Scalable and high-sensitivity readout of silicon quantum devices

Simon Schaal

A thesis submitted for the degree of
Doctor of Philosophy

University College London
Department of Electronic & Electrical Engineering
London, 30th January 2020
I, Simon Schaal, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

London, 30th January 2020

Simon Schaal
Abstract

Scalable and high-sensitivity readout of silicon quantum devices

Simon Schaal

Quantum computing is predicted to provide unprecedented enhancements in computational power. A quantum computer requires implementation of a well-defined and controlled quantum system of many interconnected qubits, each defined using fragile quantum states. The interest in a spin-based quantum computer in silicon stems from demonstrations of very long spin-coherence times, high-fidelity single spin control and compatibility with industrial mass-fabrication. Industrial scale fabrication of the silicon platform offers a clear route towards a large-scale quantum computer, however, some of the processes and techniques employed in qubit demonstrators are incompatible with a dense and foundry-fabricated architecture. In particular, spin-readout utilises external sensors that require nearly the same footprint as qubit devices.

In this thesis, improved readout techniques for silicon quantum devices are presented and routes towards implementation of a scalable and high-sensitivity readout architecture are investigated. Firstly, readout sensitivity of compact gate-based sensors is improved using a high-quality factor resonator and Josephson parametric amplifier that are fabricated separately from quantum dots. Secondly, an integrated transistor-based control circuit is presented using which sequential readout of two quantum dot devices using the same gate-based sensor is achieved. Finally, a large-scale readout architecture based on random-access and frequency multiplexing is introduced. The impact of readout circuit footprint on readout sensitivity is determined, showing routes towards integration of conventional circuits with quantum devices in a dense architecture, and a fault-tolerant architecture based on mediated exchange is introduced, capable of relaxing the limitations on available control circuit footprint per qubit. Demonstrations are based on foundry-fabricated transistors and few-electron quantum dots, showing that industry fabrication is a viable route towards quantum computation at a scale large enough to begin addressing the most challenging computational problems.
Impact Statement

Large scale quantum computing could have wide-ranging impacts on modern society. By revolutionising the way computations are performed, the computationally most challenging problems across academia and industry could be solved, from the modelling of biochemical and physical processes or the processing of vast amounts of data in the application of artificial intelligence, to the discovery of new drugs and materials.

The results demonstrated and techniques developed in this thesis bring an implementation of a large-scale quantum computer based on silicon technology one step closer to reality by demonstrating routes to a scalable and sensitive readout architecture. However, further developments, that might continue over the next decade or longer, are required to build a useful silicon-based quantum computer.

The highly-sensitive capacitance and charge sensor demonstrated in this thesis could find immediate application in other fields, where cryogenic sensing is necessary. Moreover, characterisation and modelling of foundry-fabricated transistors at cryogenic temperatures provides useful information for developing models towards cryogenic circuit design, with applications not only in the cryogenic quantum computing community, but also in deep space exploration. By adapting Josephson parametric amplifiers, developed in the field of superconducting circuits, to read-out of silicon-based devices, this thesis could additionally motivate further multidisciplinary work where Josephson parametric amplification becomes an important ingredient.
Acknowledgements

This work would not have been possible without the help and support of many great people and I would like to thank anyone who has accompanied me on my PhD journey. In the following, I highlight a few people I would like to thank in particular:

First of all, John Morton who has been a very supportive and knowledgable supervisor. While having clear research goals, I enjoyed the experimental freedom when working with John and value his scientific intuition and his remarkable talent in engaging and maintaining collaborations. I have learned a lot not only in scientific discussions with him but also by meeting many interesting people through him at conferences, meetings or dinners. Many activities outside the lab have made the last years a great experience, even though the group skiing trip that John promised at the beginning never took place.

Secondly, my industrial supervisor Fernando Gonzalez-Zalba, for inviting me to Cambridge and for continuous support throughout my whole PhD. Fernando is an exceptional and calm experimentalist and his skills in planning an experiment and deep understanding of reflectometry techniques have made my short stays in Cambridge a great success.

Previous QSD group members for their warm welcome to the group. Anasua Chatterjee and John Mansir, who have helped me a lot in getting started with experiments in the lab. Salahuddin Nur who always ensured that everyone was joining the pub call. Naitik Panjwani who has worked on a refreshingly different subject to everyone else in the group. Hee-Jin Lim and Leonid Abdurakimov for being very knowledgable in any optics or microwave related questions. Philipp Ross who kindly left me his LCN desk in exchange for a desk in the physics basement.

Current QSD members who have made all the time spent in the lab great fun. Gavin Dold and his excellent knowledge of English beer. Pierandrea Conti for being an excellent deskmate and who likes the Gordon’s cafe pizza more than an Italian should do. James who always had crazy stories to tell after another mountaineering trip and who I had a great time with exploring California. David Wise who is always up for a chat about cricket but I have no idea how it works. Christoph Zollitsch
who has been extremely helpful in any lab related questions especially when things stopped working. Together we have been through an intense but fun period and have gained lots of experience in heavy lifting (and how to avoid it) during lab refurbishment.

New QSD members and the new PhD cohort, who joined numerous, and have brought in lots of valuable knowledge and created a fun atmosphere. Michael Fogarty with his deep knowledge of silicon quantum dots, Oscar Kennedy an expert on what can and what can’t be done in the LCN clean-room and Siddharth Dhomkar who knows NV-centres like no other and has adapted from living in New York to Milton Keynes. The new students, Virginia Ciriano, Jingyu Duan and Sofia Patomäki who I have had lots of fun with, not only in helping them to get started with experiments but also exploring beautiful Sydney, Blue Mountains and Melbourne, and it is great to see their first experimental achievements.

The Hitachi Cambridge and Microelectronics group, for a very warm welcome even though both of my visits have been in winter time. Nick Lambert for letting me stay in his house in Histon, which was a great change to busy London. Alessandro Rossi who introduced me to Bluefors fridges and makes great Aperol Spritz. Jaimes Haigh who has a great understanding of general physics and who is also great at football. Emma Ball and Lisa Ibberson who I had refreshing conversations with. Imtiaz Ahmed, David Ibberson and Thomas Wagner who have been great PhD companions during my stay.

The Copenhagen group, where I have spent a great month. Special thanks to Ferdinand Kummuth for inviting me to their impressive and well-run lab, and Fabio Ansaloni and Anasua Chatterjee, the silicon soldiers in a GaAs dominated environment, and Rob McNeil and Tommaso Pregolato for welcoming me in their flat. Thanks to everyone else at the Center for Quantum Devices who has made my stay so enjoyable.

The Centre for Doctoral Training and the EPSRC for providing training and funding throughout my PhD. Andrew Fisher and Dan Browne who have been great directors. Lopa Murgai who has been a great administrator and was always fun to speak to. Most importantly my PhD cohort who has made moving to London a great experience. The hours spent together on problems but also activities such as skiing in Vancouver and many others will be unforgettable.

CEA-Leti and imec who have fabricated excellent devices. The MOSQUITO consortium who have been great partners in device characterisation. Irfan Siddiqi’s group who fabricate excellent Josephson parametric amplifiers.
I am extremely thankful to my parents and family, who have supported me unconditionally throughout my whole studies, even when I decided to move to London. Bernhard & Monika Schaal, Judith & Sébastien Langevin and Elias. Escaping the busy academic world in London for a while for exceptional home-cooked food, and very dark and quiet nights has been a treat.

Lastly, my partner Mary Raveendran, who I have spent the happiest years with. It took a bit of time adapting to living in London, but Mary has shown me the great and loving parts of London. Thank you so much for always being there and for generating a great thesis writing timetable. I am looking forward to many more years together.
Publications and Conferences

Publications and Manuscripts


Chapter 4 contains work presented in (2) and (5). Results shown in Chapter 5 have been published in (1) and (3). Finally, Chapter 6 contains work presented in (3) and (4).
Conference talks


Contents

1 Introduction ......................................................... 1
   1.1 Moore’s law ................................................. 1
   1.2 Quantum computation ....................................... 2
   1.3 Quantum computing platforms .............................. 4
   1.4 The silicon advantage towards scalable quantum computation . 6
   1.5 This thesis .................................................. 7

2 Background .......................................................... 9
   2.1 Quantum dots ................................................ 9
      2.1.1 Coulomb blockade ...................................... 9
      2.1.2 Artificial atom ......................................... 12
      2.1.3 Coupled quantum dots .................................. 14
   2.2 Silicon spin qubits .......................................... 18
      2.2.1 Silicon host material .................................. 18
      2.2.2 Single-spin qubit ....................................... 20
      2.2.3 Singlet-triplet qubit ................................... 26
      2.2.4 Multi-spin qubit ....................................... 27
      2.2.5 Readout mechanisms .................................... 28
   2.3 Radio-frequency reflectometry .............................. 32
      2.3.1 Motivation ............................................... 32
      2.3.2 Resonator ................................................ 33
      2.3.3 Single-electron transistor ............................. 37
      2.3.4 Gate-based readout ..................................... 39
         2.3.4.1 Electron exchange with reservoirs ................. 39
         2.3.4.2 Inter-dot exchange .................................. 42
         2.3.4.3 Optimization for spin readout .................... 44
   2.4 Parametric amplification ................................... 46
      2.4.1 Amplification and noise ................................ 46
      2.4.2 Noise limits ........................................... 48

XIII
2.4.3 Ultra-low noise amplification ........................................ 49
  2.4.3.1 Parametric amplification process ............................ 49
  2.4.3.2 Josephson junction devices .................................. 51
2.5 Silicon qubit summary and benchmarks ............................... 54

3 Experimental methods ..................................................... 57
  3.1 Fabrication ............................................................. 57
    3.1.1 SOI nanowire transistors ..................................... 57
  3.2 Measurement setup ................................................... 60
    3.2.1 Dilution refrigerator and wiring .............................. 60
    3.2.2 PCBs .............................................................. 65
    3.2.3 Measurement equipment ...................................... 65
    3.2.4 Reflectometry setup ........................................... 66
    3.2.5 Charge sensitivity measurement ............................... 67

4 Fast and sensitive gate-based RF-readout ................................ 71
  4.1 Readout fidelity ....................................................... 71
  4.2 Parallel high-$Q$ resonator ........................................ 74
    4.2.1 Circuit and simulations ....................................... 75
    4.2.2 Setup and implementation .................................... 81
    4.2.3 Charge sensitivity ............................................. 84
    4.2.4 Video-mode readout .......................................... 86
  4.3 Josephson Parametric Amplifier .................................... 88
    4.3.1 Setup and JPA tuning .......................................... 89
    4.3.2 $LC$ resonator and JPA operation ............................ 95
    4.3.3 QD device ...................................................... 97
    4.3.4 Charge sensitivity ............................................. 98
    4.3.5 Video-mode improvements .................................... 100
    4.3.6 Inter-dot charge transition ................................... 101
      4.3.6.1 SNR improvements ....................................... 103
      4.3.6.2 Readout fidelity estimations ............................. 104
  4.4 Impact of external magnetic fields ................................ 105
  4.5 Conclusion ............................................................ 106

5 Random-access readout .................................................. 109
  5.1 Motivation ............................................................. 109
  5.2 Conditional readout .................................................. 110
    5.2.1 Single electron memory cell .................................. 110
5.2.2 Charge sensitivity ........................................ 115
5.2.3 Electron retention time .................................. 117
5.3 Sequential readout ........................................... 119
  5.3.1 Frequency overlap ..................................... 120
  5.3.2 Operation guidelines .................................. 123
  5.3.3 Sequential readout ..................................... 129
5.4 Power dissipation ........................................... 131
5.5 Conclusion .................................................. 132

6 Scalable architectures ...................................... 133
  6.1 Motivation .................................................. 133
  6.2 Error correction architectures ............................ 134
    6.2.1 Surface code ......................................... 134
    6.2.2 Proposed surface code implementations .......... 136
    6.2.3 Surface code architecture based on multi-electron mediators 145
      6.2.3.1 Multi-electron QD as mediator ................ 145
      6.2.3.2 Surface code architecture .................... 146
      6.2.3.3 Device designs ................................ 148
  6.3 Integration ............................................... 148
    6.3.1 Electronics in a full-stack processor ............. 149
    6.3.2 Scalable readout architecture ..................... 152
      6.3.2.1 Architecture .................................. 152
      6.3.2.2 Transistor model ............................... 153
      6.3.2.3 Circuit simulations ............................ 154
      6.3.2.4 Footprint estimations ......................... 158
  6.4 Conclusion .............................................. 161

7 Conclusion and outlook .................................... 163
  7.1 Achievements .............................................. 163
    7.1.1 CMOS foundry-based quantum devices .............. 163
    7.1.2 High sensitivity gate-based sensing ............... 164
    7.1.3 Scalable quantum computing architectures ........ 164
  7.2 Future research directions ................................ 165
    7.2.1 CMOS devices ...................................... 165
    7.2.2 Gate-based readout ................................ 166
    7.2.3 Integration ........................................ 166

Appendices .................................................... 169
A Study of RF-SETs in Si-MOS technology 171
B Small signal limit of dispersive circuit QED 187
C Even parity inter-donor/dot charge transition 191
D Back-gate operation with high-quality resonators 195
E Silicon qubit summary and benchmarks 197
F Notes on experimental procedures 201
G Code 207
Bibliography 217
Chapter 1

Introduction

1.1 Moore’s law

Over the last five decades we have undergone a revolution in information technology and electronics driven by the innovation of a single circuit element – the transistor. These developments have enabled fabrication of fast and complex electronic circuits with low power consumption and ever-growing processing power, capable of storing and analysing the vast amounts of data produced by a modern digital lifestyle. Continuous transistor miniaturisation has been the key development in reaching the point we are at today. More and more transistors are integrated on the same size chip, reducing the cost per chip and allowing fabrication of larger circuits. Transistor down-sizing additionally reduces the power dissipated by a circuit, which has a dynamic and static contribution of \( P = CV_{DD}^2 f + V_{DD}I_L \) and depends on the capacitance \( C \), frequency \( f \), supply voltage \( V_{DD} \) and leakage current \( I_L \). As the power that can be dissipated on a chip is limited, reducing the transistor size and therefore the capacitance makes operation at higher frequency possible, resulting in improved performance. To scale to transistors of a size of 20 nm and less, enhancements of the carrier mobility (e.g. by using a shorter gate length) allowed for smaller operation voltages \( V_{DD} \). Moreover, leakage \( I_L \) and transistor threshold \( V_T \) has been reduced using high-\( k \) gate dielectrics and an optimised three-dimensional transistor design, consisting of a gate that wraps around a silicon nanowire (FinFEET) [1]. These developments allowed operation at an even higher frequency, overcoming power dissipation related limitations. As a result of these innovations, transistor density in integrated circuits has doubled every two years over the last 50 years, following a trend predicted by Moore, the co-founder of Intel, in Moore’s law [2]. In contrast, state-of-the-art complementary-metal-oxide-semiconductor (CMOS) circuits have reached dimensions of only a few nanometres. In 2019 some foundries started production of the 5 nm node, where a single atom being in the wrong place can
strongly affect circuit performance [3]. A fundamental physical limit has therefore been reached and the semiconductor industry has formally acknowledged that Moore’s law is coming to an end [4]. Nowadays processors are predominately becoming more powerful by using multi-core architectures. Supercomputers consist of a network of such processors and can perform more than $10^{15}$ floating-point operations per second. Still, state-of-the-art supercomputers are not powerful enough to tackle computationally hard problems where no algorithms exist to find a solution within polynomial time, such as simulating quantum chemistry, many-body physics and some optimisation tasks. A significant increase in computing power or new ideas are required to address these problems. Interestingly, defects and quantum effects limiting stability and downscaling of conventional processors are promising candidates for realising a quantum computer capable of solving computationally hard problems that seem intractable using conventional computers [5].

### 1.2 Quantum computation

With the idea in mind, that there will be a physical limit in the density of transistors, researchers had already started to investigate alternative methods to conventional computation in the 1980’s. Amongst the first to imagine that a quantum system could be used to perform a computation was Richard Feynman. He introduced the idea of performing quantum simulations by mapping a quantum system of interest onto a controllable quantum device [6]. Only a few years later, the foundation of a universal (general purpose) quantum computer with capabilities beyond traditional computation was laid out by David Deutsch [7] showing that the Church-Turing principle of universal computation and quantum theory are compatible.

The building block of a quantum computer is a so called quantum bit (qubit), realised using a quantum mechanical two-level system. As a result of the quantum
principle of superposition, any mixture of the two states, typically referred to as 0 and 1, is allowed, making quantum computing appear similar to analogue computing. However, when a qubit is measured only either a 0 or 1 is obtained. Repeated measurements enable the exploration of the probabilistic nature of these projective quantum measurements with $|\alpha|^2$ and $|\beta|^2$ being the probability to measure 0 or 1 respectively. Any superposition state of a qubit can be visualised on a so called Bloch sphere (sphere of radius 1) where the north and south pole correspond to the logical states 0 and 1.

$$|\psi\rangle = \alpha |0\rangle + \beta |1\rangle = \cos\frac{\theta}{2} |0\rangle + e^{i\varphi} \sin\frac{\theta}{2} |1\rangle$$ \hspace{1cm} (1.1)$$

Superposition gives rise to a quantum parallelism, which can be illustrated by considering a $n$-qubit superposition state which generates a vast $2^n$ dimensional Hilbert space

$$|\psi\rangle = \sum_{i=0}^{2^n-1} \alpha_i |i\rangle.$$ \hspace{1cm} (1.2)$$

Moreover, qubits can be entangled where they act as a group, rather than being fully isolated. These effects illustrates the power of a quantum computer and the difficulty of classically simulating such quantum systems.

To make use of superposition and entanglement within a computation, well-tailored algorithms are required to unleash the potential computational speed up. In gate-based quantum computation an algorithm is performed by manipulating qubits in a series of gates realised using unitary transformations acting on a single or pairs of qubits. Similar to conventional computing, multiple universal sets of gates exist which any possible operation on a quantum computer can be reduced to: e.g. the controlled-NOT gate between two qubits combined with single qubit rotations forms a universal set [8]. With Shor’s factoring quantum algorithm [9] the theoretical vision of quantum computation was brought to the world’s attention in 1994 by becoming a threat to common RSA public-key encryption due to an expected exponential speed-up compared to conventional computers in solving the factorisation problem. Later on, Grover presented an algorithm [10] that has a quadratic speed-up when searching an unstructured database. Moreover, as envisioned by Feynman, a quantum computer allows solutions to be found to many problems in chemistry and physics [11–15], such as understanding reactions and materials which could lead to efficient fertilizer production, novel pharmaceuticals and more efficient solar cells. Research on algorithms has shown that algorithms
formulated so far are based on a small number of different concepts, e.g. quantum Fourier transform (Shor), amplitude amplification (Grover), quantum walks and quantum simulation [5]. There is ongoing research in designing new algorithms.

### 1.3 Quantum computing platforms

David DiVincenzo [16] has summarised the requirements for physically implementing a qubit as follows: It needs to be a “scalable physical system with well characterized qubits” and the “ability to initialize the state of the qubits” while remaining coherent “much longer than the gate operation time”. Additionally there must be a “universal set of quantum gates” and a “qubit specific measurement capability” allowing manipulation and readout of every qubit reliably with high-fidelity. Finding and engineering a quantum system which fulfils these somewhat contradictory criteria, of a well isolated but still on-demand well controlled system, represents the challenge of building a quantum computer.

Many proposals on implementing a quantum computer have emerged [17]. Nuclear magnetic resonance (NMR) techniques of molecular spin states as well coupled ions in ion-traps have served as model systems for proof-of-principle demonstrations with up to a 7-qubit NMR implementation of Shor’s algorithm [18]. A state-of-the-art summary of the most promising technologies towards implementation of a useful quantum computer is shown in Fig. 1.2. Among these, implementations
based on superconducting circuits [19–21] and semiconductor ion-traps [22–24] are the most promising candidates to demonstrate an advantage over conventional computers based on devices with close to 100 qubits and high-fidelity but still imperfect gates. This noisy intermediate-scale quantum (NISQ) [25] regime is especially interesting for quantum chemistry applications. While superconducting qubits have a short lifetime, fast gate operations and are operated at milli-Kelvin temperature, ion-based qubits have a long lifetime, slow gate operations and can operate at room-temperature (with performance increases at cryogenic temperature).

