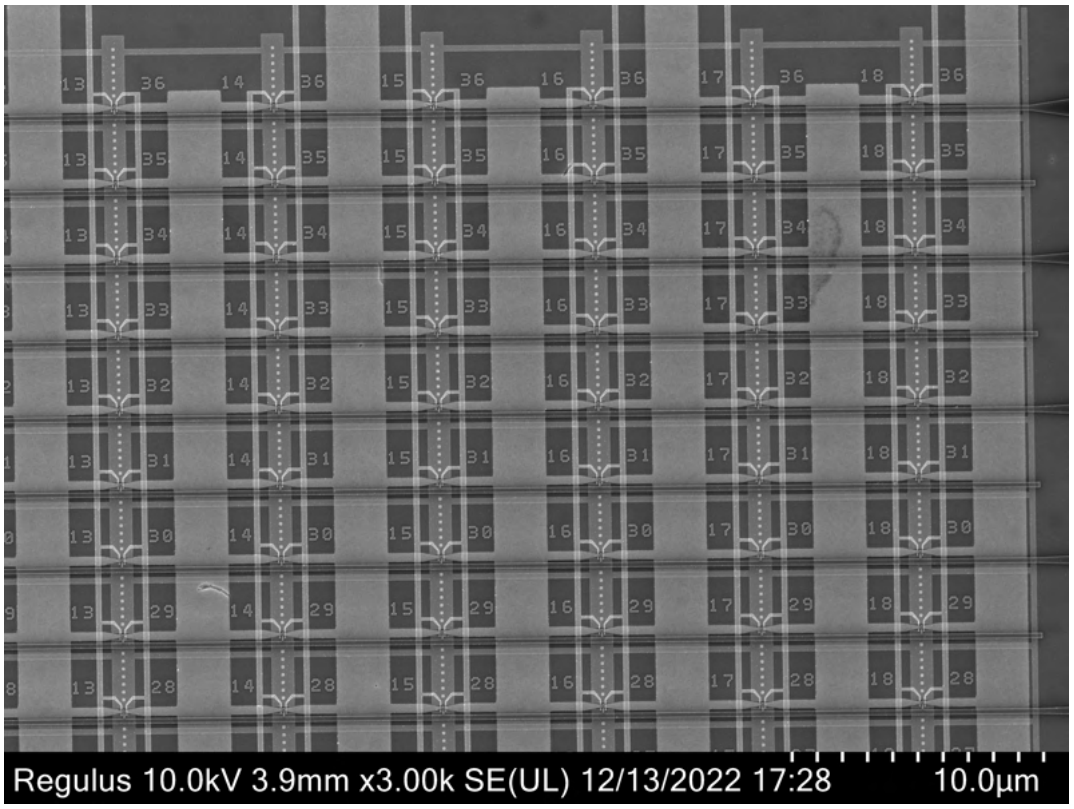


Figure B.5: SEM image 4/4 used in the dimensional characterization of plunger and barrier width

C

SUPPLEMENTARY FIGURES

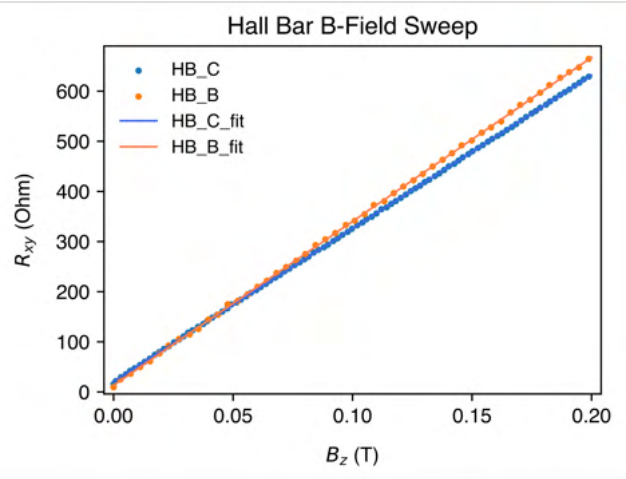


Figure C.1: Transverse resistance with perpendicular magnetic field measured in hall bars at $V_G = -0.45\,\mathrm{V}$. The linear fit is used to extract the carrier density using Eq. 2.21.