Technologies that are currently underperforming but have the potential to deliver useful quantum computers beyond the NISQ era at a large scale, are photonic and silicon-based implementations. Photons are not only excellent candidates for quantum communication [26, 27] but also quantum computation [28]. However, improvements in single photon sources and detectors (typically operating at 4 K) and reaching the necessary interaction between photons pose challenges towards a photon-based computer. In silicon-based implementations, qubits are formed using states of confined electron or nuclear donor spins [29–31]. These spin states have long coherence times and fast gate operations are achieved. Moreover, the structures necessary to form a qubit are smaller than 100 nm and qubits could be operated at cryogenic temperatures of 1 K or higher [31, 32]. Only a few qubits have been demonstrated in a photonic or silicon-based device so far, but both are excellent future prospects.

All of these technologies have seen significant commercial investments [33] with some working towards a full-stack solution and others working specifically on hardware or software. These hardware and especially software developments make quantum computers more accessible to anyone and simulating quantum physics was never as easy as it is now. Even simple algorithms can be run on a small and noisy quantum computer in the cloud hosted by companies such as IBM, Google, Rigetti and IonQ. An example of simulating a qubit experiment using just 10 lines of code within the *cirq* framework is shown in Fig. 1.3(a), which could also be run on Google’s quantum chip. In the experiment, a qubit is initialised in the $|0\rangle$ state followed by a qubit rotation with angle $\theta$ (see Fig. 1.1) and the probability of measuring the $|1\rangle$ state is obtained by repeating the experiment 100 times for a specific angle while the qubit is subject to depolarising noise. Such experiment demonstrates oscillations between the $|0\rangle$ and $|1\rangle$ state with an overall decay in the maximum probability to obtain $|0\rangle$ or $|1\rangle$ due to depolarisation noise.
1.4 The silicon advantage towards scalable quantum computation

Quantum state coherence plays an important role for a quantum computer due to the fragile and non-digital nature of quantum states. Consequently, errors in a quantum computation do not only have bit-flip character. Even a very small error in the phase can alter the result of the computation making error detection and correction and high-fidelity operations an essential ingredient for large-scale quantum computation. The error rate in a quantum computer implementation becomes a parameter as important as the number of qubits to assess how powerful a computer can be. This is captured in a metric called quantum volume [35] which is a function of the number of qubits \( N \) and the circuit depth \( d \) given by the number of operations that can be performed before an error occurs and the qubit connectivity. The quantum volume is illustrated in Fig. 1.3(b) which shows that a system in the regime of quantum supremacy, referring to demonstrating a computational advantage over conventional computers, could never reach the regime of large-scale quantum computing by increasing the number of qubits alone. The desired trajectory in the \( N-d \) space follows a diagonal path by increasing the number of qubits and the circuit depth at the same time.

Protecting qubits from errors comes at the cost of a large overhead in the number of qubits: e.g. on the order of \( 10^8 \) physical qubits are expected to be required to perform algorithms fault-tolerantly [5]. Silicon-based implementations are promising.
candidates towards realising such large-scale fault-tolerant quantum computers due to long coherence times and fast gates (see Fig. 1.2), allowing many operations to be performed before any information is lost, as well as, most importantly, due to their small size and compatibility with established industrial fabrication used to fabricate conventional computers. Modern processors contain billions of transistors on a mm-size chip and compatibility with complementary metal-oxide-semiconductor (CMOS) processes allows fabrication of qubits at a density similar to transistors. Moreover, qubits could be directly integrated with conventional circuits for efficient signal routing, generation and processing.

Two qubit devices and simple algorithms have been demonstrated on current silicon based qubit implementations [36]. Larger devices can be fabricated but coherent control is yet to be demonstrated [37, 38]. Single qubit gates can be performed at a fidelity [39–41] compatible with fault-tolerant operation when employing error correction schemes. Two qubit gates still need to be improved for fault-tolerant operation and fidelities are approaching 90% [36, 42–44]. Charge noise has been identified as the fundamental challenge to improve gate fidelities further [41, 45] and multiple schemes to overcome current limitation have been proposed [40, 46] while still being an active area of research. High-fidelity readout is demonstrated [47–49]. Proposals for scaling of current few qubit demonstrators to large-scale devices have emerged [50–55] but many concepts are yet to be tested experimentally and improvements on readout fidelity and speed in a compact gate-based architecture, compatible with large scale integration, are required [56, 57].

1.5 This thesis

As introduced above, high-fidelity qubit readout combined with a scalable architecture is an essential requirement for a large-scale quantum computer. In this thesis an approach to high-fidelity and scalable readout of silicon-based qubits is presented. The devices studied in this thesis have been fabricated using ‘CMOS friendly’ processes at semiconductor foundries. Gate-based radio-frequency reflectometry readout is a compact way of spin readout in nano-devices which does not require any additional structures next to the spin qubit. By optimising lumped element resonators used to perform gate-based readout of silicon devices and by combining with a Josephson parametric amplifier (JPA) improvements in the quality factor and noise performance are presented leading to fast high-fidelity readout. These developments bring gate-based reflectometry techniques closer to dispersive readout in the field of superconducting qubits, where high-quality transmission line
resonators fabricated alongside the qubits are used and JPAs have been developed. In contrast to transmission line resonators, the lumped element resonator presented in this thesis can be fabricated separately from the qubits allowing for high qubit density. Furthermore, efficient readout of two devices using an integrated transistor-based control circuit is shown, where devices are read out one after another using the same resonator and line. Finally, a scalable random-access readout architecture, enabling readout of dense qubit arrays, and a fault-tolerant qubit architecture based on mediated exchange, providing additional space for electron reservoirs between qubits, are presented.

This thesis is divided into chapters as follows. Chapter 2 and 3 contain background theory, necessary concepts as well as descriptions of the methods and devices used in this work.

In Chapter 4, fast and sensitive gate-based radio-frequency readout of quantum dots in nanowire transistor devices fabricated at CEA-Leti is demonstrated using an improved resonator based on a NbN spiral inductor and a JPA that operates in the sub-1-GHz regime. This work highlights that a resonator, that is fabricated separately from quantum dot devices, can serve as a route towards compact high-fidelity readout by means of circuit and noise improvements.

In Chapter 5, random-access gate-based readout is introduced and sequential readout of two devices using an integrated transistor circuit is demonstrated. This work demonstrates that the number of lines required for readout of many qubits could be reduced quadratically.

In Chapter 6, scalable architectures for silicon-based quantum computers and ways to integrate conventional electronics with qubits are introduced. A surface code architecture based on mediated exchange resilient to leakage errors and a scalable readout architecture based on random-access and frequency multiplexing that can be integrated within a dense qubit architecture is presented. By performing footprint estimations, this work provides the basis for implementation of a scalable radio-frequency readout architecture for dense quantum dot arrays.

Finally, in Chapter 7, key outcomes of this thesis and future research directions are discussed.

In the Appendices, a study of an alternative non-scalable readout method using external charge sensors, further details on optimising gate-based readout and the analysis of an even parity inter-donor/dot charge transition, impact of back-gate operation on a high quality resonator, benchmarks of silicon qubit implementations, notes on experimental procedures and programming code are presented.
Chapter 2

Background

This chapter briefly introduces the basic concepts that are required to follow the experimental results presented in this thesis. Further details on specific topics can be found in review articles and references cited throughout the text.

2.1 Quantum dots

In sub-micron size conductive islands, the model of non-interacting electrons that is almost exclusively used to describe the behaviour of electrons in the solid state in a simple framework, breaks down [58, 59]. In this regime, the behaviour of electrons is dominated by electron-electron interaction and quantum confinement resulting in atom-like electronic structures. Effects like Coulomb blockade, the electronic level structure of a quantum dot and coupling of multiple quantum dots are summarised and details can be found in Refs. [59–62].

2.1.1 Coulomb blockade

To begin, an isolated conductive island is considered. The total charge $Q$ on the island must be an integer number of the single electron charge $Q = Ne$ which provides an electrostatic energy of

$$E_{el} = \frac{Q^2}{2C} = \frac{e^2}{2C}N^2$$

with $C$ being the capacitance of the island. This illustrates that there is an energy cost associated with adding electrons to the island related to classical charge quantization and the Coulomb interaction. To observe this effect, the island is tunnel coupled to source and drain reservoirs, with voltage $V_{s/d}$ respectively, from which electrons can tunnel on and off the island. Additionally, the island is capacitively
coupled to a gate at voltage \( V_g \), that allows tuning of the potential of the island. This three electrode device as shown in Fig. 2.1(a) is known as a single electron transistor (SET) and is often described in the constant-interaction (CI) model which takes into account the Coulomb interaction and parametrises the system in terms of the electron number \( N \) on the island, the source, drain and gate voltage and capacitance. The source and drain tunnel barriers are represented by a resistance \( R_{s/d} \) that is usually larger than the resistance quantum \( R_q = h/e^2 = 25.8 \, \text{k}\Omega \). Assuming constant capacitances and neglecting any second order effects, the electrostatic energy of the SET is given by \([62]\)

\[
E_{el}(N) = \frac{(eN - C_g V_g - C_s V_s - C_d V_d)^2}{2C_{\Sigma}}
\]

\[ C_{\Sigma} = C_g + C_s + C_d \]
2.1 Quantum dots

The electrochemical potential of the island is defined as the energy required to add the \(N\)th electron to the system

\[
\mu(N) = E_{el}(N) - E_{el}(N-1) = \frac{e}{C_{\Sigma}} [(N - 1/2)e - C_g V_g - C_s V_s - C_d V_d],
\]

which shows that each voltage relates to a shift of the electrochemical potential by the amount \(e\alpha_i V_i\) with the leverarm

\[
\alpha_i = \frac{C_i}{C_{\Sigma}}.
\]

Finally, adding a single electron to the island requires the addition energy of

\[
\Delta E_{add} = \mu(N+1) - \mu(N) = \frac{e^2}{C_{\Sigma}} = E_C.
\]

The energy \(E_{el}\) as a function of \(V_g\) is shown in Fig. 2.1(c) for different numbers of electrons on the island, which results in multiple equally separated parabolas. The parabolas cross at several points representing points of energy degenerate charge configurations. When the island is tunnel coupled to a reservoir, electrons tunnel onto or off the island as a function of \(V_g\) at the degeneracy points minimizing the total energy. At any other voltage the charge configuration remains constant. This is captured in the schematic of the tunnel coupled island shown in Fig. 2.1(b), which shows the electrochemical potential \(\mu(N)\) of the island and the reservoirs \(\mu_s/\mu_d\). Vacant and filled states in the reservoirs follow a Fermi-Dirac probability distribution \(f_{FD}(E)\) leading to broadening of the electrochemical potential around a value set by the source and drain voltages leading to an effective electron temperature \(T_e\). On the island there is a ladder of discrete states available. A change in \(V_g\) shifts the electrochemical potential \(\mu(N)\) of the island linearly up and down. Transport through the island is only possible when one of the levels on the island falls within the bias window determined by the of the source \(V_s\) and drain \(V_d\) voltage. In Fig. 2.1(b) the level \(\mu(N+1)\) is aligned with the source and drain \((V_s = V_d)\) and transport is possible while at slightly different voltage (with a difference larger than the broadening of states in the source and drain) transport is blocked. This transport characteristic is known as Coulomb blockaded (CB) oscillations which are characterised by regular spikes in conductance as a function of \(V_g\) as shown in Fig. 2.1(d).

When \(V_{sd} = V_d - V_s \neq 0\), the window at which transport is possible broadens which gives rise to so called Coulomb diamonds [59, 63] as shown in Fig. 2.2. Within the diamonds transport is blocked and the occupation of the island remains constant.
The addition (and charging) energy $\Delta E_{\text{add}}$ and gate coupling $\alpha_g$ of the SET device can be determined from the diamonds as illustrated.

### 2.1.2 Artificial atom

When the dimensions of the island become smaller than the Fermi wavelength $\lambda_F = 35 \text{ nm}$ associated with the electrons in the silicon lattice, an additional contribution to the system arises due to quantum confinement [58]. This single-particle level spectrum (or orbital states) depends on the confinement geometry and can resemble the energy spectrum of an atom which is why such small islands are referred to as artificial atoms or quantum dots (QD). For example, in case of the particle in a box problem (infinite square well of length $L$) a quadratic dispersion relation is obtained for every dimension [58]

$$E_n = \frac{\pi^2 \hbar^2}{2m^*L^2} n^2. \quad (2.6)$$

Another example is a purely harmonic confinement potential ($V(r) = \frac{1}{2}m^*\omega_0^2 r^2$), where the single-particle spectrum is given by the Fock-Darwin spectrum, characterised by quantum numbers $n$ and $l$ [58]

$$E_{n,l} = (2n + |l| + 1) \hbar\omega_0. \quad (2.7)$$

---

**Figure 2.2:** Coulomb diamonds. (a) Exemplary configuration of Coulomb blockade and (b) when both source and drain are aligned with two consecutive levels on the island. (c) Illustration of the current through the SET device as a function of $V_g$ and $V_{sd}$. Within dark regions transport is possible while white regions indicated blockade. Positions of (a) and (b) and ways to determine $\Delta E_{\text{add}}$ and $\alpha_g$ are indicated. Adapted from [59].
The single-particle energy is usually smaller than the charging energy $e^2/C_\Sigma$. In addition to the orbital states, electrons that are confined in a silicon lattice have a fine-structure of the single-particle spectrum given by the spin and valley degree of freedom. Assuming that the given single-particle spectrum $E_n$ and Coulomb interaction $E_{el}$ can be treated independently, both can simply be combined

$$E(N) = E_{el}(N) + \sum_N E_n$$

(2.8)

such that an addition energy can be derived as

$$\Delta E_{\text{add}} = E_C + \Delta E$$

(2.9)

where $\Delta E = E_{N+1} - E_N$ is the single-particle level spacing. The single-particle spectrum is observed in transport measurements of quantum dots where every excited state that fall within the bias window leads to increased transport (see Fig. 2.3(a,c)). Especially in the low electron regime variations in $\Delta E_{\text{add}}$ are observed due to spin dependent filling as shown in Fig. 2.3(b,c) [64, 65]. The spin and valley filling can be studied further in magneto-spectroscopy from the slope $\frac{\partial \Delta E_{\text{add}}}{\partial B}$ [65, 66].
2.1.3 Coupled quantum dots

So far only single quantum dots have been discussed. Now a double quantum dot (DQD) is considered, where both dots can be capacitively and tunnel coupled via the mutual capacitance $C_m$ and resistance $R_m$. Additionally both dots are coupled to one gate $V_{g1(2)}$ and reservoir $V_{d(s)}$ each as shown in Fig. 2.4(a). Similar to the case of a single QD, a DQD can be described in the same model by converting into a matrix based description where every potential is a linear function of each voltage determined by the capacitance matrix $Q = CV$ from which the electrostatic energy is obtained $E_{el}(N) = 1/2VCV$. The charging energies of a DQD are then defined as [61]

$$E_{C1(2)} = \frac{e^2}{C_{1(2)}} \left( \frac{1}{C_{1(2)}} - \frac{C_m}{C_{1(2)}} \right)$$  \hspace{1cm} (2.10)$$

$$E_{Cm} = \frac{e^2}{C_m} \left( \frac{1}{C_m} - \frac{C_{1(2)}}{C_m} \right)$$  \hspace{1cm} (2.11)$$

$$C_{1(2)} = C_{d(s)} + C_{g1(2)} + C_m$$  \hspace{1cm} (2.12)$$

where $C_{1(2)}$ is the sum of capacitances relevant for QD 1(2). Figures 2.4(b-d) show the DQD occupation numbers $(N_1, N_2)$ as a function of $V_{g1}$ and $V_{g2}$ in different regimes of coupling. When $C_m \to 0$ both dots are completely decoupled and occupation of both dots is tuned independently using $V_{g1(2)}$ as shown in Fig. 2.1(b). In the opposite limit $C_m \to C_{1(2)}$, the dots are completely coupled and behave effectively as large single QD with occupation $N_1 + N_2$ that is equally coupled to $V_{g1(2)}$ as shown in Fig. 2.4(d). The most interesting is the intermediate case shown in Fig. 2.4(c), where a honeycomb pattern and triple points are observed. Only at these triple points is transport through the DQD possible, as illustrated in Fig. 2.4(e). In the high $V_{sd}$ bias regime the triple points turn into so called bias triangles from which more insights into the DQD level structure and spin configuration can be obtained [61, 63]. So far only capacitive coupling via the mutual capacitance $C_m$, giving rise to the mutual charging energy $E_{Cm}$, has been discussed. Capacitive coupling is the basis for many readout schemes as discussed in Section 2.2.5. However, fast control of charge and spin requires significant tunnel coupling between dots which is discussed in the following paragraphs.

**Charge qubit** First, the case of overall odd DQD parity is discussed with the example of a single electron within the DQD, where the single QD states are given by $|0\rangle = (1, 0)$ and $|0\rangle = (0, 1)$ representing the electron being either in the left
or right dot. With significant tunnel coupling between two QDs, electrons are no longer fully localised in single dots and form bonding and anti-bonding molecular orbitals $|\pm\rangle = \alpha_{\pm}|0\rangle \pm \alpha_{\mp}|1\rangle$ leading to bending of the honeycomb lines. Defining $\varepsilon$ as the detuning parameter that sets the energy splitting via tuning of a gate voltage and can take the system from the (1,0) to the (0,1) state, the Hamiltonian of this two level system is given by $H = -\varepsilon \sigma_z$. The tunnel coupling $t_c = \Delta_c/2$ provides the off diagonal terms representing transitions between both states $H_{\Delta_c} = -\frac{\Delta_c}{2}\sigma_x$ such that the overall Hamiltonian and eigenenergies are [62]

$$H(t) = -\frac{\Delta_c}{2}\sigma_x - \frac{\varepsilon}{2}\sigma_z \quad (2.13)$$

$$E_{C_{\pm}} = \pm \frac{1}{2}\sqrt{\varepsilon^2 + \Delta_c^2} \quad (2.14)$$

where $\sigma_i$ denote the Pauli spin matrices. At zero detuning, the bonding and anti-bonding states are shifted from the (avoided) crossing by $\pm t_c$ leading to an overall splitting of $2t_c = \Delta_c$ as shown in Fig. 2.5(a) and the system has equal probability to be found in the (1,0) or (0,1) state. At zero magnetic field, bonding and anti-bonding states are spin degenerate. Using microwaves with an energy that matches the $E_{C_{\pm}}$ splitting the system can be coherently driven between both states, through a process known as photon assisted tunnelling and energy level spectroscopy and measurements of the charge coherence time can be performed [67]. When driven

---

**Figure 2.4: DQD.** (a) Schematic of a DQD. Each QD is coupled to one reservoir and gate and there is mutual coupling between the QDs. Transport characteristic through a DQD as a function of both gate voltages in the regime of weak (b) intermediated (c) and strong (d) coupling. Adapted from [61].
Figure 2.5: DQD energy diagram. (a) DQD in the odd parity configuration at the example of a (1,0) to (0,1) charge transition. Tunnel coupling between the dots leads to an avoided crossing. (b) DQD in the even parity configuration at the example of a (1,1) to (0,2) charge transition. Spin singlet and triplet states are formed (symmetric triplet states not shown). The triplets remain in the (1,1) charge configuration as a function of \( \varepsilon \). (c) Complete spectrum of an even parity DQD which includes the symmetric triplet states split off by \( E_z \), the triplets of (0,2) charge configuration split off by \( E_{ST} \) and the triplet avoided crossings at larger \( \varepsilon \). Coupling between different spin number is neglected due to small spin-orbit and hyperfine couplings.