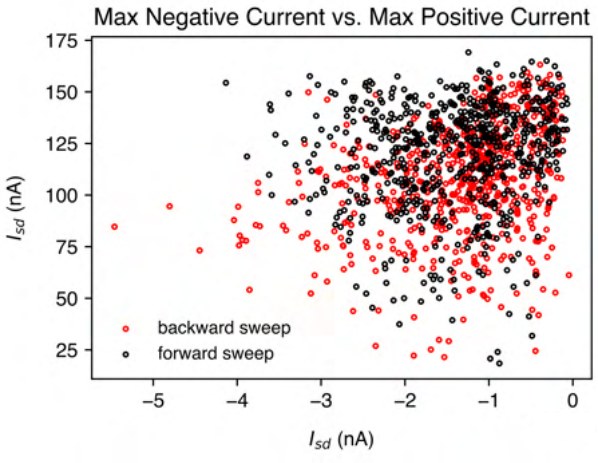


Figure C.2: Correlation between maximum measured positive and neagtive currents over forward and backwards sweeps of the grid.

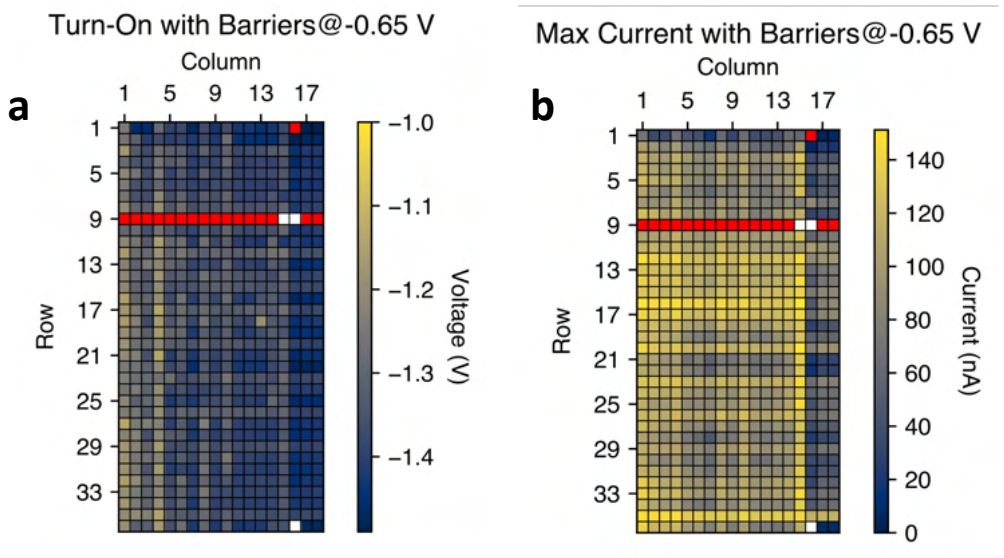


Figure C.3: **a)** Turn-on threshold voltages and **b)** maximum currents measured over the grid. Bias conditions were: $V_{screen} = -500\,\mathrm{mV}$, $V_{barriers} = -650\,\mathrm{mV}$, $V_G = -1.0\,\mathrm{V} \rightarrow -1.5\,\mathrm{V}$, $V_{sd} = 0.5\,\mathrm{mV}$. Owing to the screening gate bias, less leakage is observed than in Fig. 5.5 at similar barrier voltages.

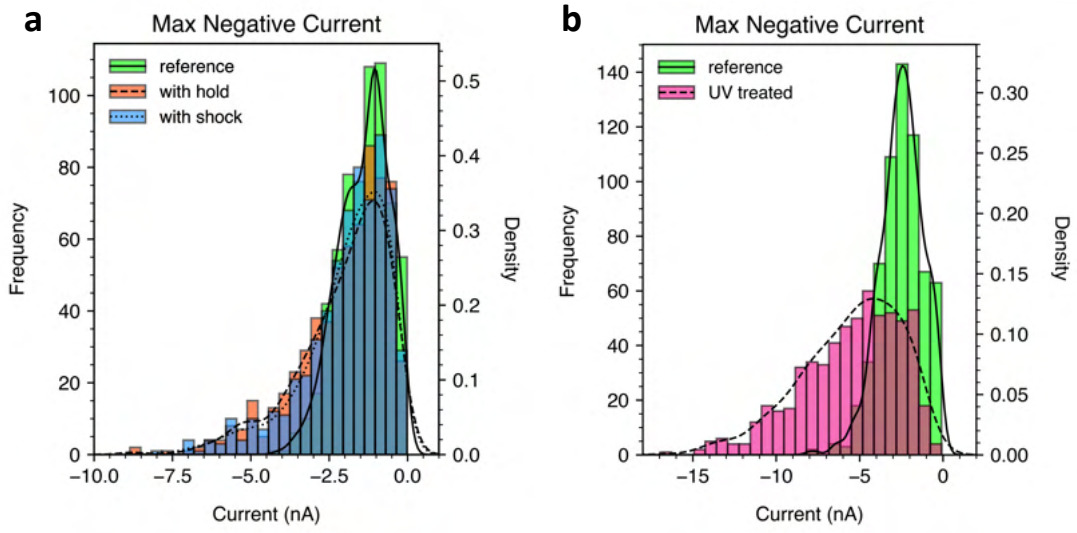


Figure C.4: Comparison of maximum currents measured over the grid **a)** using 'shock' and 'hold' scheme, and **b)** following UV treatment. Bias conditions were: $V_{screen} = -500\,\mathrm{mV}$, $V_{barriers} = -650\,\mathrm{mV}$, $V_G = -1.0\,\mathrm{V} \rightarrow -1.5\,\mathrm{V}$, $V_{sd} = 0.5\,\mathrm{mV}$. Owing to the screening gate bias, less leakage is observed than in Fig. 5.5 at similar barrier voltages.