Pauli spin blockade For even DQD parity (even total number of electrons), electron transfer in a coupled DQD becomes spin dependent and singlet-triplet states

fast (non-adiabatically) through the anti-crossing using an oscillating voltage \( \varepsilon(t) \) within the charge coherence time, Landau-Zener interferometry can be performed where electron interference is observed [68].
are formed. The spin singlet and triplet states are defined as follows \[63\]

\[
|S\rangle = (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)/\sqrt{2}, \tag{2.15}
\]

\[
|T_0\rangle = (|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle)/\sqrt{2}, \tag{2.16}
\]

\[
|T_+\rangle = |\uparrow\uparrow\rangle, \tag{2.17}
\]

\[
|T_-\rangle = |\downarrow\downarrow\rangle. \tag{2.18}
\]

In addition to these spin configurations, the electrons can be configured across the two QDs in the (1,1) or (0,2) charge state due to finite tunnel coupling. As inter-dot tunnelling preserves spin, only (1,1) singlet(triplet) states couple to (0,2) singlet(triplet) states. When both electrons are in the same dot (0,2), singlet and triplet states are split by \(E_{ST}\) which can be on the order of 1 meV \[69\] as the triplet occupies an excited single-particle level to fulfill fermionic wave-function antisymmetry (Pauli exclusion principle). However, deep in the (1,1) charge configuration singlet and triplet have a different (usually much smaller) splitting given by \(J_0 = 4t_c^2/E_{Cm}\) in the Hubbard approximation \[63\]. As a result of spin conservation and different singlet-triplet splittings for the (1,1) and (0,2) configuration, the avoided crossing for the triplets is found at a much different value of detuning than the avoided crossing for the singlet as shown in Fig. 2.5(c). Additionally, at finite magnetic field the three triplet states are split by the Zeeman energy \(E_z = g\mu_B B\) and the diagram reduces to a two level \(S\) and \(T_0\) system in a small range of detuning as shown in Fig. 2.5(b) with energies \[63, 70\]

\[
E_{S\pm}(1,1) = E_{S\pm}(0,2) = \pm \frac{1}{2} \sqrt{\epsilon^2 + \Delta_c^2} \tag{2.19}
\]

\[
E_{T_0}(1,1) = \epsilon/2 \tag{2.20}
\]

where \(E_{T_0}(1,1) = \epsilon/2 \pm g\mu_B B\) and \(E_{T_0}(0,2) = -\epsilon/2 + E_{ST}\) are split off and the triplet (0,2) anti-crossings are neglected. These features of a DQD with even parity result in a phenomenon called spin blockade, where tunnelling is forbidden for the triplet states (locked into the (1,1) configuration) due to spin conservation, while the singlet can transition between a (1,1) and (0,2) charge configuration. Spin blockade can be observed as a region of blocked transport in bias triangles in one direction while for the opposite sign of \(V_{sd}\) transport is possible \[69\]. Spin blockade plays an important role for realising spin qubits and provides a way to distinguish spin states via the charge configuration as discussed in Section 2.2.5.
Electron exchange and spin blockade has not only been demonstrated in silicon QDs [69, 71, 72] but also between donors [73] and hybrid donor-dot devices [74, 75].

2.2 Silicon spin qubits

In semiconducting devices, the charge and spin degree of freedom of single electrons provides a suitable two level system to form a single qubit. Compared to charge, the spin degree of freedom couples much more weakly to fluctuations in the electric and magnetic environment leading to longer coherence times. However, the weak coupling makes manipulation and readout of a single spin more challenging. Nevertheless, spin qubits are the preferred qubit implementation in silicon.

Electron spin qubits in silicon can be formed using electrons that are confined in QDs, as introduced in the previous section, or using electrons that are bound to a donor atom which provides a natural confinement potential. Concepts that apply to electrons confined in QDs also apply to electrons bound to donors. However, the confinement reached using donors is typically stronger (e.g. the wave function of an electron bound to $^{31}$P has an expected spread of 2.38 nm) and the number of electrons that can be added to a donor at cryogenic temperature can be limited (e.g. no electron, one electron and two electrons). Moreover, the electron spin bound to a donor is subject to hyperfine coupling with the nuclear spin and the nuclear spin itself can be used to form a spin qubit. Furthermore, spin transfer from the electron spin to the nucleus spin is possible.

In the following essential concepts on the formation, manipulation and readout of different spin qubit implementations in silicon are discussed. Detailed reviews can be found in Ref. [63, 76].

2.2.1 Silicon host material

Semiconductor based qubits were first implemented in GaAs based systems. These materials can be grown with extremely high purity and atomic layer precision, which has led to the formation of two-dimensional electron gases (2DEG) of very high quality. Additional electron confinement is reached by device patterning via etching or using voltages applied to metal structures at the surface, where confinement in all remaining dimensions can be reached resulting in so called quantum dots [58]. Silicon has similar properties but offers some advantages as a host material for spin-based qubits, specifically in terms of qubit coherence and large scale fabrication,
However, some challenges had to be overcome in order to implement spin qubits in silicon devices [76].

To begin, the reduced effective mass of electrons in silicon compared to other materials (e.g. factor of $\sim 3$ compared to GaAs) requires scaling down of gate structure dimensions in order to reach electron confinement, which makes fabrication more challenging. Due to advances in lithography techniques, quantum devices are now routinely fabricated in university and industrial clean-rooms [37, 77]. Moreover, bulk silicon is a semiconductor with indirect band gap and cubic symmetry leading to a six-fold degenerate conduction band minimum. This valley degeneracy could in principle be problematic when forming a quantum bit that relies on a non-degenerate two-level system, however, a sufficient valley splitting of up to 1.5 meV (that is to some degree controllable) is achieved for spin qubits formed in donors and nanostructures [78–81] due to confinement, electric fields and strain.

Having overcome these challenges, the excellent mechanical and electrical properties and highly developed large scale fabrication industry make silicon an ideal host material for fabrication of large-scale quantum processors. Not only can the qubits benefit from industrial scale fabrication when fabricated in a CMOS compatible process, but also integration with conventional electronics is facilitated which can provide solutions for efficient control, readout and fast feedback when transitioning from a small-scale to a full-scale quantum computer. Additionally, silicon-based spin qubits offer very long coherence and relaxation times compared to other solid-state platforms such as superconducting qubits. Relaxation of a spin due to energy exchange is characterised by the timescale $T_1$ and is usually mediated by phonons. Phonons can not directly induce spin relaxation of pure spin states as angular momentum is not conserved. However, spin-orbit coupling, that arises for electron spins in a lattice potential and electron spins confined in quantum dots [82, 83], couples spin levels to orbital levels lifting phonon-related relaxation rules. Silicon has no piezoelectric phonons and an overall weak spin orbit coupling (due to no bulk inversion symmetry, small atomic number and large band gap), resulting in a long spin relaxation time on the order of seconds in electron spin qubit devices at cryogenic temperature [32, 39, 80]. Finally, silicon has a low natural abundance of nuclear spin carrying isotopes $^{29}$Si (5%) and can be purified to mostly (e.g. 800 ppm) nuclear spin free isotopes $^{28}$Si and $^{30}$Si. The hyperfine interaction between a spin qubit and constantly fluctuating spins in the host lattice provides a large contribution to decoherence time described by $T_2$ (irreversible loss of coherence), which is strongly suppressed in the ‘semiconductor-vacuum’ of isotopically purified silicon, where $T_2$ can be charge noise limited [41]. Consequently, even at cryogenic
temperature up to 1.45 K operation of a spin qubit formed using a single electron spin is possible \cite{32}. In this regime, spin relaxation increases and is dominated by second-order phonon processes but relaxation times longer than reported dephasing $T^*_2$ (reversible loss of coherence) and most decoherence $T_2$ times are achieved \cite{84}. Ionized donors can even realise quantum memories with decoherence times exceeding minutes at room temperature \cite{31}, highlighting the prospects of silicon based quantum computing at elevated cryogenic temperature.

### 2.2.2 Single-spin qubit

A qubit based on a single spin $1/2$ is one of the most natural two level systems that is formed by the Zeeman split spin up and down levels $|0\rangle = |\downarrow\rangle$ and $|1\rangle = |\uparrow\rangle$ in an external magnetic field $B_z$ (usually $> 1$ T) with Hamiltonian $H_0 = g\mu_B B_z \hat{S}_z = \frac{\hbar\omega_0}{2} \sigma_z$.

**Coherent control** Coherent oscillations between these levels can be achieved using pulsed nuclear or electron spin resonance (NMR/ESR) which is often performed using an ac magnetic field $B_{ac}$ (perpendicular to the external field $B_z$) at frequency $\omega \approx g_e(n)\mu_B B_z/\hbar$ (typically radio or microwave frequency) where $g_e(n)$ is the electron(nuclear) $g$ factor. For a donor based spin qubits, there can be an additional hyperfine splitting \cite{85, 86}. The time-dependent Hamiltonian of a spin $1/2$ subject to such an external field $B_z$ and drive $B_{ac}$ is given by

$$H(t) = \frac{\hbar\omega_0}{2} \sigma_z + \hbar\Omega \cos(\omega t + \delta) \sigma_x$$

with $\Omega = g\mu_B B_{ac}/\hbar$ and $\Omega \ll \omega_0, \omega$. This Hamiltonian is solved by transforming into the reference frame rotating at frequency $\omega$ of the ac field $B_{ac}$ and neglecting faster oscillating components \cite{87} which leads to an effective Hamiltonian

$$H_{rot} = \frac{\hbar}{2} \begin{pmatrix} \Delta & \Omega e^{-i\delta} \\ \Omega e^{i\delta} & -\Delta \end{pmatrix}.$$  

(2.22)

Thus, $z$-rotations are performed at angular frequency given by $\Delta = \omega_0 - \omega$, while rotations around an axis in the $x$-$y$ plane (determined by the phase $\delta$) are performed at angular frequency $\Omega$ (determined by $B_{ac}$). Figures 2.6(a,b) show an exemplary donor and quantum dot spin qubit device respectively. In the donor device a donor nuclear and electron spin is implanted close to the silicon surface while the electron spin in the quantum dot is confined either at a Si/SiO$_2$ interface or in a Si/SiGe...
heterostructure using voltages applied to the gates at the surface (see Section 3.1 for fabrication details). In both devices spin manipulation is achieved using a shorted on-chip waveguide (strip-line) that ends very close to the spin, where the oscillating current provides an oscillating magnetic field. Multiple spins can be controlled with the same strip-line (as long as close enough). Local $g$ factor anisotropies and/or hyperfine interaction [88–90] can lead to a distribution in manipulation frequencies. High fidelity control (99.6% and better) of multiple spins has been achieved [91], however the frequency of coherent spin oscillations is so far limited to $\sim 1$ MHz due to the magnitude of $B_{ac}$ delivered at the spin. Transition frequencies of donor and electron spin can be tuned using electric fields [91–93] such that many spins could be controlled using a global field at a single frequency given that electric tuning is fast and sufficient magnitude $B_{ac}$ is achieved.

Additionally, electron spins can be manipulated using electric fields via electric dipole spin resonance (EDSR) [94] through mechanisms that indirectly couple spin to electric fields. EDSR is achieved based on spin-orbit coupling [95] and asymmetries in the local environment, which can be in the form of a local difference in $g$ factor [89, 90], a different local nuclear spin environment [96] or using a micro-magnet [66, 97–101]. The latter is the most successful approach as it can provide local field differences on the order of $\Delta B_z = 30$ mT between spins resulting in coherent oscillations up to 30 MHz [100, 101]. An exemplary micro-magnet is shown in Fig. 2.6(c) and manipulation is achieved using an oscillating voltage that translates into an effective oscillating magnetic field in the rest frame of the electron spin which is moved quickly through an anisotropic environment.

**Qubit operation and characterisation** For qubit operation, spins are initialised using spin-selective tunnelling [102] or by loading a random spin from a reservoir followed by waiting for time $t_{\text{wait}} \gg T_1$ until the spin relaxes. Relaxation can be enhanced by pulsing to a specific relaxation hot-spot [36, 80]. Single qubit operation and characterisation is performed in pulsed experiments as shown in Fig. 2.6(d) for the example of an electron spin confined in a $^{28}$Si/SiGe based QD device. By initialising a spin-up state followed by a measurement after time $t_{\text{wait}}$ the relaxation time $T_1$ is obtained from the exponential decay of the spin-up probability. Coherent (Rabi) oscillations between the spin-up and down state can be observed in an experiment where the duration $t_p$ of a microwave drive is varied. The in-plane dephasing time $T_2^*$ is obtained from the free induction decay in a Ramsey sequence, where the spin is initialised, brought into a superposition state for a varying free evolution time $\tau$, and brought back followed by a measurement. The decoherence
time $T_2$ is obtained in a Hahn-echo sequence, which is similar to the Ramsey experiment but includes a spin-flip in the middle of the free evolution time which results in refocussing of the spin. Multiple spin-flips can be incorporated for further improvements known as Carr-Purcell-Meiboom-Gill (CPMG) sequence.

Dephasing and decoherence times of $T_2^* = 120 \mu$s and $T_2^{\text{CPMG}} = 28$ ms for electron spins in QDs [91], and $T_2^* = 270 \mu$s, $T_{2e}^{\text{CPMG}} = 560$ ms and $T_2^{n} = 1.75$ s and $T_{2n}^{\text{CPMG}} = 35$ s for the electron and nuclear spin of donors [39] in isotopically purified $^{28}$Si have been demonstrated combined with control fidelities exceeding 99.6% in all implementations and reaching up to $> 99.9\%$ using pulse optimisation [40] demonstrating that a spin $1/2$ systems in silicon provides an excellent basis for spin qubits and memories.

**Two qubit gate** The interaction between electron spins is the basis for two qubit gates in donor and QD devices. The complete two electron spin Hamiltonian can be formally derived in second quantized form including contributions from the on-site energy $H_\varepsilon$, the Zeeman energy $H_z$, valley degeneracy $H_v$, Hubbard-style Coulomb repulsion $H_U$ and tunnel coupling $H_t$. Assuming sufficient valley splitting and ignoring the higher energy (0,2) triplets the Hamiltonian can be reduced into the basis of $\{(1,1) : |\uparrow, \uparrow\rangle, |\uparrow, \downarrow\rangle, |\downarrow, \uparrow\rangle, |\downarrow, \downarrow\rangle; (0,2) : \frac{1}{\sqrt{2}} |0, \uparrow\downarrow - \downarrow\uparrow\rangle\}$ and written as [104, 105]

$$H = \begin{pmatrix} E_z & 0 & 0 & 0 & 0 \\ 0 & \delta E_z/2 & 0 & 0 & t_c \\ 0 & 0 & -\delta E_z/2 & 0 & -t_c \\ 0 & 0 & 0 & -E_z & 0 \\ 0 & t_c & -t_c & 0 & U - \varepsilon \end{pmatrix}. \quad (2.23)$$

The individual terms in this Hamiltonian can be recognized from the discussion in Section 2.1.3 and the energy diagram in Fig. 2.5(c) with $E_z$ being the Zeeman energy, $\delta E_z$ the difference in Zeeman energy between both dots, $t_c$ the tunnel coupling, $\varepsilon$ the detuning and $U$ the charging energy for moving both electrons into the same dot. Using a Schrieffer-Wolff transformation [106] and assuming $U - \varepsilon \gg t_c$ the Hamiltonian can effectively (to first order) be reduced to the well known Heisenberg interaction in the basis of individual (1,1) spin states
2.2 Silicon spin qubits

Figure 2.6: Spin qubit device operation and characterisation. Donor device consisting of a single nuclear and electron spin (a) and QD device consisting of two trapped electron spins (b). Both devices include a microwave strip-line for control and SET for readout. Electrons are loaded onto the QD via a reservoir. (c) QD device as in (b) but with a cobalt micro-magnet instead of a strip-line allowing for electrical control. (d) Set of experiments for spin characterisation. Measurement of the relaxation time, manipulation speed through Rabi oscillations, dephasing time in a Ramsey sequence and decoherence time in a Hahn echo experiment. Adapted from [36, 44, 66, 91, 103].
Figure 2.7: Two qubit gate. (a) Probability of $|↑↓⟩$ and (b) energy difference between $|↑↓⟩$ and $|↓↑⟩$ as a function of $δE_z$. SWAP gate is performed at small $δE_z$ where singlet and triplet are the eigenstates, while a CPHASE gate is obtained at large $δE_z$. (c) Illustration of resonant drive between two qubit states for $J = 0$ and $J > 0$. At $J > 0$ oscillations of one spin conditional on the other spin can be performed. (d) Quantum state tomography of a Bell state prepared using a two-qubit gate based on large Zeeman difference. (e) Change in drive frequency of qubit 1 as qubit 2 oscillates between two states over time demonstrating conditional resonant driving. Adapted from [36, 44, 105].

$$\{ |↑⟩, |↑⟩, |↓⟩, |↓⟩ \} \quad [104]$$

$$H_{int} = \begin{pmatrix} E_z & 0 & 0 & 0 \\ 0 & \frac{δE_z}{2} - \frac{J}{4} & \frac{J}{2} & 0 \\ 0 & \frac{J}{2} & -\frac{δE_z}{2} - \frac{J}{4} & 0 \\ 0 & 0 & 0 & -E_z \end{pmatrix} \quad (2.24)$$

with $J = \frac{2\varepsilon^2}{U−\varepsilon−δE_z} + \frac{2\varepsilon^2}{U−\varepsilon+δE_z}$ being the Heisenberg exchange interaction strength. Based on this interaction there are three ways to construct a two qubit gate.
2.2 Silicon spin qubits

**Exchange gate** For Heisenberg exchange only, when $\delta E_z \ll J$, the off diagonal terms drive a flip-flop of the two spins and singlet-triplet states are eigenstates which allows realisation of a SWAP gate. Combined with single qubit rotations along the $z$ axis $R_z^{(1)/2}(\theta)$ (of qubit 1 and 2) a CNOT can be constructed in five steps: $U_{\text{CNOT}} = R_z^{(1)}(\pi/2)R_z^{(2)}(-\pi/2)U_{\text{SWAP}}^{-1/2}R_z^{(1)}(\pi)U_{\text{SWAP}}^{-1/2}$. The exchange interaction strength $J$ can be tuned electrically via $t_c$ or $\varepsilon$. The latter has so far been preferred in experiments in coupled QDs and donors [42, 107, 108] and sub-ns $\sqrt{\text{SWAP}}$ operations has been achieved with a fidelity up to 90%. Gate fidelity is charge noise limited and could be improved using symmetric operation, where exchange is tuned via $t_c$ instead of $\varepsilon$ which has demonstrated improved noise performance [46].

**Phase gate** When $\delta E_z \gg J$ dominates, $\ket{\uparrow, \downarrow}$ and $\ket{\downarrow, \uparrow}$ are eigenstates of the system and a phase of equal magnitude but different sign is acquired over time between the states, which allows for a controlled phase gate that is achieved when the phase difference equals $\pi$. The CPHASE gate can be constructed in a single fast pulse by tuning $J(\varepsilon)$ which is one of the reasons why it has been the preferred experimental implementation of a two qubit gate in silicon so far and has been performed based on an intrinsic difference [104] in $\delta E_z$ and using a micro-magnet [36]. With the latter a rotation can be performed within 50 ns and a Bell state fidelity of $F = 89\%$ has been achieved which is a measure of the gate fidelity. An exemplary state-tomography for the $\Psi^\pm$ Bell state is shown in Fig. 2.7(d) where any contributions besides the four central contributions are due to imperfections.