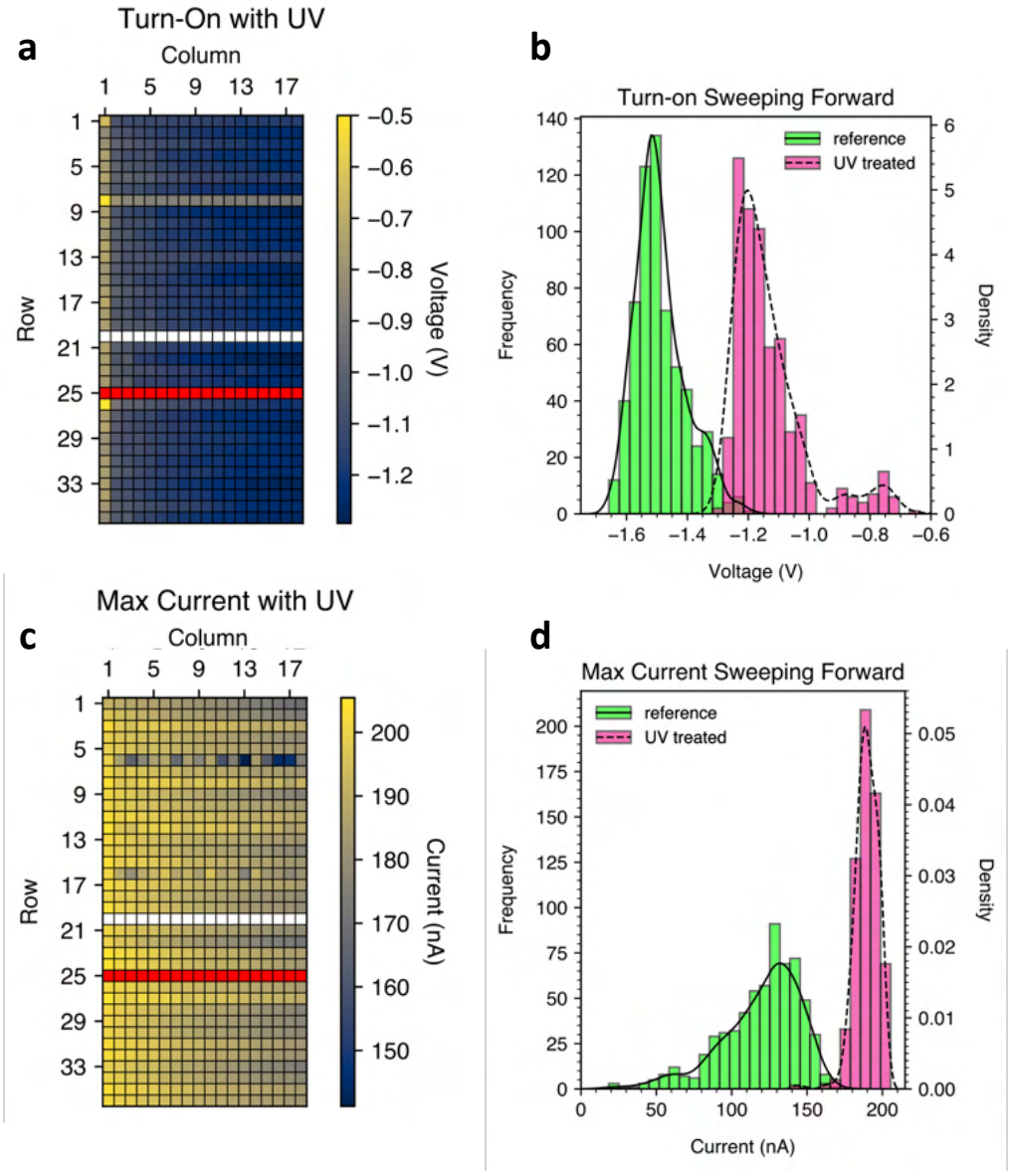


Figure C.5: Transport through devices measured in a grid following 30 minutes of UV treatment: **a)** turn-on threshold voltages, **b)** histogram comparing thresholds with the reference grid, **c)** maximum source-drain currents, **d)** histogram comparing maximum currents with the reference grid. In the reference grid, bias conditions were: $V_{screen} = -500\,\text{mV}$, $V_{barriers} = -500\,\text{mV}$, $V_G = -1.0\,\text{V} \rightarrow -1.7\,\text{V}$. In the UV grid, bias conditions were: $V_{screen} = -200\,\text{mV}$, $V_{barriers} = -400\,\text{mV}$, $V_G = -0.5\,\text{mV} \rightarrow -1.4\,\text{V}$. The S/D bias is always $V_{sd} = 0.5\,\text{mV}$. The white rows in a) and b) likely correspond to a broken wire bond on the plunger line.

Due to biasing conditions being different between reference and UV samples, a definitive comparison cannot be made between grids. Nonetheless, we point out the sharp distribution of maximum current, reminiscent of those found under 'shock' and 'hold' schemes, although no such scheme was used here. Moreover, row-wise decay in turn-on thresholds appears to progress much more rapidly. This suggests UV treatment may have introduced more defects, which leads to an initially faster population of traps. The collection of devices around $-0.8\,\text{V}$ in Fig. C.5b are those from the beginning of each row measurement. These are least affected by the filled trap population.