**Resonant gate** Similar to the hyperfine interaction between the donor nuclear and electron spin, giving rise to 4 level system with distinct drive frequencies, an alternative resonantly driven two qubit gate for electron spins in QDs can be constructed based on the tunable exchange interaction $J$ and spin resonance techniques (ESR/EDSR). While at $J = 0$ single spins can be driven unconditional on the second spin due to an equal energy difference, at $J > 0$ distinct drive frequencies form depending on the second spin allowing for single spin manipulation conditional on the second spin as shown in Fig. 2.7(c). This has been demonstrated [44, 109] and is illustrated in Fig. 2.7(e) where two different drive frequencies for qubit 1 are observed over time while qubit 2 is oscillating between the up and down state. A controlled rotation with gate fidelity of up to $F = 98\%$ within 2 $\mu$s has been demonstrated. A faster gate could be performed when combined with EDSR and strong micro-magnets [100, 109].
2.2.3 Singlet-triplet qubit

Instead of encoding quantum information in a single spin, the singlet-triplet ($S-T_0$) qubit makes use of the $S_z = 0$ subspace of the combined two electron spin states as defined in Eq. (2.16) and shown in Fig 2.5(b), where the symmetric triplets are split off by an external magnetic field. This is appealing because it allows for single-qubit gates via the electrically controlled exchange interaction between the two spins, which drives two qubit gates in a qubit encoded in a single spin as discussed above [107, 108, 110]. Moreover, the $S_z = 0$ subspace is protected from overall magnetic field fluctuations which has led to longer coherence times especially in early GaAs based implementations [111]. By transforming Eq. (2.23) into a singlet triplet basis and making similar approximations in the regime $U - \varepsilon \gg t_c$ and neglecting higher energy states, the $(1,1)$ singlet and triplet Hamiltonian is given by [104]

\[
H_{ST_0} = \begin{pmatrix}
J(\varepsilon) & \Delta E_z \\
\Delta E_z & 0
\end{pmatrix},
\]

(2.25)

Here $J(\varepsilon) = \frac{1}{2}(U - \varepsilon) + \sqrt{2t_c^2 + \frac{1}{4}(U - \varepsilon)^2}$ is the exchange interaction and $\Delta E_z = \delta E_z \cos \theta$ is the local Zeeman energy difference with $\theta(\varepsilon, t_c)$ being an angle introduced by the transformation. In this system, the exchange interaction generates a phase (rotations along the $z$ axis), while the local Zeeman difference drives oscillations between the $S$ and $T_0$ state (along the $x$ axis) providing full electrical single qubit control as shown in Fig. 2.8(a). Another advantage of the $S-T_0$ qubit is the possibility of electrostatic coupling between two $S-T_0$ qubits due to the difference in charge configuration between a singlet, that can be found in $(0,2)$, while the triplet is only found in the $(1,1)$ configuration as shown in Fig. 2.8(b). However, such a gate is sensitive to charge noise and has not shown high fidelity so far [112]. Similar to a spin $1/2$ qubit, $S-T_0$ qubits can additionally be coupled via the exchange interaction [113, 114]. Pauli spin blockade also allows for initialisation into a singlet state at $\varepsilon > 0$. For high-fidelity single qubit gates a controlled Zeeman difference $\delta E_z$ is desired which can be achieved via dynamical nuclear spin polarisation [110] or using a micro-magnet [66, 97-101]. The system of a donor and quantum dot electron spin as shown in Fig 2.8(c) provides an excellent $S-T_0$ qubit due to the hyperfine field generated by the nuclear spin leading to a well defined local field difference between the donor and dot electron spins [74], especially when the nuclear spin is large. For a phosphorus donor-dot $S-T_0$ qubit manipulation at 57 MHz is
2.2 Silicon spin qubits

Figure 2.8: Singlet-triplet qubit. (a) Bloch sphere representation of a $S-T_0$ qubit. (b) Difference in the preferred charge configuration of a singlet and triplet state. (c) Illustration of a donor-dot electron spin hybrid $S-T_0$ where the donor spin can serve as a memory, the hyperfine coupling provides $\Delta E_z$ and capacitive or exchange based gates between two cells could be achieved. Reproduced from [75, 112].

demonstrated [75]. Detailed methods for operation of $S-T_0$ qubits can be found in Ref. [115].

2.2.4 Multi-spin qubit

A single qubit can also be encoded in states involving more than two spins. The so called ‘exchange-only’ qubit is formed using the two-dimensional subspace ($S = 1/2, s_z = 1/2$) of three singly occupied quantum dots (1,1,1) [116]: $|1\rangle = \sqrt{\frac{3}{2}} |\downarrow\rangle \otimes |\uparrow\uparrow\rangle - \frac{1}{\sqrt{6}} |\uparrow\rangle \otimes (|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle)$ and $|0\rangle = \sqrt{\frac{1}{2}} |\uparrow\rangle \otimes (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)$. In this basis, the exchange interaction between the left and central spin and between the right and central spin provides two axis for fully electrical universal qubit control. Two qubit couplings using exchange interaction and electrostatic coupling are suggested [117]. Full control of such an exchange-qubit has been demonstrated in GaAs [114] and all electrical control makes it attractive. However, complex pulse sequences are required and the full Hilbert space spanned by a triple quantum dot system provides many leakage states outside the computational subspace.

A similar type of exchange only qubit can be formed in a double quantum dot using three spins in the (1,2) configuration. The logical qubits states are identical but due to the singlet and triplet configuration in the second there is an additional
energy difference between $|0\rangle$ and $|1\rangle$. As a result of this single qubit pulses must be very fast, but the two anti-crossings where the excited state $|E\rangle = |S \uparrow\rangle$ couples to $|1\rangle$ ($(1,2)$ configuration) and $|0\rangle$ ($(2,1)$ configuration) allow for full electrical control [119]. Charge noise can be problematic due to charge based manipulation. Single qubit has been demonstrated [120, 121] and Coulomb based two qubit gates are suggested [119].

Overall, encoding of a qubit using three spins allows for all electrical control at the cost of higher complexity, more leakage states and larger device footprint.

### 2.2.5 Readout mechanisms

In the previous section, different ways of forming and controlling spin qubits in silicon quantum devices were discussed. This section, focusses on the mechanisms for readout of the charge and spin of a single electron in a quantum dot or dopant device. The magnetic moment of a spin, which is proportional to the Bohr magneton $\mu_B = 9.274 \times 10^{-24} \text{J/T}$, is very weak. For a spin $1/2$ the energy difference between spin up and down is $\sim 100 \mu\text{eV}$ at a field of $1$ T. Conventional spin resonance techniques allow detection of at least $10^{13}$ spins [122]. Using micro-resonators a sensitivity of 240 spins has been demonstrated [123, 124]. NV-centres allow sensing of single-spins in close proximity [125] with a sensitivity to external magnetic fields (which includes the field generated by nearby spins) on the order of $10$ nT/$\sqrt{\text{Hz}}$ [126, 127]. Combined with scanning techniques magnetometry at sub-nanometre preci-
sion can be performed using NV-centres [125, 128–130]. The key to readout of single spins in nanostructures such as quantum dots is to convert from the spin degree of freedom to the charge degree of freedom, with the single electron charge being $e = 1.6 \times 10^{-19}$ C, combined with detection through highly-sensitive charge sensors. The energy related to a single charge in nanostructures is typically a few meV.

**Charge sensors** A current of a nano-ampere corresponds to movement of millions of electrons on a timescale of milliseconds, illustrating that the current originating from a single electron could not be measured directly. However, very precise charge sensors capable of detecting single electron tunnelling, can be formed in nanostructures that exhibit forms of quantized transport. Charge sensors are either formed in one-dimensional constrictions (quantum point contact, QPC) or using charge islands such as SETs and QDs, and are placed in close proximity to the qubit device as shown in Fig. 2.10(a-b). When biased at the point of maximum transconductance, changes in the local charge distribution transfer into measurable changes in the QPC/SET current as illustrated in Fig. 2.10(c-f). Charge sensing is possible even when the tunnel rate of a transition is too opaque to be measured in transport.

**Spin-to-charge conversion** Energy or spin-selective tunnelling allows readout of the electron spin in a QD by coupling to a reservoir as illustrated in Fig. 2.11(a). The discrete spin up and down states in the QD are split using a magnetic field $B_z$ into $\mu_\uparrow$ and $\mu_\downarrow$ and for readout the edge of the continuous states in the reservoir, the Fermi level $\mu_{\text{res}}$, is aligned in the centre between the up and down state. As a result of this, an electron in the spin up state prefers to tunnel onto the reservoir, followed by a spin down electron tunnelling back into the QD. On the contrary, an initial spin down electron remains in the QD [47, 102]. Thus, for spin-selective tunnelling, the QD is pulsed from the region of stable charge configuration close to the dot-to-reservoir transition as indicated for the (1,0) and (0,1) configuration of a DQD in Fig 2.11(c). Following the pulse, a change in the signal is observed when the spin up electron tunnels out (after time $t_{\text{out}}$) of the QD, which leaves the QD empty, and remains until a spin down electron tunnels back in ($t_{\text{in}}$) as shown in Fig 2.11(d). The key concept for spin readout using this method is a difference in tunnel rate for a spin up and down electron, which can also be achieved when both spin up and down are brought above the Fermi level and spin up and down is distinguished using a time window at which tunnelling occurs due to different
Figure 2.10: Charge sensors. (a) QD structure coupled to a quantum point contact (QPC) sensor and (b) single electron transistor (SET) sensor. Both current through the QD and sensor can be measured. Transport characteristic of (c) the sensor QPC and (d) SET with the operation point indicated. (e) Differential sensor current and (f) QD current as a function of gate voltage. Adapted from [63, 101, 131].

tunnel rates (tunnel-rate-selective readout) [132]. Both approaches require tuning of the dot-to-reservoir tunnel rates to be within the detection bandwidth.

In a DQD, Pauli spin blockade allows for spin to charge conversion without an external reservoir, which distinguishes between the spin singlet and triplet states and allows direct readout of a $S$-$T_0$ qubit [107] or readout of a single spin using an ancillary spin [133]. Typically $T_{+}$ states are split off using a small magnetic field and the $S$-$T_0$ states form the computational states. At zero detuning ($\varepsilon = 0$) there is a small splitting between the $S(1,1)$ and $T_0(1,1)$ state and the $S(1,1)$ and $S(0,2)$ charge states are degenerate, while the $T_0(0,2)$ is split off due to Pauli exclusion, see Fig 2.11(b). At positive detuning ($\varepsilon > 0$) a $(0,2)$ charge state becomes energetically favourable, and the $S(1,1)$ state can transition to $S(0,2)$ via elastic tunnelling (after time $t_{\text{out}}$), while the $T_0(1,1)$ state remains blocked until lifted via spin relaxation (after time $T_1$). The detuning axis $\varepsilon$ and the region of spin blockade is indicated in Fig 2.11(c) and the change in charge configuration of the singlet is detected by a nearby charge sensor as shown in Fig. 2.11(e), allowing for singlet-triplet readout, when the sensor is well aligned with the DQD charge dipole as the overall number of charges in the DQD does not change. When a DQD is additionally coupled to a
2.2 Silicon spin qubits

$\mu_1 \uparrow \mu_\downarrow \mu_\text{res} \mu_\uparrow \mu_\downarrow \mu_\text{res}$

$B_z k T \gg B_T$ $T_0(0,2)$ $T_0(1,1)$ $T_0(0,2)$ $T_0(1,1)$ $S(1,1)$ $T_0(1,1)$ $S(1,1)$ $S(0,2)$

$\varepsilon > 0$

\[ \varepsilon = 0 \]

\[ \varepsilon > 0 \]

$\mu_1 \uparrow \mu_\downarrow$ $\mu_\text{res}$ $\mu_1 \uparrow \mu_\downarrow$ $\mu_\text{res}$

Figure 2.11: Spin to charge conversion. (a) Spin selective tunnelling of Zeeman energy split spin up and down states in a QD to a reservoir. (b) Singlet-triplet Pauli spin blockade in a tunnel coupled DQD. (c) Stability diagram of a DQD with charge occupation indicated by numbers and colouring. Arrows indicate pulses related to reservoir and spin blockade based readout. Charge sensor signal for spin up and down in a spin-selective tunnelling (d) and singlet-triplet spin blockade measurement (e). Adapted from [103].
reservoir the $S(0,2)$ state can be selectively mapped to the (1,0) charge configuration (direct enhanced latching readout) or the $T_0(1,1)$ can be mapped to the (1,2) charge configuration (reverse enhanced latching readout) [48, 134] as indicated by dashed arrows in Fig. 2.11(c). This enhanced latching mechanism improves the signal, due to state selective change of the DQD occupation by a whole electron, and extends signal lifetime by changing from spin dominated relaxation to a metastable charge state.

To maximise readout fidelity in reservoir based readout, the Zeeman splitting ($\sim 0.1 \text{meV/T}$) should be much larger than the thermal energy ($\sim 0.09 \text{meV/K}$). For spin-blockade based readout the singlet-triplet (or valley) splitting is the energy scale limiting fidelity due to relaxation. A singlet-triplet splitting on the order of 1 meV is observed [69, 71, 135] even at low fields. Moreover, singlet-triplet readout forms the basis for parity measurements required in error-correction schemes [136].

### 2.3 Radio-frequency reflectometry

In Section 2.2.5, the physical mechanisms for readout of semiconductor spin qubits based on spin-to-charge conversion were presented. In this section, limits in detection bandwidth and sensitivity of charge sensors are discussed and improvements using radio-frequency (RF) techniques are introduced.

#### 2.3.1 Motivation

Readout that is faster than the relaxation time ($T_1$) of the qubit is essential as any information would otherwise be lost. Consequently, in order to achieve high-fidelity, a sufficient readout sensitivity and bandwidth are required. Non-invasive charge sensors close to the spin qubit device provide spin readout using spin-to-charge conversion (as discussed in Section 2.2.5), where the readout signal is given by a change in the current through the sensor. The change in sensor current can be maximised, apart from careful sensor design and operation [137], using cryogenic current amplifiers [138].

The charge sensitivity $\delta q$ provides a benchmark on charge sensor performance and depends on the current noise spectral density $S_I(\omega)$ and the sensitivity of the current to changes in the charge environment ($\partial I/\partial Q$) [139]

\begin{equation}
\delta q = \sqrt{\frac{S_I(\omega)}{\partial I/\partial Q}}.
\end{equation}
2.3 Radio-frequency reflectometry

$S_I(\omega)$ can have contributions from thermal (Johnson-Nyquist) [140, 141] noise and shot noise (Poissonian tunnelling events) [142]. At electron temperatures achieved in a dilution refrigerator ($T \approx 100$ mK), shot noise dominates and a theoretical limit of $\delta q = 2.7 \mu$e/\sqrt{Hz}$ has been derived for experimental parameters [143, 144]. However, in experiments at a bandwidth up to 4.4 kHz a sensitivity of only $20 \mu$e/\sqrt{Hz}$ has been achieved [145]. This can be explained by an additional contribution to $S_I(\omega)$ from charge noise originating from fluctuations in the occupancy of multiple charge traps nearby. Different switching rates between individual traps lead to a time-dependent charge offset with a spectral density that has a $1/f$ dependence [45, 146]. Considering the resistance (e.g. 100 kΩ) of the sensor and capacitive contribution from the device and cables (on the order of 0.1–1 nF), the bandwidth of measurements based on transport through such sensors is limited to frequencies on the order of kHz [147], where charge noise limits the sensitivity.

Using RF techniques, an SET can be operated at frequencies larger than 100 MHz, where bandwidth and sensitivity limitations can be overcome. Additionally, RF techniques can also be used to form a gate-based sensor that turns the metallic gates that define the qubits themselves into a sensor. The mechanisms behind RF readout are explained in the next sections starting with the introduction of resonators.

2.3.2 Resonator

Impedance is an important factor in RF circuits. In order to efficiently transfer power from a source to a load, impedance matching is required. Matching networks are capable of matching the potentially large impedance of a device under test to the impedance of the coaxial line ($Z_0 = 50 \Omega$). In microwave engineering [148], the power transfer is characterised by the scattering parameter ($S$-parameter). A two port network, as shown in Fig. 2.12(a), with incoming waves $V_{in}^1$ and $V_{in}^2$ from port 1 and 2 is represented as

$$
\begin{pmatrix}
V_{out}^1 \\
V_{out}^2
\end{pmatrix} = \begin{pmatrix}
S_{11} & S_{12} \\
S_{21} & S_{22}
\end{pmatrix} \begin{pmatrix}
V_{in}^1 \\
V_{in}^2
\end{pmatrix}.
\] (2.27)

The matrix elements $S_{xy}$ determine what fraction of the wave originating from port $y$ is transferred to port $x$ ($S_{xy} = \frac{V_{out}^x}{V_{in}^y}$). Thus, $S_{11}$ and $S_{22}$ are the reflection coefficients of port 1 and 2, while $S_{12}$ and $S_{21}$ are the reverse and forward transmission. The complete $S$-matrix of a circuit can be measured using a vector-
network-analysers (VNA), which is capable of measuring the magnitude and phase of each matrix element. In such measurements, a test signal is applied to a device under test using a source with tunable frequency and power. The test signal power is typically measured in logarithmic units $P \text{(dBm)} = 10 \cdot \log_{10} \left( \frac{P}{1 \text{mW}} \right)$ with respect to 1 mW. Similarly, the magnitude of the $S$-matrix elements, representing gain and loss, is typically measured in decibels $S_{xy} \text{(dB)} = 10 \cdot \log_{10} \left( \frac{P_{x \text{out}}}{P_{y \text{out}}} \right) = 20 \cdot \log_{10} \left( \frac{V_{x \text{out}}}{V_{y \text{out}}} \right)$.

In RF reflectometry, one port of the network is terminated by the device which is characterized by a large impedance $Z_d$. In this configuration, the parameters $S_{11}$ and $S_{21}$ are the most relevant which describe the reflection $\Gamma$ of the signal and the transmission $T$ of the signal to the device (see Fig. 2.12(b)). Given the input impedance of the line, $Z_0$, and the impedance of the device and matching network, $Z$, the reflection and transmission coefficient is defined as

$$\Gamma = \frac{Z - Z_0}{Z + Z_0} \quad (2.28)$$

$$T = \frac{2Z}{Z + Z_0} \quad (2.29)$$

Without any matching to the line $Z = Z_d \gg Z_0$ and $\Gamma \rightarrow 1$ and $T \rightarrow 0$, which means that no signal is delivered to the device, hindering any readout (at high frequency.) However, using a $LC$ impedance transformer the large device impedance can be
matched to the line, allowing readout at high frequency, which can be achieved by simply adding an inductor $L$ into the circuit which forms a resonator with the parasitic capacitance $C_p$, device capacitance $C_d$ and resistance $R_d$ which is shown in Fig. 2.12(c). Both $C_d$ and $R_d$ depend on device parameters. The impedance of the resonant circuit in Fig. 2.12(c) is given by

$$Z = i\omega L + \frac{R_d}{1 + i\omega R_d C_0}, \quad (2.30)$$

$$C_0 = C_p + C_d. \quad (2.31)$$

Resonance occurs when the reactance is zero ($\text{Im}(Z) = 0$). This results in a resonant frequency of

$$\omega_r = \sqrt{\frac{1}{L C_0} \left(1 - \frac{L}{R_d^2 C_0}\right)} \approx \sqrt{\frac{1}{L C_0}} \quad (2.32)$$

which can be approximated for typical experimental parameters $L/R_d^2 C_0 \ll 1$. Substituting Eq. (2.32) in Eq. (2.31), the equivalent impedance of the circuit close to resonance at $\omega = \omega_r + \Delta \omega$ is [148]

$$Z = \frac{L}{C_0 R_d} + i\omega L + \frac{1}{i\omega C_0} \quad (2.33)$$

$$= \frac{L}{C_0 R_d} + \frac{i}{w} \left(\omega^2 L - \frac{L}{LC_0}\right) \quad (2.34)$$

$$\approx \frac{L}{C_0 R_d} + \frac{i}{w_r} \left(2\Delta \omega \omega_r L\right) \quad (2.35)$$

$$= R_{\text{eff}} + i2\sqrt{\frac{L}{C_0} \frac{\Delta \omega}{\omega_r}} \quad (2.36)$$

with $R_{\text{eff}} = L/C_0 R_d$ which illustrates how the resonator can match the high resistance of the device to an effective resistance. When $R_{\text{eff}} = 50 \, \Omega$ the resonator is perfectly matched to the coaxial line such that at resonance $\left|\Gamma\right| \to 0$ and away from resonance $\left|\Gamma\right| \to 1$. An example of a perfectly matched resonator is shown in Fig. 2.13(a). In addition to the matching, another important parameter is the quality factor of a resonator which characterises the rate of energy losses compared to energy stored in the resonator. When coupled to an external line there is an external quality factor $Q_{\text{ext}}$, internal quality factor $Q_{\text{int}}$ and the overall loaded quality factor $Q_{\text{load}}$. For a series $LCR$ circuit as shown in Fig. 2.12(d) these are given
by [148]

\[ Q_{\text{ext}} = \frac{\sqrt{L/C_0}}{Z_0} \]  
\[ Q_{\text{int}} = \frac{\sqrt{L/C_0}}{R_{\text{eff}}} = R\sqrt{\frac{C_0}{L}} \]  
\[ Q_{\text{load}} = \left( \frac{1}{Q_{\text{int}}} + \frac{1}{Q_{\text{ext}}} \right)^{-1}. \]

The loaded quality factor can be experimentally obtained from the ratio of the full-width half maximum (FWHM) and the resonance frequency \( Q_{\text{load}} = \omega_r/\Delta\omega_{1/2} \). The FWHM is indicated in red in Fig. 2.13(a) and determines the detection bandwidth of reflectometry readout. Based on this, the coupling coefficient which is a measure of impedance mismatch can be defined as \( \beta = Q_{\text{int}}/Q_{\text{ext}} \). For \( \beta = 1 \) the resonator is perfectly matched to the line. For \( \beta < 1 \) the resonator is under-coupled.
and the power is mostly dissipated in the resonator \((R_{\text{eff}} > Z_0)\), while for \(\beta > 1\) the resonator is over-coupled and most of the power is dissipated in the external line \((R_{\text{eff}} < Z_0)\). \(\beta\) can be obtained from the depth of the magnitude of the reflection coefficient \(|\Gamma|\) at resonance [148]

\[
\beta = \begin{cases} 
1 - |\Gamma|, & \text{if under-coupled } (\beta < 1) \\
1 + |\Gamma|, & \text{if over-coupled } (\beta > 1) \\
1, & \text{if impedance matched } |\Gamma| = 0.
\end{cases}
\] 

(2.40)

When the resonator is over-coupled there is an additional phase shift of 180°. This is shown in Fig. 2.13(d) for \(\beta = 1.25\) (and \(\beta = 0.5\)). Detection of small changes in the parametric resistance \(R_d\) or capacitance \(C_d\) of the device connected to the resonator is the basis for RF reflectometry. Figure 2.13 illustrates that small changes in \(R_{\text{eff}}\) result in changes of the magnitude of the reflected signal when operating at a fixed frequency close to resonance, while small changes in the capacitance of the resonator result in a shift of the resonance which is most apparent in the phase of the reflected signal. How such parametric changes relate to readout of a spin qubit device is explained in the following sections.

### 2.3.3 Single-electron transistor

An SET that is operated at the point of maximum transconductance at one of the Coulomb blockade oscillations realises a sensitive charge sensors for readout of spin qubits (see Section 2.2.5). At such operation point, the resistance of the SET is on the order of a resistance quantum \(R_Q = h/e^2 = 26\, \text{k}\Omega\). Thus, a resonator for RF readout, that connects to the source of the SET device, should be matched to a resistance \(R_d\) on the order of 100 kΩ. This is typically achieved using an inductance on the order of 100–1000 nH. Any changes in the electrostatic environment of the SET, such as changes in the charge configuration of a nearby quantum dot (qubit), result in a change of the SET operation point which translates into a change in resistance, which in turn changes the reflection coefficient \(\Gamma\), see Eq. (2.29). Thus, RF readout is performed by monitoring resonator damping, rather than directly measuring the current through the SET as shown in Fig. 2.14(a). The absolute value of the change in reflection coefficient, which determines the sensitivity of this
method, can be calculated as [149]

\[
|\Delta \Gamma| = \frac{|\partial \Gamma / \partial R_d| \Delta R_d}{R_{\text{eff}} Z_0} \left(\frac{R_{\text{eff}} + Z_0}{2 R_d}\right)^2.
\]  

(2.41)

(2.42)

The first term in Eq. (2.42) represents the matching of the resonator and is maximised for \( R_{\text{eff}} = Z_0 \), while the second term represents the fractional change of the resistance. Thus, \( |\Delta \Gamma| \) is maximised when the resonator is well matched and the change in resistance is large.

RF electrometers have first been realised in Al based metal islands and junctions, led by the pioneering experiment by Schoelkopf et al. [147], where sensitivities down to \( \delta q = 1 \mu e/\sqrt{\text{Hz}} \) were achieved [153, 154] in the superconducting and normal state, limited by the noise of the first amplification stage, with a measurement bandwidth up to 10 MHz. RF techniques for readout of semiconductor qubits were first introduced in GaAs/AlGaAs, where qubit coherence is shorter and in such improvements in measurement bandwidth allowed probing the qubit state in the sub-microsecond regime [155]. RF-QPCs [150, 156] and RF sensor QDs [151] with a sensitivity of \( \delta q = 146 \mu e/\sqrt{\text{Hz}} \) have been demonstrated. Shortly afterwards, the first RF SET in Si-MOS devices, shown in Fig. 2.14(b), was demonstrated [152].
with a sensitivity better than $\delta q = 10 \text{µe/}\sqrt{\text{Hz}}$. Using variable capacitors, resonant
frequency tuning and tuning into perfect matching, for further improvement of the
charge sensitivity, have been demonstrated [157–159]. Techniques to determine the
RF charge sensitivity are presented in Section 3.2.5 and a study of RF-SETs in
Si-MOS devices is shown in Appendix A.

2.3.4 Gate-based readout

Inspired by the success of RF electrometers, gate-based RF sensors have been
demonstrated, representing a compact and scalable way for read out of spin qubits.
Rather than using RF techniques to read out a nearby QPC or SET, in gate-based
readout the RF signal is sent directly to the spin qubit through a resonator that
connects to one of the gates defining the nanostructure. In this configuration the
resistance $R_d$, as introduced in Fig. 2.12(c), represents dissipation and is usually
large for a gate that is only capacitively coupled to the QD. The device capacitance
$C_d$ has a geometric contribution (voltage-independent) that simply arises from the
capacitances in the nano-scale device and can often be estimated using parallel plate
approximations. Additionally, there are parametric contributions (voltage depen-
dent) that arise from single-electron tunnelling in the device. As changes in the
impedance originate from the ability of a single electron to tunnel between two
states when driven by the RF signal, there are specific regimes in which detection
is possible depending on how the RF frequency compares with the tunnelling and
relaxation rate in which either a parametric capacitive or dissipative term can dom-
ninate [70, 160–164]. In contrast, SET charge sensors are able to detect changes in
the local charge environment at any tunnelling rate, given that tunnelling occurs
within the sensor detection bandwidth, but sensitivity of an SET is limited at low
frequency due to charge noise. The origin of the parametric signal is discussed below
at the example of a single QD coupled to reservoirs and two coupled QDs (DQD).

2.3.4.1 Electron exchange with reservoirs

The gate-based signal of a resonantly driven QD based on particle exchange with
a reservoir is well understood in terms of a state-dependent tunnelling capaci-
tance [165] and Sisyphus resistance [160] and has been studied in aluminium based
electron/cooper-pair boxes [160, 166, 167], GaAs QDs [168] and silicon QDs [162,
163]. Tunnelling occurs between the zero-dimensional density of states of the QD
and those of the reservoir, the latter having a probability distribution with electron
temperature $T_e$. Such reservoir transitions allow for direct readout of the charge
Figure 2.15: Electron exchange with reservoir. (a) Top: Energy diagram of a QD as a function of reduced gate voltage for different occupation number \(N\). When the QD is tunnel-coupled to a reservoir, the QD transitions from \(E_0\) to \(E_1\) at the indicated degeneracy point. Bottom: Enlarged view of the degeneracy point. A small RF signal with amplitude \(\delta n_g\) can drive the QD across the transition which can lead to relaxation as shown by tunnelling rates \(\gamma_{+/-}\) as a function of \(n_g\). (b) Phase and magnitude response of the resonator as a function of the QD potential relative to the Fermi level \(\varepsilon\). (c) Expected and experimentally obtained capacitive and dissipative response at a \(\varepsilon = 0\) as a function of tunnel rate. Adapted from [162, 163].

The energy of a QD that is coupled to a gate and holds \(N\) electrons is \(E = E_C(N - n_g)^2\), see Eq. (2.2), where \(n_g\) is the effective gate voltage charge. Near a degeneracy point (e.g. \(n_g = 0.5\)) as shown in Fig. 2.15(a), the QD can be understood as a tunnel-coupled two-level system with levels \(E_0\) (dot empty) and \(E_1\) (dot occupied). Consequently, when \(n_g^0 \approx 0.5\) an RF signal with amplitude \(\delta n_g = C_g V_{RF} / e\) (which is state of a QD device (as well as spin related effects, see Section 2.2.5). Additionally, a charge sensor similar to an RF-SET can be formed, where small changes in the gate-based signal when operating at the point of largest slope at a dot-to-reservoir transition are detected. Due to a physically different mechanism of the gate-based signal, such a device would only require a single reservoir (often referred to as single-electron box [160]) as no transport is necessary which results in a different noise spectral density limited by Sisyphus noise [162].
small compared to the energy spacing of the QD) can move the system, that is in the
ground state \( E_0 \), to \( n_0 + \delta n_g \geq 0.5 \) where \( E_0 \) is the excited state. The system can be
described using a Hamiltonian \( H = \tilde{\varepsilon} \sigma_z \) where the eigenenergies \( E_{0/1} = \mp \frac{\varepsilon}{2} \) depend
on the detuning parameter \( \varepsilon = -e\alpha(V_g - V_0) \) that is voltage controlled. The details
of how the system behaves and what kind of signal is generated depends on three
main parameters: the tunnel rate \( \gamma_{+/−} \), the drive frequency \( f_0 \) and the temperature
of the electrons in the lead \( T_e \). When the parameters allow for relaxation via inelastic
tunnelling between the dot and reservoir an energy of \( \Delta E = E_1 - E_0 = \varepsilon \) can be
dissipated and a change in resistance is observed (Sisyphus resistance). Otherwise
the signal is strongly capacitive. Detail are discussed in the following paragraphs.

**Low tunnel rates** At low tunnel rate, \( \hbar \gamma_0 \ll k_B T_e \), the line-shape of the capacitive
\( C_t \) and resistive \( R_{\text{sis}} \) signal is given by [162, 163, 169]

\[
C_t = \frac{e^2 \alpha^2}{4k_B T_e} \frac{1}{1 + \omega_0^2/\gamma_0} \cosh^{-2} \left( \frac{\varepsilon}{2k_B T_e} \right) \tag{2.43}
\]

\[
R_{\text{sis}} = \frac{4k_B T_e}{e^2 \alpha^2} \frac{1 + \gamma_0^2/\omega_0^2}{\gamma_0} \cosh \left( \frac{\varepsilon}{2k_B T_e} \right) \tag{2.44}
\]

with a full-width half maximum (FWHM) that is dominated by the electron tem-
perature, \( \varepsilon_{1/2} = 4 \ln(\sqrt{2} + 1)k_B T_e \), allowing for local thermometry of nanoscale
deVICES [170]. The peak value of the capacitance \( C_t \) goes to zero for \( \gamma_0 \ll \omega_0 \) and
tends to \( e^2 \alpha^2/4k_B T_e \) for \( \gamma_0 \gg \omega_0 \) (with an inflection point at \( \gamma_0 = \omega_0 \)). The peak
value of dissipation, given by \( 1/R_{\text{sis}} \), tends to zero (infinite gate resistance) for both
\( \gamma_0 \ll \omega_0 \) and \( \gamma_0 \gg \omega_0 \), as the electron either doesn’t follow the drive or tunnels as
soon as possible minimising dissipation, and has a maximum at \( \gamma_0 = \omega_0 \). Thus, the
tunnel rate can be estimated depending on whether the signal appears mostly in the
phase of the reflected signal or both in phase and amplitude. This is demonstrated
in an experiment [163] and shown in Fig. 2.15(b), where the resonator response in
the magnitude and phase for different dot-to-reservoir transitions of two QDs are
shown. The peak phase and magnitude as a function of \( \gamma_0 \) extracted from these
measurements follows the trend expected from Eq. (2.44) as shown in blue and red
in Fig. 2.15(c).

**High tunnel rates** At tunnel rates larger than the electron thermal energy, \( \hbar \gamma_0 \gg k_B T_e \), a different regime is reached and can observed in transition \( N_2 = 3 \leftrightarrow 4 \) in
Fig. 2.15(b) which appears much broader compared to other transitions. In this
regime, electron tunnelling occurs elastically and out of phase with the drive and
only produces a capacitive response \[163, 169, 170\]

\[
C_t = \frac{e^2 \alpha^2}{\pi} \frac{\hbar \gamma_0}{(\hbar \gamma_0)^2 + \varepsilon^2}
\] (2.45)

with a FWHM of \(\varepsilon_{1/2} = 2\hbar \gamma_0\), from which the tunnel rate can be obtained. This solely capacitive response is shown in green in Fig. 2.15(c).

As illustrated in the previous paragraphs, a gate-based sensor uses resonant techniques that allow the electron occupation of a SET or QD device to be determined directly (in-situ), similar to transport measurements, without the need for additional sensor nanostructures. Detection is even possible when tunnel barriers to reservoirs are too opaque to observe a sufficient current in transport \[171\] as long as the resonant frequency is comparable to the tunnelling rate as illustrated in Fig. 2.15(c). When a device is operated as a charge sensor, a charge sensitivity of up to \(\delta q = 37 \mu e/\sqrt{\text{Hz}}\) \[162\] has been demonstrated using gate-based techniques, compared to \(\delta q = 1 \mu e/\sqrt{\text{Hz}}\) \[154\] achieved using RF-SETs.

### 2.3.4.2 Inter-dot exchange

**Sisyphus resistance and tunnelling capacitance** The gate-sensor signal originating from an inter-dot charge transition of a DQD is similar to the case of a dot-to-reservoir charge transition, where charge relaxation gives rise to Sisyphus resistance and tunnelling capacitance, and similar expressions can be derived by expanding changes in the charge configuration for small modulations \[164\].

\[
C_t = \frac{(e \alpha')^2}{4k_B T_e} \left( \frac{\varepsilon}{\Delta E} \right)^2 \frac{1}{1 + \omega_0^2 / \gamma_0^2} \cosh^{-2} \left( \frac{\Delta E}{2k_B T_e} \right) \] (2.46)

\[
R_{\text{sis}} = \frac{4R_Q k_B T_e}{h \alpha'^2} \left( \frac{\Delta E}{\varepsilon} \right)^2 \frac{1 + \gamma_0^2 / \omega_0^2}{\gamma_0} \cosh \left( \frac{\Delta E}{2k_B T_e} \right)
\] (2.47)

One important difference to note is that the geometric coupling factor \(\alpha'\) for a DQD is given by the difference in coupling of the sensor gate to each QD: \(\alpha' = \alpha_{\text{QD1}} - \alpha_{\text{QD2}}\). Moreover, the states involved are the energy levels \(E_{C\pm}, E_{S\pm}\) and \(E_{T0}\) of a DQD, as shown in Fig. 2.16(a,b), which form an avoided crossing at the degeneracy point \(\varepsilon = 0\) with \(\Delta E = \sqrt{\varepsilon^2 + \Delta^2}\). As already seen in the case of dot-to-reservoir transitions, \(R_{\text{sis}}\) and \(C_t\) depend on the electron temperature \(T_e\) and are related to phonon induced processes providing the same asymptotic behaviour as a function of inter-dot tunnelling (charge relaxation) rate \(\gamma_0\). \(1/R_{\text{sis}}\) and \(C_t\) of an inter-dot transition show two symmetric maxima around \(\varepsilon = 0\), rather than a single maximum.
Quantum capacitance In addition to $C_t$ and $R_{sis}$ there is fundamentally different contribution related to the energy band curvature of the DQD known as the quantum capacitance [161]. Quantum capacitance has been observed in GaAs [161], InAs [172], carbon nanotube [173] and silicon [72, 163] based DQD devices. The quantum capacitance has been described theoretically [70, 169] and has been unified with Sisyphus resistance and tunnelling capacitance in a small-signal equivalent model [164].

For each charge and spin state of a DQD the quantum capacitance $C_q$ can be obtained from the curvature of the energy as a function of detuning $E_x(\varepsilon)$ (see

---

**Figure 2.16: Quantum capacitance.** (a) Energy diagram and quantum capacitance $C_q$ as a function of level detuning $\varepsilon$ with effectively odd (charge states) and (b) even (singlet-triplet states) DQD occupancy. (c) Line-shape of $C_q$ as a function $\varepsilon$ for different electron temperature $T$ ($k_BT_e/\Delta_c = 0.1, 0.5, 1, 2, 10$ from black to turquoise). (d) Magnetic field response of $C_q$ for even occupancy in the low temperature and slow relaxation regime $k_BT_e/\Delta_c = 0.1$. Adapted from [70, 161].

at $\varepsilon = 0$, as the energy of an absorbed and emitted phonon are on average the same [164].
Eq. 2.14 & 2.20) \cite{70,164}

\[ C_q^x = -e^2 \alpha'^2 \partial^2 E_x / \partial^2 \varepsilon \]  
\[ C_{q^\pm}^S = \mp e^2 \alpha'^2 \Delta c^2 / \Delta E^3 \tanh \left( \frac{\Delta E}{2k_B T_e} \right) \]  
\[ C_{q^\pm}^{T_n} = 0 \]

and is independent of the relaxation rate. The energy diagram and quantum capacitance \( C_q \) for a DQD of odd and even occupation is shown in Fig. 2.16(a,b) respectively. In the even parity configuration, only the singlet state produces a quantum capacitance, while both the excited and ground state of the odd parity configuration have a non-zero quantum capacitance. The maximum signal is obtained at low temperature as shown in Fig. 2.16(c), while the signal disappears when \( k_B T > \Delta c \). Consequently, at low temperature the quantum capacitance dominates over phonon induced Sisyphus and tunnelling contributions. Moreover, fading of the quantum capacitance signal can be observed when applying a large external magnetic field \( B_z \) (larger than the singlet-triplet splitting) where the \( T_\pm \) becomes the ground state for

\[ \varepsilon > \frac{\Delta c^2 - (2g_B B)^2}{4g\mu_B B} \]  

as shown in Fig. 2.16(d), illustrating a simple method for DQD parity measurements \cite{172}. Finally, the state-dependent quantum capacitive signal allows for direct singlet-triplet readout of a DQD when spin and charge relaxation is slow compared to the drive \( T_1^c, T_1^s \gg \omega_0^{-1} \).

\subsection*{2.3.4.3 Optimization for spin readout}

In the previous section, it was shown that the response of a gate-based sensor to electron tunnelling between a single QD and reservoir, or within a DQD, is mostly capacitive and only has a dissipative contribution when relaxation rates are comparable to the drive frequency. For spin readout of a DQD, the regime of small relaxation rate (compared to the drive frequency) is of particular interest, because spin information is conserved. In this regime the signal is purely capacitive and the resonator response \( |\Delta \Gamma| \) to small changes in the device capacitance \( \Delta C_d \) at
2.3 Radio-frequency reflectometry

Figure 2.17: Signal optimisation. (a) Tuning of the matching $\Gamma_{\text{res}}$ and coupling coefficient $\beta$ as a function of voltage applied to a variable capacitor $V_t$. (b) Signal-to-noise ratio as a function of $\beta$. (c) Total capacitance and (d) quality factor as a function of $V_t$. (e) Equivalent resonant circuit including a tunable capacitance $C_t$ and $C_m$. Reproduced from [174].

In contrast to RF SETs, where the signal is maximised by achieving perfect matching and maximising the relative resistive change, see Eq. (2.42), the dispersive signal additionally depends on the internal quality factor as well as the matching and relative change in capacitance. The interplay between minimising the overall capacitance, achieving perfect matching and maximising the quality factor for achieving maximal sensitivity has been demonstrated in an experimental setup that includes variable capacitors (see Fig. 2.17(e) for the equivalent circuit). Variable capacitors allows for in-situ tuning of the matching $\Gamma_{\text{res}}$ and coupling coefficient $\beta$ using a bias.
voltage $V_t$ as shown in Fig. 2.17(a). Maximum signal-to-noise ratio $\text{SNR} \propto |\Delta f|^2$ is achieved slightly below perfect matching (indicated by a dashed line) as shown in Fig. 2.17(b) due to a reduction in total capacitance $C_{\text{Tot}}$ and increase in internal quality factor $Q_0$ when slightly under-coupled $\beta < 1$, as shown in Fig. 2.17(c,d) respectively.