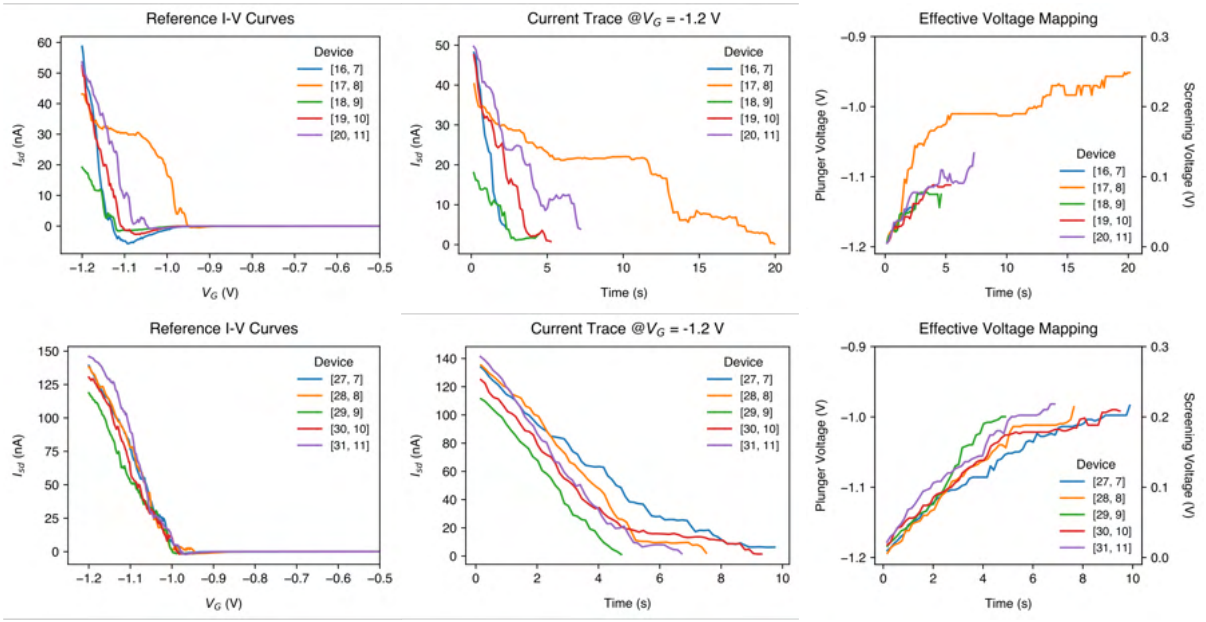


Figure C.6: From left to right: reference I-V curves, current trace, voltage mapping. The top row is of the reference grid and the bottom is of the UV treated grid.