## 2.4 Parametric amplification

To increase the small signal originating from the quantum device above the noise level of room temperature measurement electronics amplification is essential. This section provides an overview on the operating principles and limits of amplification with regard to readout of quantum devices. Further details into (parametric) amplification can be found in topical review articles [175–177] and theses in the field of superconducting qubits [178–180].

### 2.4.1 Amplification and noise

A linear amplifier is an electronic device that takes an input signal and outputs a copy of the signal with increased power. Such amplification by a factor $G$ (amplitude is increased by $\sqrt{G}$) comes at an inevitable cost, meaning that not only the signal (and its noise) power is increased but additionally the amplifier adds noise (the limit is discussed further below).

The process of amplification is illustrated in Fig. 2.18 which shows exemplary power spectra at the amplifier input and output. For an input with signal power $P_{\text{in}}^s$ and associated noise power $P_{\text{in}}^n$, the amplifier outputs a signal with power $G \cdot P_{\text{in}}^s$ and noise $G \cdot (P_{\text{in}}^n + P_{\text{amp}}^n)$. Thus, the signal amplitude is increased at a cost of a reduction in signal-to-noise ratio (SNR) which inherently links noise to amplification. Amplifiers usually have a limited operation bandwidth, where a given gain is achieved within a small tolerance, and can process input signals up to a specific power level, beyond which the linear range of the amplifier is exceeded and reduced gain is observed (compression point).

Typically, noise intensity is characterised as a function of frequency $f$ through the noise power spectral density $S(f)$. In conductors noise is tied to thermal fluctuations of charge carriers at a temperature $T$, known as Johnson-Nyquist or ‘white’ noise [140, 141], with a constant noise power spectral density of $S(f) = k_B T$ (for a
2.4 Parametric amplification

A matched load), such that noise power is related to a noise temperature

$$P_n = k_B T_n B$$

(2.55)

where $B$ is the given bandwidth. The noise temperature is another way of expressing the noise added by an amplifier to the input system noise $P_{\text{out}}^n = Gk_B(T_{\text{sys}} + T_n)B$ and illustrates how the amplifier effectively increases the input noise temperature. For two stages of amplification the output noise power for a given input system noise temperature $T_{\text{sys}}$ and amplifier noise temperature $T_{n1,2}$ is

$$P_{\text{out}}^n = G_2(G_1 k_B [T_{\text{sys}} + T_{n1}] + k_B T_{n2})B.$$  

(2.56)

Consequently, the effective system noise temperature $T_{\text{noise}}$ after a $m$-stage amplification chain is given by [178]

$$T_{\text{noise}} = \frac{P_{\text{out}}^n}{k_B B G_1 G_2 \ldots G_m} = T_{\text{sys}} + T_{n1} + \frac{T_{n2}}{G_1} + \ldots + \frac{T_{nm}}{G_1 G_2 \ldots G_m}.$$  

(2.57)

For large gain $G_i$ such that $T_{ni} \gg T_{ni+1}/G_i$, the noise added by the amplifier chain is dominated by the noise temperature of the first amplifier in the chain. Commercial cryogenic amplifiers typically operate at 4 K and with a noise temperature of a few Kelvin and a gain of 30 dB ($G_{\text{dB}} = 10 \log_{10} G$). Equation (2.57) shows that such gain is sufficient to limit the system noise temperature to a few Kelvin when
performing measurements at room temperature. Moreover, by using an additional cryogenic amplifier of lower noise temperature (operating at a lower temperature stage within a dilution refrigerator), the SNR of cryogenic measurements is further improved. Commercially available cryogenic low noise amplifiers are based on high-electron-mobility transistors (HEMT) with a power consumption of $\sim 20 \text{ mW}$ making operation at lower temperatures infeasible, where cooling power is limited ($< 1 \text{ mW}$). Superconducting devices offer low power dissipation and combined with the non-linear nature of Josephson junctions ultra-low noise amplifiers have been demonstrated, which are discussed in the next section. Alternatively, RF SET devices can be operated as low temperature amplifier, however these are not discussed further [181].

### 2.4.2 Noise limits

In the previous section, the noise power spectral density of a dissipative load was shown to be proportional to temperature, has been introduced. This would suggest that at temperatures approaching zero, the noise intensity is close to zero. Instead, at low temperatures quantum effects arising from the fluctuation-dissipation theorem [175] lead to a non-zero spectral density of

$$S(\omega) = \frac{\hbar \omega}{2} \coth \left( \frac{\hbar \omega}{2k_B T} \right).$$

(2.58)

For $k_B T \gg \hbar \omega$ this expression yields $S(\omega) \approx k_B T$, the classical high-temperature limit, while for $k_B T \ll \hbar \omega$, $S(\omega) \approx \frac{\hbar \omega}{T}$. This describes the zero-point energy of the electromagnetic field and is an example of a quantum mechanical oscillator which is never completely at rest (even at $T = 0$) due to the non-zero ground state energy. Consequently, the effective noise temperature can be different from the physical temperature in the low temperature limit and has a quantum mechanical limit of $T_{\text{eff}} = \frac{\hbar \omega}{2k_B}$. When operating in the sub-1-GHz regime this limit can be very close (e.g. 11 mK at 500 MHz) to the physical base temperature of a dilution refrigerator (10 mK), however, when operating at a few GHz, the minimum system noise temperature can be significantly larger (25 mK/GHz).

These limits on the effective temperature set by quantum fluctuations apply not only to the qubit itself, but also to the amplification process. The limit on amplification can be derived [182, 183] when converting a single classical input mode to the quantum mechanical form $E(t) = \frac{1}{\sqrt{2}} \left( \hat{I} \sin(\omega t) + \hat{Q} \cos(\omega t) \right)$ with quadrature operators $\hat{I}$ and $\hat{Q}$ that obey the canonical commutation relation $[\hat{I}, \hat{Q}] = i$. Similar
relations apply for the magnitude and phase of the signal. In the amplification process a transformation to quadratures with a power gain of \( G \), \( \hat{I} \rightarrow \hat{I}' = \sqrt{G} \hat{I} \) (and similarly for \( \hat{Q} \)) is made. It turns out, that in order to fulfil the commutation relations an additional operator is required in the transformation which represents the amplifier added noise. The variance of this additional operator can be bound to \( 1/2 \), thus, an amplifier adds at least half a photon of noise when measuring one quadrature, which is also encoded in the quantum mechanical uncertainty principle

\[
\sigma_I \sigma_Q \geq \frac{1}{4}. \quad (2.59)
\]

In case of the amplifier, this can be interpreted that in order to produce an amplified copy, the amplifier is required to measure both quadratures. As these are bound to an uncertainty relation, an additional uncertainty (noise) is introduced on the amplified copy.

In this context it makes sense to introduce the distinction between a phase-preserving and phase-sensitive amplifier. An amplifier that amplifies both quadratures of the input signal in the same way preserves the phase in the output signal. For such an amplifier the quantum limit on the variance of \( 1/2 \) applies to each quadrature according to Eq. (2.59). When only one quadrature is amplified the amplifier is phase-sensitive. This can be useful when all the information of a signal is encoded in a single quadrature, such as the homodyne detection scheme. In this case it is possible to reduce the uncertainty on the quadrature of interest below the limit of \( 1/2 \) using squeezing which results in de-amplification of the second quadrature in order to fulfil Eq. 2.59.

### 2.4.3 Ultra-low noise amplification

#### 2.4.3.1 Parametric amplification process

Parametric amplification [184] describes the process of achieving gain by harmonically varying a parameter of the amplifier system. Energy from this modulation, which is usually strong and referred to as pump signal, is transferred onto a weak input signal. The most common example is given by children on swing, who can reach great height by pumping with their legs. Parametric amplification is well known in the field of non-linear optics [185] but has been studied extensively even before at much lower frequency (radio-frequency and microwave) in electronics [186] due to excellent noise properties. It has been shown that phase-preserving para-
A non-linear device or medium is the basis for parametric amplification which is achieved in a wave mixing process where energy from the pump is transferred onto the signal and idler mode.

Physical amplifiers can operate at the quantum limit of added noise [182, 183] and even below for phase-sensitive amplifiers due to squeezing.

Physically, parametric amplification is achieved in a mixing/scattering process, where a weak input signal is mixed with a strong pump signal in a non-linear device or medium as illustrated in Fig. 2.19. Depending on the non-linearity so called three-wave ($\chi^{(2)}$ non-linearity) or four-wave mixing ($\chi^{(3)}$ non-linearity)

\[
\omega_{\text{pump}} = \omega_{\text{signal}} + \omega_{\text{idler}} \tag{2.60}
\]
\[
\omega_{\text{pump}} + \omega_{\text{pump}} = \omega_{\text{signal}} + \omega_{\text{idler}} \tag{2.61}
\]

is achieved, where either one or two pump photons are transferred into a signal and idler photon. The idler is an additional mode which represents energy conservation.

In this context it makes sense to discuss the notion of a degenerate and non-degenerate parametric amplifier which is illustrated in Fig. 2.20. A parametric amplifier is degenerate (in frequency) when the signal and idler modes are degenerate within their bandwidth $\omega_{\text{signal}} \approx \omega_{\text{idler}}$. A non-degenerate amplifier, where signal and idler are separated in frequency, can be realised in a three-wave mixing process. Three-wave mixing produces two resonant modes at which gain is produced (signal and idler) that are typically separated in frequency due to energy conservation. Only when the pump is operated at twice the frequency of the signal, signal and idler become degenerate and a degenerate amplifier is realised. In a four-wave process the amplifier is always degenerate. When operating the pump
at a single frequency, the amplifier is even doubly degenerate as signal, idler and pump modes all coincide. A parametric amplifier can also be spatially degenerate when signal and idler are travelling through the same spatial port or transmission line. Often a four-wave amplifier uses a single port which can result in both spatial and frequency degeneracy. In this case additional components to direct the strong pump signal in order to avoid disturbance of the signal are required. Alternatively, such a degenerate amplifier can be operated using two well balanced pump tones as shown in Fig. 2.20.

All parametric amplifiers can be operated either in phase-preserving or phase-sensitive modes. A parametric amplifier is phase-sensitive when the signal information is distributed over both the signal and idler. A non-degenerate amplifier is necessarily phase-preserving as both signal and idler are separated in frequency. A degenerate amplifier is phase-sensitive when $\omega_{\text{signal}} = \omega_{\text{idler}}$ such that signal and idler are indistinguishable and the signal information is distributed over both modes. In Fig. 2.20 phase-preserving operation of a degenerate amplifier is shown, where the signal (green arrow) is on one side of the gain profile (left of $\omega_0$) which generates a copy of the amplified signal on the opposite side (red arrow) as the idler (right of $\omega_0$). When the signal is exactly at the centre of the resonant mode (directly at $\omega_0$) the amplifier is phase-sensitive.

### 2.4.3.2 Josephson junction devices

The characteristic of dissipation-less transport make superconductors ideal candidates for achieving parametric quantum limited amplification. The Josephson effect [187] in a superconductor-insulator-superconductor tunnel junction (Josephson...
junction) gives rise to a non-linear inductance

\[ L_J = \frac{h/2e}{2\pi I_0 \cos \varphi} \]  

(2.62)

where \( I_0 \) and \( \varphi \) is the superconducting critical current and phase difference across the junction. Parametric amplification [188, 189] and squeezing [190, 191] have been observed in superconducting devices based on Josephson junctions early on, but have not found many applications until the development of superconducting circuits for quantum computation [19]. Readout in experiments using superconducting circuits has been greatly improved [176, 192–203] through further advances in amplifier stability and noise performance.

Degenerate parametric amplifiers have been realised in lumped element non-linear resonators based on junction arrays or with junctions in the superconducting quantum interference device (SQUID) geometry [198, 199]. When flux-driven via a second port connected to a flux line, three-wave mixing (pumped at twice the signal frequency) is realised [201, 204]. An effective four-wave mixing process is achieved when pumped through the signal port (current driven) [197, 204]. Phase-preserving amplification near the quantum limit [197, 205, 206] (100%–64% efficiency) and squeezing [193, 207] has been demonstrated. In a \( \lambda/4 \) resonator shunted by a SQUID [202] a similar four-wave mixing amplifier that is pumped through the signal port has been fabricated. Optimal noise performance in such implementations of a degenerate amplifier is achieved when using two pump tones in current driven mode or by operating using a flux drive which comes at the cost of an additional microwave port [204, 206]. Besides such degenerate amplifiers, a non-degenerate amplifier based on a four junction ring modulator realising three-wave mixing has been demonstrated (Josephson parametric converter, JPC) [200, 208–210], and operation near the quantum limit is achieved and squeezing is observed. A JPC has four microwave ports where pump, signal and idler travel through spatially separated transmission lines. A travelling wave parametric amplifier (TWPA) exhibits a distributed non-linearity and has been demonstrated using a non-linear kinetic inductance material [211] and using a large amount of Josephson junctions (JTWPA) [203, 212]. Compared to non-linear resonator based JPAs, a (J)TWPA achieves amplification at increased bandwidth with higher dynamic range, however, due to losses in such increasingly complex device the quantum limit is often not fully reached. Typical values are 22% and 75% efficiency for the TWPA and JTWPA respectively. Due to the large bandwidth and distributed non-linearity, the classification of (non)degenerate amplification is usually not applied to (J)TWPAs. A
2.4 Parametric amplification

Figure 2.21: Josephson junction based amplifiers. Summary of Josephson junction based amplifiers including schematics, micrographs and main parameters. (a) Lumped-element and (b) $\lambda/4$ based non-linear resonator, (c) Josephson parametric converter, (d) SQUID amplifier and (e) JTWPA. Images reproduced from [197, 202, 203, 210, 218].

theory for generation of squeezed states in a JTWPA has been developed [213]. Finally, another type of microwave amplifier based on a DC current biased SQUID has been demonstrated [214, 215]. To achieve amplification, the SQUID is current biased above the critical current where the input signal modulates the flux through the SQUID at a given flux bias point. Noise performance close to radio-frequency powered JPAs in phase-preserving mode has been achieved [216], however, no squeezing can be performed. A summary of ultra-low noise Josephson junction based amplifiers is given in Fig. 2.21.

Most JPAs operate in the few GHz regime compatible with superconducting circuits for quantum computation, while SQUID amplifiers perform best in the sub-1-GHz regime but operation at higher frequency is possible [217, 218]. Recently, JPAs operating in the sub-1-GHz regime have been demonstrated [219]. While JPAs have superior noise properties due to a parametric amplification process, SQUID amplifiers require fewer microwave components as they are operated in transmission using DC currents. Compared to HEMT amplifiers, JPAs are not as straightforward to operate and require additional circuit components in addition to well-tuned and low noise signals to operate at the quantum limit of noise.


## 2.5 Silicon qubit summary and benchmarks

In this section, an overview of different spin qubit implementations in silicon-based devices including state-of-the-art benchmarks are given. Only the most successful experimental results are shown here and details are presented in tabular form in Appendix E. Table 2.1 summarises different ways to implement spin qubits in silicon including physical mechanisms for control and readout and (dis)advantages.

<table>
<thead>
<tr>
<th>spin qubit state</th>
<th>single qubit gate</th>
<th>two qubit gate</th>
<th>readout methods</th>
<th>+ advantages, − disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>\uparrow\rangle$</td>
<td>$S_z = \pm1/2$</td>
<td>$J, \delta E_z$</td>
<td>SDT, ancilla PSB</td>
</tr>
<tr>
<td>$</td>
<td>\uparrow\uparrow\rangle$</td>
<td>$S_z = 0$</td>
<td>$J, \Delta E_z$</td>
<td>capacitive, exchange</td>
</tr>
<tr>
<td>$</td>
<td>\uparrow\uparrow\uparrow\rangle$</td>
<td>$S_z = 1/2$</td>
<td>exchange, capacitive, exchange</td>
<td>PSB</td>
</tr>
</tbody>
</table>

**Table 2.1: Spin qubit implementations.** Summary of different qubit implementations with number of spins and sub-spaces indicated. Physical mechanisms for control and readout (either spin dependent tunnelling, SDT, or Pauli spin blockade, PSB, based) are compared and (dis)advantages are highlighted.

In most experimental spin qubit implementation encoding of the quantum information in a single-spin 1/2 has been preferred, as long coherence times and high fidelity control are achieved for single spins in silicon. In Fig. 2.22(a) coherence and control fidelities of phosphorus donor and QD based single-spin qubits (either MOS or Si/SiGe) are summarised. Implementations using isotopically purified $^{28}\text{Si}$ (realised in $^{31}\text{P}$ and Si-MOS) show especially long coherence times, while implemen-
tations with a micro-magnet (realised in Si/SiGe) and manipulation using EDSR show increased (one order of magnitude) qubit control speed. In all implementations high-fidelity single qubit operation is typically limited by charge noise and highest fidelity is achieved using pulse optimisation. The fidelity of two-qubit gates still needs to be improved to allow for fault-tolerant operation.

In Fig. 2.22(b) spin-qubit readout using different types of charge sensors and readout methods is summarised. Highest readout fidelities are achieved using Pauli spin blockade. Latched Pauli spin blockade readout provides further improvements in signal-to-noise resulting in increased readout fidelity at shorter integration time. For spin-dependent tunnelling the available integration time is limited by the spin tunnelling rates.

In Fig. 2.22(c) gate-based electron spin readout using a resonator that is capacitively coupled to a DQD structure formed using donors or QDs is summarised. Readout is performed directly (in-situ) based on inter-dot Pauli spin blockade which does not require any additional sensor structures or a reservoir (see Section 2.2.5). High fidelity at a short integration time is obtained in Si/SiGe using an on-chip NbTiN $\lambda/2$ resonator with large quality factor $Q_{\text{load}}$ operating at high frequency $f_0$ (low loss and low parasitics), where the readout fidelity was relaxation limited. Additional parameters to optimise are the resonator matching $\beta$ and gate coupling $\alpha$ as well as noise temperature $T_n$ (limited to $\sim 4$ K in these experiments), see Section 2.3.4.3 & 2.4 and Appendix E.
Figure 2.22: Summary of qubit benchmarks. (a) Qubit control and coherence benchmarks for $^{31}$P (in isotopically purified $^{28}$Si, manipulated using ESR), QDs in Si-MOS (in isotopically purified $^{28}$Si, manipulated using ESR) and Si/SiGe (in natural silicon, manipulated using EDSR). (b) Readout benchmarks for readout using external charge sensors (SET) based on spin-dependent tunnelling (SDT), Pauli spin blockade (PSB) and latched PSB. (c) Readout benchmarks of compact gate-based sensors for a $^{31}$P (using an off-chip resonator based on NbTiN inductor), Si-MOS QD (using an off-chip resonator based on SMD components) and Si/SiGe QD (on-chip NbTiN $\lambda/2$ resonator) device.
Chapter 3

Experimental methods

In this chapter important experimental methods and techniques for fabrication and characterisation of silicon-based nano-scale devices are introduced. This includes a basic description of fabrication processes developed at CEA-Leti for devices presented in this work, to aid the understanding of the device behaviour, and common methods for cryogenic device characterisation.

3.1 Fabrication

A common key objective in fabricating quantum dots is to reach single electron confinement loosely analogous to the natural confinement potential of donor atoms. Due to the reduced effective mass in silicon (compared to GaAs for example) and short range exchange interaction between QDs, fabrication of gate structures of small width and high density is necessary to realise multi-QD structures. Typically a gate width of 50 nm and less with a pitch of the same order is required. In this section, fabrication of such small structures using SOI technology combined with deep-ultra-violet (DUV) lithography and resist trimming is introduced.

3.1.1 SOI nanowire transistors

Fully-depleted silicon-on-insulator (FD-SOI) nanowire technology allows formation of QDs in a transistor (wrap-around gate) geometry. Fabrication processes for such single-electron devices have been developed at CEA-Leti [89, 220–222] on a 300 mm CMOS platform using 193 nm DUV lithography. One of the key achievements to define structures small enough to form QDs using these optical techniques has been resist trimming.