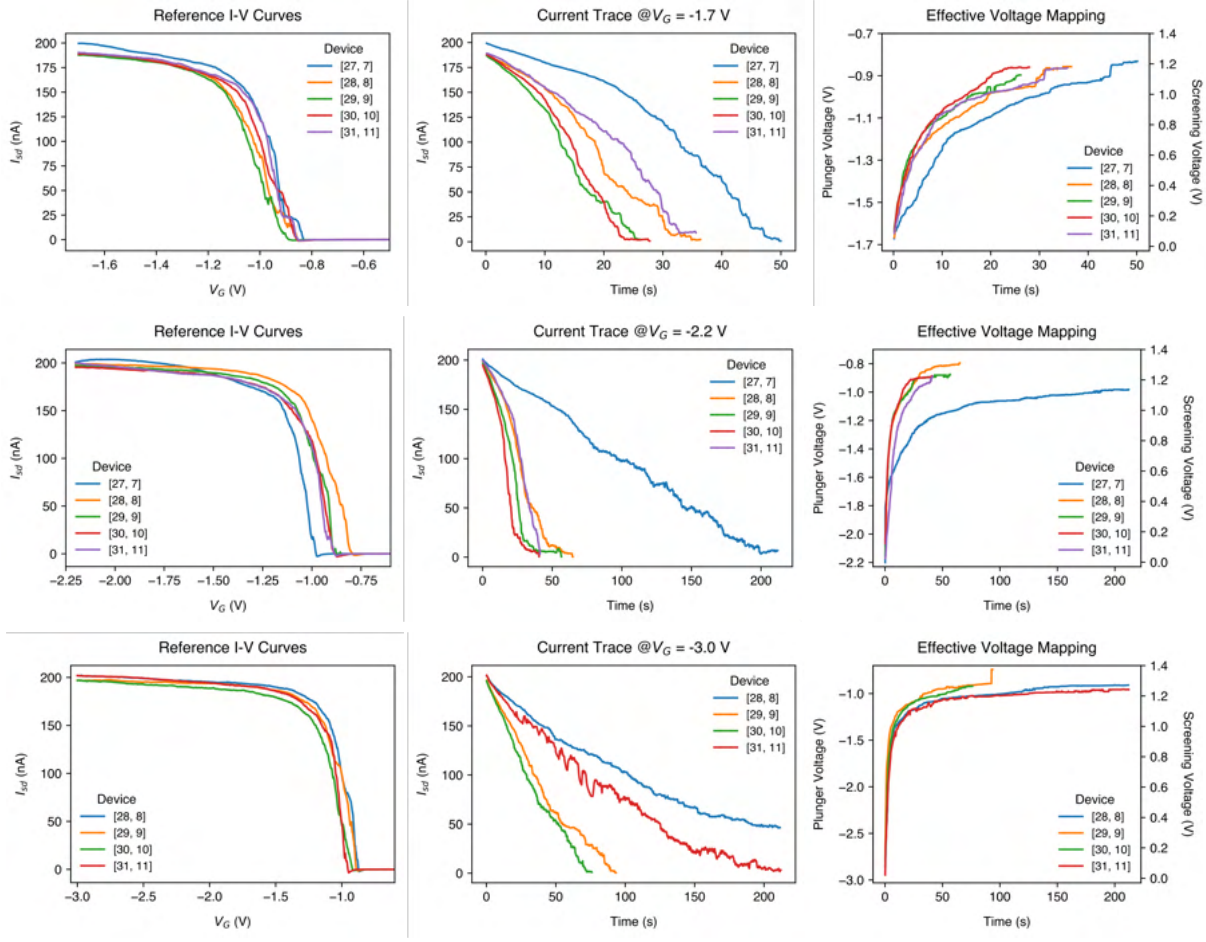


Figure C.7: From left to right: reference I-V curves, current trace, voltage mapping. Going down: holding voltages at $-1.7\,\mathrm{V}$, $-2.2\,\mathrm{V}$ and $-3.0\,\mathrm{V}$ after sweeping from $0\,\mathrm{V}$.

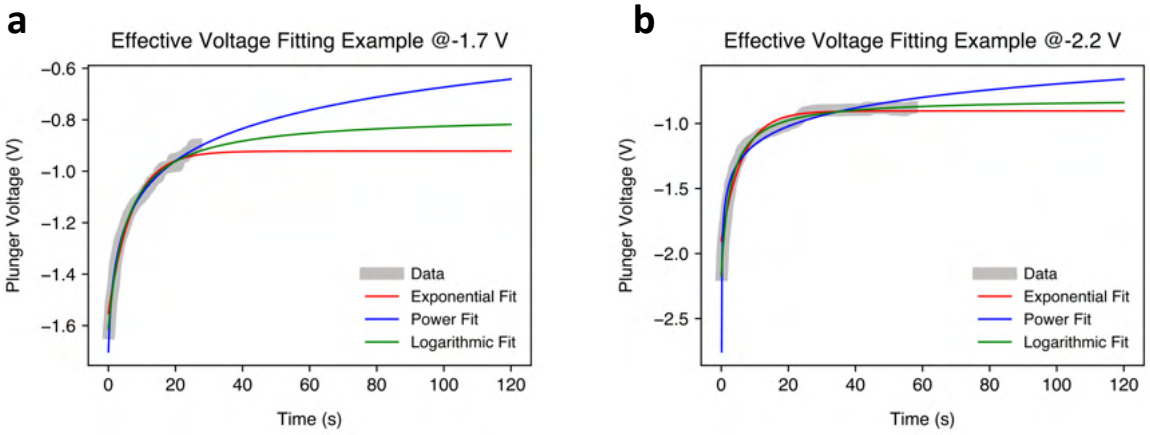


Figure C.8: Examples of exponential, power-law and logarithmic fits to effective voltage mappings using plunger holding voltages of **a)** $-1.7\,\mathrm{V}$ and **b)** $-2.2\,\mathrm{V}$

The function used for the exponential fit is:

$$y = -a\exp(-b(x+c)) + d \tag{C.1}$$

The function used for the power fit is:

$$y = -a(x+b)^c + d \tag{C.2}$$

The exponential fit usually struggles to capture the correct envelope of the voltage mapping, while the power fit decays unreasonably slowly, extends steeply near $t = 0$, and also yielded great variability between fitted measurements. Moreover, these functions require 4 fitting parameters, whereas the logarithmic function of Eq. 5.18 requires 3.

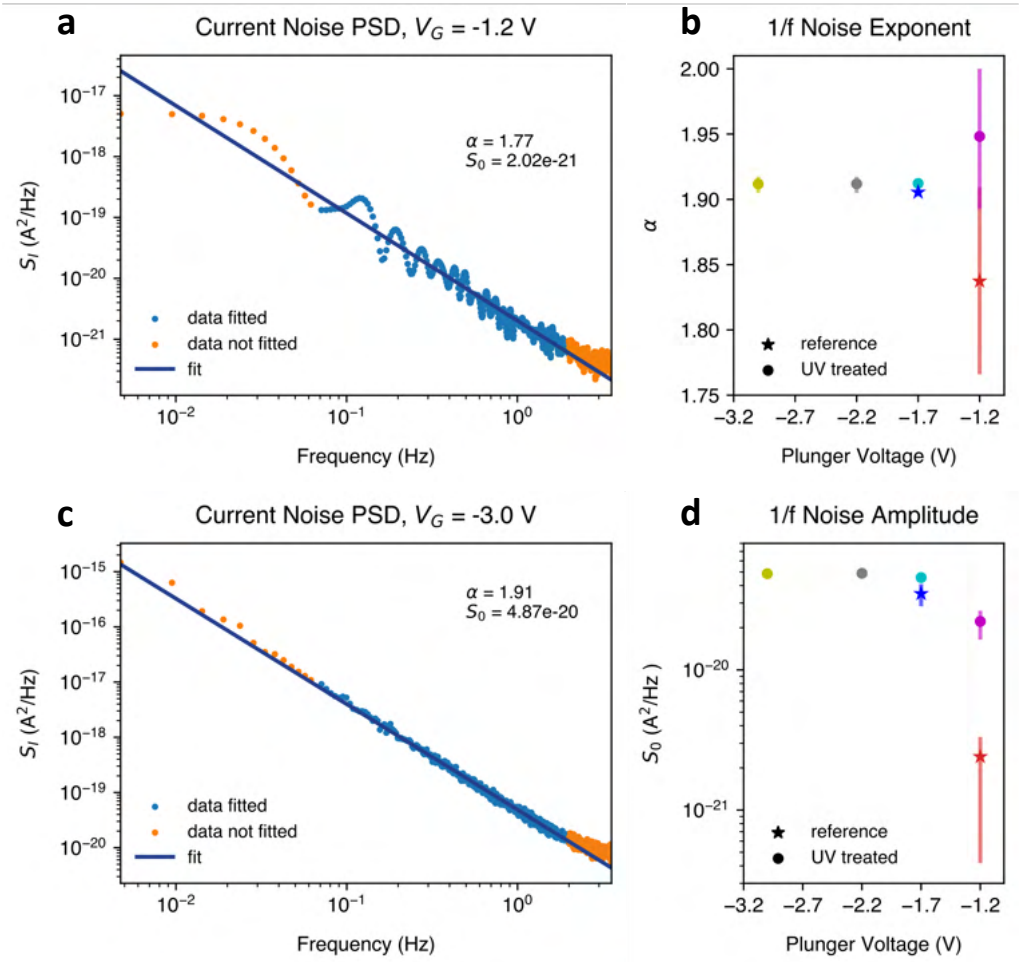


Figure C.9: Example current noise power spectra acquired from current traces when holding the plunger at **a)** $V_G = -1.2\,\mathrm{V}$ and **c)** $-3.0\,\mathrm{V}$. Fittings noise parameters using the Hooge model (Eq. 2.37) for **a)** noise exponent α and **b)** the noise amplitude S_0 for the current traces collected in Figs. C.6 & C.7. Points are the averages of 5 measurements and error bars mark upper and lower bounds of the data. Fittings were carried out used a least-squares algorithm. The lowest frequency pint is limited by the data collection period and the highest frequency point is the Nyquist frequency, which is half of the sampling rate over the current trace.

While we were unable to measure charge noise, we found it informative to collect the low-frequency noise power spectra of the current through SHTs held at different plunger voltages. We note that since there is no way to confidently subtract the noise background without tuning devices into Coulomb blockade, this data also includes the noise coming from our cryogenic set-up, the current amplifier and room temperature electronics.

Despite this, we found interesting behaviour, where the noise at low voltages showed oscillations, which were mostly absent at higher plunger biases. Since there are fewer traps activated at lower voltage, we suspect that oscillation peaks may correspond to the characteristic switching frequencies of individual TLFs. The large spread in the frequency exponents at this voltage reflects a non-uniform distribution of active TLFs, as expected at this low voltage. In line with conclusions in the main text, we also find that the noise amplitude in UV-treated samples is higher, which could indicate the presence of more traps.