Fabrication is based on a SOI wafer consisting of a 850 μm thick Si substrate, 145 μm thick buried oxide (BOX) and ~10 nm thick Si layer at the surface. The thin
layer and substrate are doped with phosphorus at a concentration of $5 \times 10^{17} \text{ cm}^{-3}$ and $1 \times 10^{15} \text{ cm}^{-3}$ respectively. To form a nanowire transistor device, a nanowire is etched followed by gate deposition and source/drain formation as shown in Fig. 3.1. For each patterning step using DUV lithography, a bottom anti-reflective coating (BARC) and resist is spun onto the SOI wafer. The smallest feature size that can be obtained just using DUV is 80 nm, however, using resist trimming features down to 5 nm can be obtained as illustrated in Fig. 3.1. After the resist (and BARC) is trimmed to the desired width, a nanowire is etched and any resist is removed. A gate is formed using a similar DUV lithography and resist trimming process. The gate stack consists of a thin 0.8 nm layer SiO$_2$, 1.9 nm layer of high-k dielectric HfSiON followed by a 5 nm thick layer of TiN and 50 nm of poly-Si with
an equivalent oxide thickness of 1.3 nm. To fabricate source and drain reservoirs, at first a silicon-nitride spacer is formed followed by silicon epitaxy and low dose doping (LDD). Finally a second spacer and salicidation allows to contact the device using standard tungsten interconnects followed by Cu contact vias. Devices with a nanowire width $W$ and gate length $L_g$ ranging from 10 µm down to 10 nm are fabricated on the same chip. Transmission electron microscope (TEM) images of a silicon nanowire and gate stack (view along the gate) and an image of a contacted device along the source/drain are shown in Fig. 3.2(a,b) respectively.

When operating the nanowire transistor at a positive gate voltage, transport is initiated via channels that form at the top corners of the nanowire due to the enhanced electric field. This corner effect is the basis for formation of QDs. Combined with disorder due to surface roughness and charge traps, such channels turn into localised QDs at cryogenic temperature [222, 223]. Depending on the nanowire width and voltage applied to the Si substrate that acts as a back-gate, a single or a double QD can be formed in a single nanowire. At more positive back-gate voltages, typically a single QD is formed as shown in Fig. 3.2(c), while at more
negative back-gate voltages, double QDs are formed as shown in Fig. 3.2(d). In devices of $W < 60 \text{ nm}$, formation of double QDs with increased disorder and strong coupling is observed and tuning into a single QD using the back-gate is typically not possible.

3.2 Measurement setup

Once a device has been fabricated, connections from the nano-scale device, that is operated at cryogenic temperature, to the control and measurements electronics at room-temperature need to be made. In this section, essential parts of the cryogenic and room-temperature setup for control and readout of silicon quantum devices are presented. Additional information on procedures can be found in Appendix F.

3.2.1 Dilution refrigerator and wiring

The energy scales associated with spin qubits in semiconductor devices are small compared to room-temperature. Consequently, spin qubits need to be operated at cryogenic temperature, typically at milliKelvin temperature in a dilution refrigerator, and spin readout and control is performed using low-power signals. Dilution refrigerators are expensive to operate such that pre-screening of devices is often performed in a cryogenic probe-station or cryostat using a flow of liquid helium or a cryo-cooler based on gas compression and heat-exchange. These systems operate at 4 K unless $^4\text{He}$ gas is pumped to a reduced pressure which allows operation down to 1.6 K. The main experimental results in this thesis have been obtained in a dilution refrigerator at a temperature of 10–50 mK. Multiple system at different locations have been used ranging from a $^4\text{He}$ bath based Kelvinox K400 (Hitachi Cambridge, described in [224, 225]) to Oxford Instruments Triton (UCL, described in [226]; Niels Bohr Institute Copenhagen) and Bluefors LD (Cambridge University, UCL) and XLD (UCL) cryo-free systems.

A summary of a typical dilution refrigerator measurement setup is presented in Fig. 3.3, where small differences in the implementation can occur between the aforementioned setups. In a dilution refrigerator there are different cooling stages, each with an individual expected cooling power, and with the device sitting at the lowest stage that is thermally connected to the mixing chamber of a $^3\text{He}/^4\text{He}$ dilution unit which is able to reach temperatures below 100 mK. In the dilution unit the helium mixture separates into a $^3\text{He}$ rich (normal fluid) and $^3\text{He}$ light phase
(super-fluid) at a temperature \(< 860 \text{ mK}\) and cooling is achieved based on further dilution of the \(^3\text{He}\) light phase via pumping which involves heat absorption [227].

While the device reside at the coldest part of the cryostat, control and measurements electronics is operated at room-temperature (RT), such that wiring requires careful considerations in terms of noise and heating requirements. Good practices on wiring of a dilution refrigerator are given in [228, 229] and in the following a typical setup is discussed. First of all, filtering, thermalisation and attenuation of wiring carrying signals from room-temperature to the device at milliKelvin temperature is required in order to avoid excitation or heating of the device due to high-temperature (black body radiation) and high-frequency noise. Moreover, electrical conductivity and thermal expansion are important material properties to consider. In a dilution refrigerator, there is a distinction between low and high-frequency lines. Low-frequency lines are designed to carry frequencies typically up to 100 kHz and are realised using twisted pairs arrange in a loom made from either constantan or PhBr (low thermal conductivity), which is thermalised and filtered at multiple stages. Filtering can be performed using multi-stage RC/\(\pi\) type filters (100 kHz cut-off, home-made or commercially available from QDevil ApS or Aivon Oy) and/or copper powder [230]/Eccosorb\(^\text{®}\) (> 5 GHz cut-off) at cryogenic temperature and can include further filtering at room-temperature using commercial RC/\(\pi\) type filters. A dilution refrigerator setup typically contains 24–96 low-frequency lines which end in a Fisher type connector at room-temperature and microD at cryogenic temperature carrying bias voltages and transport signals of small currents (pA–nA).

High-frequency wiring in the microwave regime uses semi-rigid coaxial cables that carry signals up to 18 GHz (SMA connectors) or 40 GHz (k-type connectors). To achieve low attenuation at high frequency a large electrical conductivity is desired, however large electrical conductivity usually is accompanied by large thermal conductivity which can lead to a strong heat exchange between stages that are at different temperature. Consequently, different cable materials are chosen depending on the temperature stage and depending on if a line is used for readout or control signal delivery. From 300 K to 4 K stainless steel (SS) or BeCU wiring is typically selected due to the reduced thermal conductivity at the cost of increased losses. For control signal lines Cu or CuNi is chosen at stages below 4 K and for readout lines superconducting Nb(Ti) is preferred for minimal signal loss.

For control and readout of a spin qubit device lines tailored for a specific function can be identified. First of all, there are lines dedicated for radio-frequency (RF) reflectometry readout. For RF readout two lines are required. One attenuated and filtered drive line, that delivers a small modulation signal, and a second line for
Figure 3.3: Fridge wiring. Exemplary dilution refrigerator setup for silicon qubit experiments. Signals are delivered to the device at milliKelvin temperature via low-frequency loom and high-frequency coaxial lines. A typical RF readout setup consisting of a drive and readout line and exemplary pulse and ESR lines for qubit control are shown.
readout, that amplifies the signal reflected by the device. Both signals are combined at milliKelvin temperature using a cryogenic directional coupler. Additionally, a cryogenic circulator might be used to isolate the device from noise originating from the amplifier. The readout line connects to the device via a printed circuit board PCB (that sits in a sample box for shielding and thermalisation), where the readout signal is combined with a low-frequency signal and connects to an LC resonator.

For control, fast pulsing of voltages on gates within a desired voltage parameter space and bandwidth is required. Pulses are delivered using a pulse line, which is typically an attenuated high-frequency line that is combined with a low-frequency line on the PCB using a bias tee. Through such a pulse line setup, voltage offsets and low frequency signals are provided through the low-frequency part and fast pulses are delivered via the microwave line. Typically, voltages pulses for control of silicon based devices only requires signals up to 1 GHz (due to long coherence times). Thus, an alternative approach to a pulse line is a non-attenuated line that uses a material with intrinsically large attenuation at high-frequency such as flexible graphite-CuNi cables [231] (to reduce high-frequency noise). Additional filtering and thermalisation (indicated by filters and 0 dB attenuators) allows to further reduce noise on such a pulse line which can be used to deliver voltage offsets and pulses up to 1 GHz (depending on the filtering setup) using a single line [232]. Finally, there is one additional type of line to deliver control signals to a shorted waveguide for electron spin resonance (ESR). Such an ESR line is typically attenuated and connects to an on-PCB waveguide which can be extended further onto the chip to deliver strong ac magnetic fields.

Exemplary characterisation measurements of microwave components of a typical spin qubit setup are shown in Fig. 3.4. Measurements of the gain of two different cryogenic amplifiers at RT are shown in Fig. 3.4(a) with a typical gain on the order of 30 dB at frequencies < 1 GHz. The characteristic of a 30 dB attenuator, line attenuation and 1 GHz filter (MC-VLF-1000+) at RT typical for a drive line is shown in Fig. 3.4(b). Figure 3.4(c) shows the transmission through a cryogenic coupler (Krytar 158020-810) and circulator (Quinstar QCY-007020UM00, shielded) at RT which demonstrates that the benefits of using a circulator in terms of noise isolation comes at the cost of a reduced measurement bandwidth of 600–800 MHz. A measurement of a complete readout circuit (from drive input to readout output including circulators) at RT and 10 mK is shown in Fig. 3.4(d), where a further reduction in bandwidth is observed at base temperature which could be due to a change in the circulator parameters at low temperature which operates based on ferromagnetic resonances.
Figure 3.4: Microwave components. Characteristics of typical microwave components through transmission measurements at room-temperature (RT) unless stated otherwise. (a) Comparison of the gain of two amplifier models used for reflectometry measurements. (b) 30 dB attenuator, transmission line and filter (MC-VLF-1000+) characteristic of the drive line. (c) Cryogenic coupler and circulator used for reflectometry measurements. (d) Complete reflectometry setup (from drive line input to readout line output including circulators) at RT and 10 mK.
3.2 Measurement setup

3.2.2 PCBs

As already illustrated in Fig. 3.3, the main function of the PCB is to hold the chip with the quantum devices and provide electrical connections to the low and high-frequency signal and measurement lines. The PCB can hold additional components for filtering, combining of signals (bias tee) and LC resonators. Different types of PCBs are used to perform measurements presented in this thesis.

Figure 3.5 shows the different types of PCBs at accurate relative dimensions. The smallest PCB, an exemplary ‘Cambridge’ type PCB, is shown in Fig. 3.5(a), which typically holds up to three high-frequency and < 10 low frequency lines (see [224] for details). Figure 3.5(b) shows a ‘London’ type PCB which holds up to four high-frequency and 24 low frequency lines. These PCB are based on a Rogers 4003 laminate of up to 0.8 mm thickness for improved high-frequency performance (dielectric constant of 3.38 and loss tangent of 0.0021) with tracks made of 1 oz of copper and silver immersion finish for reliable bonding. High-frequency lines are connected via SMP connectors and low-frequency lines use simple molex pin connectors. Thin-film resistors (e.g. Susumu RR0816P-104-B-T5), multi-layer ceramic capacitors (e.g. Vishay VJ0402A101JNAAJ or 0402JR05Z4STR500 for 50 fF) and ceramic core inductors (e.g. Epcos B82498B, Coilcraft 1206CS) are used to form on-PCB bias tees and resonant circuits. A blue LED (Broadcom ASMT-CB00) is used for carrier activation, at an operation voltage of 4 V and a current of 2 µA at milliKelvin temperature, when changing the voltage on the substrate of SOI nanowire devices that acts as a back-gate.

The ‘Copenhagen’ (Qdevil) board shown in Fig. 3.5(c) follows a daughterboard on motherboard approach [233] and holds 48 low frequency lines (microD connector) and 16 high-frequency lines (mini-Coax connector at the back, not shown). The motherboard includes filtering for all low-frequency lines and bias tees for all high-frequency lines. Connections to the daughterboard are made using an interposer and so-called fuzz buttons. The daughterboard holds four multiplexed LC resonators for readout. PCBs are enclosed in or mounted onto (oxygen free) copper boxes or blocks for good thermal contact when connected to the mixing chamber of the dilution refrigerator.

3.2.3 Measurement equipment

Simple transport measurements and leakage tests are performed using a Keithley 2400 source-meter. Low-temperature transport measurements are performed using a current pre-amplifier (SR570, Femto or Ithaco) followed by digitisation using a
digital multimeter (Agilent, HP) or lock-in amplifier (SR830) with a typical gain of 100 nA/V and an integration time of 30 ms or longer. Gate voltages are supplied using a variety of sources depending on the experiment such as HP 32545A universal source, SIM rack (SIM928), decadac, QDAC, BILT source (BE602) or using waveform generators in DC mode. Simple waveforms are supplied using Tektronix AFG3102 and Keysight 33522B series generators. The Yokogawa GS200 is an excellent current source for providing the bias current for a Josephson parametric amplifier.

High-frequency characterisation including reflection and transmission measurements are done using vector-network-analysers (Rohde & Schwarz ZVB, ZVD or Keysight E5071C or Copper Mountain Planar/Cobalt series). Spectrum analysers (Rohde & Schwarz FSV, Agilent CSA) are used to perform noise analysis and charge sensitivity measurements.

### 3.2.4 Reflectometry setup

The cryogenic part of the reflectometry measurement setup consists of a readout and drive line that connect to a readout resonator as introduced in Fig. 3.3. Additionally, room-temperature microwave components are required to implement a homodyne detection scheme based on IQ demodulation as shown in Fig. 3.6.
3.2 Measurement setup

An RF signal at fixed (but tunable) frequency is provided by a signal source which is split using a directional coupler. Most of the signal is guided to the local oscillator (LO) port of the IQ mixer (polyphase microwave), while a small part of the signal (−15 dB) enters the drive line after further attenuation (typically −40 dB). The RF signal reflected by the resonator and delivered via the readout line, is further amplified (typically 60 dB) and connects to the RF input of the IQ mixer. As both the signal on the LO and RF port have the same frequency, a low-frequency (DC) signal (and a signal at twice the RF frequency)\(^1\) is generated at the in-phase (I) and quadrature (Q) port which contains the demodulated RF magnitude and phase information. The IQ signal is further amplified, filtered and acquired using an oscilloscope (Waverunner 44Xi-A, LeCroy DHO4054A, Keysight DSO/MSO) or digitizer (AlazarTech 9462, Spectrum M4i-4421).

To take reflectometry measurements for fast device characterisation, gate voltages are ramped using a waveform generator. The digitizer/oscilloscope is simultaneously triggered using the generator and the I and Q signal is acquired over time and a point in time translates to a specific gate voltage configuration.

3.2.5 Charge sensitivity measurement

Charge sensitivity is a benchmark to assess the performance of a charge sensor. Charge sensors are typically operated at the point of maximum transconductance,

\(^1\)https://www.markimicrowave.com/blog/how-to-think-about-iq-mixers/
such as a sharp edge in a step of conductance in a quantum point contact (QPC) or a Coulomb oscillations in a single-electron device (SET or QD), where a small change in the electrostatic environment leads to a measurable change in the device conductance. When operating charge sensors at radio-frequency signals using an LC resonator, the change in conductance manifests in resonator dampening (due to a change in resistance when connected to the source of a device) or a shift of the resonance (when gate-based, due to a change in the capacitance), see Section 2.3.

Two exemplary single-electron transitions in the phase response of a gate-based sensor are shown in Fig. 3.7(a). The charge sensitivity in the RF domain is measured by applying a sinusoidal modulation onto the gate that biases the device at the point of maximum transconductance, which causes an amplitude modulation of the reflected RF carrier \( f_{RF} \) at the modulation frequency \( f_{AC} \). Assuming a purely linear response, two side-bands around the RF carrier are generated

\[
A_1(t) = A_{RF} \sin(2\pi f_{RF} t) \\
A_2(t) = A_{AC} \sin(2\pi f_{AC} t) \\
A(t) = [1 + A_2(t)]A_1(t) = [1 + A_{AC} \sin(2\pi f_{AC} t)]A_{RF} \sin(2\pi f_{RF} t) \\
A(t) = A_{RF} \sin(2\pi f_{RF} t) + \frac{A_{RF} A_{AC}}{2} \left[ \sin(2\pi [f_{RF} + A_{AC}] t + \phi) + \sin(2\pi [f_{RF} - A_{AC}] t - \phi) \right]
\]

where \( A_1(t) \) is the RF drive, \( A_2(t) \) is the modulation and \( A(t) \) is the resulting amplitude modulated wave. To obtain the charge sensitivity, the amplitude of the modulation is calibrated to a charge equivalent root-mean-square amplitude \( \Delta q_{\text{rms}} \), typically 0.01\( e \) or less, and the height of side-bands appearing at \( f_{RF} \pm f_{AC} \) in the frequency spectrum is determined, which represents the signal-to-noise ratio (SNR) of the charge sensitivity measurement. From this the charge sensitivity is obtained as

\[
\delta q = \Delta q_{\text{rms}} / (\sqrt{2BW} \times 10^{\frac{3\text{SNR}}{20}}),
\]

where \( BW \) is the spectrum analyser resolution bandwidth.

In an experiment, the RF carrier and gate voltage are combined at milliKelvin temperature using a bias tee, while the constant gate voltage offset is joined with the sinusoidal modulation at room temperature using a voltage divider and adder box as shown in Fig. 3.7(b). The purpose of the divider and adder box is to provide the required voltage resolution and stability in addition to joining of both signals. A large division is desired at the modulation (AC) input (of 5000–100) compared to a small division (of \( \approx 10 \)) at the voltage offset (DC). Simple resistor and capacitor
3.2 Measurement setup

**Figure 3.7:** Charge sensitivity measurement. (a) Single electron transitions in the phase response used to perform a charge sensitivity measurement. (b) Experimental setup for a charge sensitivity measurement. (c) Frequency spectrum close to $f_{RF}$ as a function of gate voltage offset showing side-bands appearing at the points of maximum transconductance. (d) Calibration of the AC amplitude by splitting of a transition as a function of frequency $f_{AC}$. (e) Line cuts of the spectrum in (c) at the points indicated in blue and orange showing the height of the side-bands corresponding to the SNR.
based circuits provide this functionality [228]. A resistor-only based circuit, as shown in Fig. 3.7(b), provides a simple frequency independent solution with the advantage that both the DC and AC input contribute to the overall offset voltage. The AC input offset can be used for fine tuning of the offset and the division factors are given by $d_{DC} = \frac{R_1 + R_2}{R_2}$ and $d_{AC} = \frac{R_3 + R_2}{R_2}$. A circuit that includes a capacitor provides a frequency dependent solution with useful noise filtering functions, such as filtering of high frequency noise and quick steps in the DC offset voltage at the cost of not being able to fine tune the gate offset using a high division AC offset. The division factors for such a circuit, which is also shown in Fig. 3.7(b), are given by $d_{DC} = \frac{R_1 + R_2}{R_2}$ and $d_{AC} = \frac{R_3 + R_4}{R_4}$ and the frequency response can be estimated from the $RC$ low and high-pass. In charge sensitivity experiments used throughout this thesis, the resistor based circuit has been implemented due to the additional offset fine tuning, combined with low-pass filtering of the DC input prior to the divider circuit for increased stability.

Figure 3.7(c) shows the appearance of side-bands at the points of maximum transconductance in the frequency spectrum. The exact modulation signal amplitude at a given frequency is obtained by applying an amplitude large enough to observe broadening and splitting of the selected transition as shown in Fig. 3.7(d), which accounts for any signal filtering along the circuit and lines. Based on this, the modulation amplitude is calibrated to a charge equivalent amplitude from the separation in voltage of the selected and next transition at higher voltage, which corresponds to the addition of a single electron charge as shown in Fig. 3.7(a). After calibration of the modulation amplitude, the SNR is obtained from the height of the side-bands as shown in Fig. 3.7(e) and typically an optimisation as a function of AC amplitude, RF power and frequency is performed to obtain the device specific maximum charge sensitivity that is calibrated to the gate charge.
Chapter 4

Fast and sensitive gate-based
RF-readout

A qubit readout protocol is an essential ingredient for all quantum computing technologies. In this chapter, two significant improvements in gate-based readout techniques are presented based on a parallel high-quality resonant circuit and Josephson parametric amplifier. A charge sensitivity of $0.25 \mu e/\sqrt{Hz}$ is demonstrated and spin readout above fault-tolerant threshold within 1 $\mu$s is expected. These results pave the way towards compact high-fidelity readout of spin qubits in silicon. Most relevant concepts for this chapter are introduced in Section 2.3 and 2.4. A study of an alternative non-scalable readout method using external charge sensors is presented in Appendix A.