D

SRH MODEL

Trap filling is modeled using Fermi-Dirac statistics. To this end, a parameter f_{pt} is introduced to represent the probability that a trap state is empty (filled by a hole), and thus in a position to capture an electron.

$$f_{pt} = 1 - f_t = \exp\{(E_t - F_n)/k_B T\} \tag{D.1}$$

Here, f_t is the probability that the trap is already filled by an electron, E_t is the trap energy and F_n is the quasi Fermi-level of electrons. Using this, a net rate of electron capture is constructed that depends on: the number of trapping centers per unit volume N_t, the number of electron states $N(E)dE$ in the energy range dE, the fraction of occupied trap states $f(E)$, and the probability per unit time that an electron in the range dE is captured by an empty trap $c_n(E)$,

$$dR = f_p N_t c_n(E) f(E) N(E) dE. \tag{D.2}$$

The net rate of capture is found by integrating this equation over the energy interval dE between the conduction band edge E_c and infinity, then subtracting the contribution of electron emission from the traps. The electron emission rate takes the same form as Eq. D.2, but with $f_{pt} \to f_t$, $f(E) \to f_p(E)$, and $c_n(E) \to e_n(E)$. The result for the net rate of electron capture is then,

$$R_{cn} = \left(1 - \exp\{(F_t - F_n)/k_B T\}\right) f_{pt} n C_n, \tag{D.3}$$

In the above, F_t is the quasi-Fermi-level of the traps, C_n is the probability per unit time that an electron in the conduction band will be captured for the case that all traps are empty, and n is the density of electrons in the conduction band, given by

$$n = N_c \exp\{(F_n - E_c)/k_B T\}. \tag{D.4}$$

N_c is the effective density of states at the conduction band edge. n_1 is defined as the electron density at the conduction band if the Fermi level were to fall at the trap level E_t,

$$n_1 = N_c \exp\{(E_t - E_c)/k_B T\}. \tag{D.5}$$

By substituting Eq. D.4 into Eq. D.3, we acquire the net rate of electron capture by traps.

$$R_{cn} = C_n f_{pt} n - C_n f_t n_1. \tag{D.6}$$

Through an identical treatment, one may arrive at an expression for the net rate of electron

emission from traps, which is equivalent to the net rate of hole capture by traps, given by,

$$R_{cp} = C_t f_t p - C_t f_{pt} p_1. \tag{D.7}$$

The corresponding hole terms are,

$$p = N_v \exp\{(E_v - F_n)/k_B T\}, \tag{D.8}$$
$$p_1 = N_v \exp\{(E_v - E_t)/k_B T\}, \tag{D.9}$$

where N_v is the effective density of hole states at the valence band edge E_v.

In the steady state, the rate of electron capture must equal the rate of hole capture. Under these conditions, there is a fixed rate of trap-assisted electron and hole recombination, R. To find R, R_{cn} and R_{cp} are equated to pull out expressions for f_t and f_{pt}

$$f_t = \frac{C_n n C_p p_1}{C_n(n + n_1) + C_p(p + p_1)} \tag{D.10}$$
$$f_{pt} = \frac{C_n n_1 C_p p}{C_n(n + n_1) + C_p(p + p_1)} \tag{D.11}$$

Substituting these quantities into either of Eq. D.3 or Eq. D.7, yields

$$R = \frac{C_n C_p (np - n_1 p_1)}{C_n(n + n_1) + C_p(p + p_1)}, \tag{D.12}$$

where it turns out that $n_1 p_1$ is independent of the trap energy level and equivalent to the intrinsic carrier concentration n_i^2.

$$\begin{aligned} n_1 p_1 &= N_c N_v \exp\{(E_v - E_c) k_B T\} \\ &= N_c N_v \exp\{(-E_g) k_B T\} \\ &= n_i^2 \end{aligned} \tag{D.13}$$

This yields Eq. 5.9 shown in the main text.

The probabilities C_n and C_p are directly related to the carrier velocities and trapping cross-sections, which in turn depend on carrier kinetics and trap chemistry. At higher electron velocities, the relaxation time is smaller, and the likelihood of an electron being scattered toward a trap per unit time is greater. Shallow traps will capture charges faster due to their lower activation energy requirement, however, traps that occupy larger volumes, which are typically deeper, have an enhanced Coulomb interaction.