Measurements have been performed at UCL in a Oxford Instruments Triton, Bluefors XLD and LD system using ‘London’ type PCBs under supervision of Prof J. J. L. Morton. The author acknowledges contributions from Dr I. Ahmed, Dr J. A. Haigh and Dr M. F. Gonzalez-Zalba for the development of the parallel resonator design and fabrication of NbN spiral inductors. Dr J. W. A. Robinson and his group for sputtering of NbN films and Dr Irfan Siddiqi and his group for developing and providing a low-frequency Josephson parametric amplifier and CEA-Leti for fabrication of nanowire transistor devices.

4.1 Readout fidelity

In Section 2.3.1, prospects of radio-frequency techniques towards sensitive and fast readout were introduced. To motivate further sensitivity improvements, the concept of readout fidelity is discussed which combines the signal-to-noise ratio (SNR) for a given integration time of a measurement with the relaxation and coherence requirements of the qubit system and characterises readout performance using a
single number. For example, a measurement with \( \text{SNR} \gg 1 \) that is performed only within an integration time \( \tau_{\text{int}} \) close to the relaxation time \( T_1 \) of the qubit does not represent a high-fidelity measurement: for many measurements, the information is lost during the measurement time and false readings are produced, representing low readout fidelity. A model for readout fidelity has been introduced in [234] and is discussed at the example of capacitive gate-based readout, where a signal difference between two states (i.e. singlet and triplet states) \( \Delta \phi = \phi_S - \phi_T \) provides the measurement signal amplitude, that is subject to noise of amplitude \( \sigma \). The SNR of such measurement is defined as the ratio of the signal and noise power

\[
\text{SNR} = \frac{P_{\text{signal}}}{P_{\text{noise}}} = \left( \frac{\Delta \phi}{\sigma} \right)^2.
\]

Assuming Gaussian noise, the singlet and triplet probability density for an experiment consisting of \( N_{\text{total}} \) initialisation and single-shot readout sequences is modelled as two Gaussian distributions separated by \( \Delta \phi \) and broadened by the measurement noise \( \sigma \) with a histogram given by

\[
N(\Delta \phi) = N_{\text{total}}((1 - p_T)P_S(\Delta \phi) + p_T P_T(\Delta \phi)) \Delta \phi_{\text{binsize}}.
\] (4.1)

Here \( p_T \) is the triplet probability over all outcomes and \( \phi_{\text{binsize}} \) is the histogram bin size. When the singlet is the ground state, the probability density is given by a simple Gaussian

\[
P_S(\Delta \phi) = \frac{1}{\sqrt{2\pi} \sigma} \exp \left( -\frac{(\Delta \phi - \phi_S)^2}{2\sigma^2} \right),
\] (4.2)

while the triplet probability includes additional terms accounting for relaxation during the integration time \( t_{\text{int}} \) given a relaxation time \( T_1 \) of the triplet excited state to the singlet ground state (also known as Pauli spin blockade time)

\[
P_T(\Delta \phi) = \frac{1}{\sqrt{2\pi} \sigma} \left[ e^{-t_{\text{int}}/T_1} \exp \left( -\frac{(\Delta \phi - \phi_T)^2}{2\sigma^2} \right) \right. \]
\[+ \left. \int_{\phi_S}^{\phi_T} \frac{t_{\text{int}}}{T_1(\phi_S - \phi_T)} \exp \left( -\frac{t_{\text{int}}}{T_1} \frac{\phi - \phi_S}{\phi_S - \phi_T} \right) \exp \left( -\frac{(\Delta \phi - \phi)^2}{2\sigma^2} \right) d\phi \right].
\] (4.3)

When both probability distributions are well separated (multiple \( \sigma \)), both states can be identified with unit fidelity by defining a threshold centred between \( \phi_S \) and \( \phi_T \), where for a measurement signal amplitude below threshold, a singlet is identified and above threshold, a triplet. However, when the distributions have non-zero overlap, measurements can be misidentified and the optimal threshold is not necessarily in the centre. The singlet and triplet fidelity is obtained by integrating over the
misidentified measurements given a threshold $\phi_{\text{th}}$

$$F_S = 1 - \int_{\phi_{\text{th}}}^{\infty} P_S(\Delta \phi) d\Delta \phi \quad F_T = 1 - \int_{-\infty}^{\phi_{\text{th}}} P_T(\Delta \phi) d\Delta \phi$$

and the average fidelity and visibility is defined as

$$F_{\text{avg}} = \frac{F_S}{2} + \frac{F_T}{2} \quad V = F_S + F_T - 1.$$  \hfill (4.6)

Example probability distributions $P_{S/T}$ are shown in Fig. 4.1(a) for SNR = 5 = $\frac{T_1}{\Delta t_{\text{int}}}$. In case of the triplet, there is a non-zero probability to measure a signal corresponding to a singlet due to relaxation. Consequently, the optimal threshold $\phi_{\text{th}}$ (dashed vertical line) is found slightly closer to the singlet rather than in the centre between the singlet and triplet distribution and a visibility $V = 0.95$ is obtained in this example. Using the fidelity model, limits on how fast readout should be performed can be estimated for a given relaxation time $T_1$. Moreover, limits on the required sensitivity of the sensor can be identified. First the measurement time $t_{\text{min}}$ is defined to be the minimum time required to perform a measurement $t_{\text{min}}$, at which SNR = 1 is obtained. In such, $t_{\text{min}}$ is a measure of readout sensitivity and depends on the signal and noise level obtained in an experiment. For an integration time $t_{\text{int}}$ that is longer than $t_{\text{min}}$ the SNR improves linearly, assuming averaging over uncorrelated noise, due to a decrease in noise amplitude as $\sigma = \sqrt{\frac{t_{\text{min}}}{t_{\text{int}}}}$. Figure 4.1(b)
Chapter 4  Fast and sensitive gate-based RF-readout

shows the visibility $V$ as a function of $t_{\text{int}}$ for a given $t_{\text{min}}$ and $T_1$. For $T_1 = 10 \text{ ms}$, similar to what is observed for two spins in a silicon DQD (see Section 2.5), and $t_{\text{min}} = 0.1 \mu\text{s}$, a visibility $> 99\%$ is obtained between $2.5 \mu\text{s} < t_{\text{int}} < 400 \mu\text{s}$ as shown in Fig. 4.1(b). At shorter $t_{\text{int}}$ the fidelity is sensitivity-limited, while for large $t_{\text{int}}$ fidelity is relaxation-limited. When $t_{\text{min}}$ is increased by one order of magnitude or $T_1$ is decreased by one order of magnitude the window for $t_{\text{int}}$ at which high-fidelity readout is obtained reduces roughly proportionally as shown in Fig. 4.1(b). Overall, high-fidelity readout ($> 99\%$) is approximately obtained for $25t_{\text{min}} < t_{\text{int}} < T_1/25$. Consequently, assuming that $T_1 = 10 \text{ ms}$, a detector sensitivity that achieves $t_{\text{min}} < 16 \mu\text{s}$ is required in order for high-fidelity readout $> 99\%$ to be possible. At larger values of $t_{\text{min}}$ the region of sensitivity-limited readout starts to overlap with the region of relaxation-limited readout making high-fidelity measurements impossible. Finally, to implement error correction codes, the readout must be faster than the decoherence time $T_2$ rather than the relaxation time $T_1$. Without continuous spin-refocussing (such as Hahn-echo or CPMG sequences) decoherence is limited to $T_2^* = 20–120 \mu\text{s}$ [41, 91]. Assuming 100 $\mu\text{s}$, this yields $t_{\text{min}} < 160 \text{ ns}$, highlighting the importance of fast readout.

In the following sections, methods to achieve fast readout with high-fidelity ($> 99\%$) using gate-based reflectometry are presented based on reducing the minimum measurement time $t_{\text{min}}$. Specifically, the signal power is improved using a high-quality parallel resonant circuit, that increases the absolute change of the reflection coefficient $|\Delta \Gamma|$ for a given change in device capacitance, and the noise power $P_n$ is reduced using a Josephson parametric amplifier. Both leads to an increase in SNR as

$$\text{SNR} = \frac{|\Delta \Gamma|^2 P_{RF}}{P_n}$$

and a reduction in $t_{\text{min}}$.

4.2 Parallel high-$Q$ resonator

As introduced in Section 2.3, gate-based reflectometry was first demonstrated using resonant circuits that have been developed for radio-frequency single-electron transistors. For an RF-SET, the goal is to match the resistance $R_d$ of the SET, which is on the order of $100 \text{ k}\Omega$, to the line-impedance of $50 \Omega$. This is typically achieved using a surface mount inductor $L$ of $80–400 \text{ n}H$ combined with the parasitic capacitance $C_0$ of $0.3–0.8 \text{ pF}$ (typically not tunable and given by PCB and device
4.2 Parallel high-$Q$ resonator

parasitics), and yields a resonator with resonance frequency of 200–800 MHz, effective resistance $R_{\text{eff}} = \frac{L}{C_0 R_d}$ and quality factor of $\sim 30$ [162, 163, 235]. For gate-based readout, the gate-resistance is on the order of 300 kΩ or larger [162]. Moreover, both good matching and also a large internal quality factor are important for achieving a large signal, as shown in Eq. (2.53). In order to match the resonant circuit to a larger resistance, an inductance of 1–10 µH would be necessary. However, this in turn reduces the internal quality factor $Q_{\text{int}} = R_d \sqrt{C_0}$, demonstrating that a simple resonant circuit consisting of an inductor, that directly connects the RF signal from the coaxial line onto the device gate, and the parasitic capacitance in parallel, as shown in Fig. 2.12(c), is not well suited for optimising gate-based readout sensitivity. In the following, an improved resonant circuit is presented that combines a coupling capacitor and high-quality NbN spiral inductor in a parallel resonator circuit, whose effective component values allow to form well-matched high-quality resonators using similar physical values of $L, C$ and $R$.

4.2.1 Circuit and simulations

Circuit In this section, the new parallel resonator configuration is discussed including simulations and benefits for gate-based readout. In contrast to the series configuration [162, 163, 235], as shown in Fig. 2.12(c), in the parallel configuration [236] the inductor $L$ is connected in parallel with the quantum device, modelled via $R_d$ and $C_d$ (state-dependent), and parasitic capacitance $C_p$ and coupled to the coaxial line using a coupling capacitor $C_c$, as shown in Fig. 4.2(a). By defining the variable capacitance $C_0 = C_p + C_d$ a capacitively coupled parallel $LR_dC_0$ resonator can be identified. The impedance $Z$ of this circuit is given by

$$Z = \frac{1}{i \omega C_c} + \left( \frac{1}{R_d} + \frac{1}{i \omega L} + i \omega C_0 \right)^{-1}$$

(4.8)

$$= \frac{1}{i \omega C_c} + \frac{i \omega L}{1 + \frac{i \omega L}{R_d} - \omega^2 L C_0}$$

(4.9)
which determines the reflection coefficient $\Gamma = \frac{Z - Z_0}{Z + Z_0}$. The resonance frequency $\omega_r$ and quality factors of such a circuit are [236]

\[
\omega_r \approx \frac{1}{\sqrt{L(C_0 + C_c)}}, \quad (4.10)
\]

\[
Q_{\text{ext}} = \frac{C_0 + C_c}{C_c^2 Z_0} \sqrt{L(C_0 + C_c)}, \quad (4.11)
\]

\[
Q_{\text{int}} = \frac{\sqrt{C_0 + C_c}}{L} R_d. \quad (4.12)
\]

The real and imaginary part of the impedance at resonance for typical experimental parameters, for which $L/R_d^2 C_0 \ll 1$ and $C_c < C_0$ is fulfilled, are

\[
\text{Re}[Z(\omega_r)] = \frac{L}{R_d(C_0 + C_c)} \left( 1 - \frac{C_0}{C_0 + C_c} \right)^2 \approx \frac{L}{C_c^2 R_d} (C_0 + C_c), \quad (4.13)
\]

\[
\text{Im}[Z(\omega_r)] = -\frac{1}{\sqrt{C_c/(C_0 + C_c)}} \left( 1 - \frac{C_0}{C_0 + C_c} \right)^2 + \frac{\sqrt{L/C_c}}{\sqrt{C_c/(C_0 + C_c)}} \left( 1 - \frac{C_0}{C_0 + C_c} \right)^2, \quad (4.14)
\]

\[
\approx \frac{1}{\omega_r} \left( \frac{C_0 + C_c}{C_c} \omega_r^2 - \frac{1}{C_c} \right) \quad (4.15)
\]
which yields an equivalent impedance close to resonance at \( \omega = \omega_r + \Delta \omega \) of

\[
Z_{eq} = \frac{L}{C^2 R_d} (C_0 + C_c) + \frac{i}{\omega} \left( \frac{C_0 + C_c}{C_c} L \omega^2 - \frac{1}{C_c} \right) \tag{4.16}
\]

\[
\approx R_{eff} + \frac{i}{\omega_r} \left( \frac{C_0 + C_c}{C_c} 2 L \omega_r \Delta \omega \right) \tag{4.17}
\]

\[
= R_{eff} + i 2 \sqrt{(C_0 + C_c)L / C_c} \frac{\Delta \omega}{\omega_r} \tag{4.18}
\]

\[
= R_{eff} + i 2 \sqrt{L_{eff} / C_c} \frac{\Delta \omega}{\omega_r}. \tag{4.19}
\]

This shows that the parallel circuit can be approximated by an equivalent \( L_{eff} R_{eff} C_c \) series resonant circuit at resonance as shown in Fig. 4.2(b). The effective values provide an understanding on how the resonance is affected by changes in physical circuit parameters. The expression for \( R_{eff} \) shows that \( C_c \) allows tuning into matching \( (R_{eff} \approx 50 \, \Omega) \), independently of \( C_0 \) and \( L \). Moreover, the parallel configuration achieves an effectively larger inductance (when \( C_0 > C_c \)), while the internal quality factor still scales with the inverse of the physical value \( \sqrt{L} \). This analysis of the effective values for a series circuit illustrate how the parallel circuit provides a way to achieve good matching combined with improved quality factor using similar physical values of \( C_0, L \) and \( R_d \). This has been demonstrated in an experiment [170], where good matching combined with an improved loaded quality factor of \( Q_{load} = 141 \) is achieved compared to a quality factor of 30 in previous experiments that use a series resonator configuration with similar surface mount components [162, 163]. The loaded quality factor is given by \( Q_{load} = \frac{f_r}{\Delta f} = \left( \frac{1}{Q_{int}} + \frac{1}{Q_{ext}} \right)^{-1} \), with \( \Delta f \) being the resonator (readout) bandwidth, such that \( Q_{load} \) is equal to \( \frac{Q_{int}}{2} = \frac{Q_{ext}}{2} \) when matched. A simulation using the experimental parameters is shown in Fig. 4.2(c).

**Simulations** The resonator response is determined by the absolute differential change in the reflection coefficient \( |\Delta \Gamma| \) arising from changes in \( C_0 \). In the limit of a small signal \( (Q_{load} \Delta C/(C_0 + C_c) \ll 1) \) \( |\Delta \Gamma| \) is given by [236]

\[
|\Delta \Gamma| = \left| \frac{\partial \Gamma}{\partial C_0} \Delta C_0 \right| = \frac{2 R_{eff} Z_0}{(R_{eff} + Z_0)^2} Q_{int} \frac{\Delta C_0}{C_0 + C_c} \tag{4.20}
\]

\[
= \frac{1}{2} \frac{\beta}{(1 + \beta)^2} Q_{int} \frac{\Delta C_0}{C_0 + C_c} \tag{4.21}
\]

for this parallel circuit. Fig. 4.3 shows the dependence of \( |\Delta \Gamma| \) on physical parameters using circuit simulations based on typical values of \( L = 150 \, \text{nH}, \; C_0 = 0.4 \, \text{pF} \).
and $R_d = 1 \, \text{M}\Omega$ to illustrate guidelines on how to maximise the signal reflected by the resonator. Using small values ($< 50 \, \text{fF}$) of $C_c$ the resonator can be tuned into perfect matching as shown in Fig. 4.3(a), where $|\Delta \Gamma|$ reaches a maximum. Figures 4.3(b-e) show that the reflected signal $|\Delta \Gamma|$ is maximised by reducing dielectric losses (represented by $1/R_d$) and capacitance ($C_0$) of the resonant circuit and follows the expected trend of $|\Delta \Gamma| \propto R_d C_0 / \sqrt{L(C_0 + C_c)}$ when perfectly matched. Furthermore, the value of $C_c$ required for perfect matching decreases with increasing $R_d$ and decreasing $C_0$. The internal quality factor $Q_{\text{int}}$ increases with $C_0$ ($Q_{\text{int}} \propto \sqrt{C_0}$), while $|\Delta \Gamma|$ decreases as $C_0$ is increased ($|\Delta \Gamma| \propto 1/\sqrt{C_0}$). Finally, $Q_{\text{int}}$ increases linearly with $R_d$. Overall, a large signal is achieved by reducing the resonator capacitance $C_0$, achieving high $Q_{\text{int}}$ by reducing dielectric losses and by tuning into matching using $C_c$. These guidelines and the interplay between experimentally tuning some of these parameters is confirmed in an experiment where maximum signal is achieved close to matching and by reducing the overall capacitance [174] (see Fig. 2.17). Following these guidelines, where the capacitance is reduced more than the inductance, a large resonator impedance $Z = \sqrt{L/(C_0 + C_c)} > 50 \, \Omega$ (high-impedance resonator) is achieved.

**Power at the device** The power delivered at the device is an important parameter as the SNR scales linearly with input power (see Eq. (4.7)) up to a limit where power broadening reduces the reflected signal which depends on the transition that is sensed. Besides parameters of the resonator itself, the gate coupling $\alpha$ to the device plays an important role as $\alpha$ not only determines the magnitude of $\Delta C_0$ but also sets the maximum RF amplitude to be delivered at the gate. For an inter-dot charge transition the quantum capacitance related to the tunnelling of a single electron in the device is $\Delta C_0 = \alpha^2 e^2 / 2 \Delta \c$, see Eq. (2.49). To avoid power broadening, the energy equivalent of the RF amplitude at the device gate should not exceed the tunnel coupling $v_{\text{gate}} < \Delta e / \hbar e$. Based on the circuit in Fig. 4.2(a), $v_{\text{gate}}$ can be expressed as a function of the input power $P_0 = \frac{v_{\text{in}}^2}{2Z_0}$ and voltage $v_{\text{in}}$ at the coaxial line to determine the maximum resonator input power below broadening. Considering the fraction of input power that is transferred to the device $1 - |\Gamma|^2$
4.2 Parallel high-\(Q\) resonator

\[ C_0 = R_d = \Gamma_{\text{maxmax}} \]

\[ C_0 = R_d = fF \]

Figure 4.3: Resonator optimisation. (a) \(|\Delta \Gamma|\) and reflection coefficient at resonance \(\Gamma_{\text{min}}\) as a function of coupling capacitor. \(C_c\) tunes the matching and \(|\Delta \Gamma|\) is maximised at perfect matching (here: \(C_c = 35\) fF). \(|\Delta \Gamma|\) as a function of \(C_c\) for different values of (b) \(C_0\) and (c) \(R_d\). \(|\Delta \Gamma|_{\text{max}}\) (at perfect matching) as a function of (d) \(C_0\) and (e) \(R_d\). The \(C_c\) required to achieve perfect matching and the internal quality factor are shown at the top and right axis respectively. Largest \(|\Delta \Gamma|_{\text{max}}\) is obtained for small \(C_0\) and large \(R_d\) combined with small \(C_c < 50\) fF for perfect matching. For these simulations \(L = 150\, \text{nH}\), \(C_0 = 0.4\, \text{pF}\) and \(R_d = 1\, \text{M\Omega}\) unless varied and a capacitive shift of \(\Delta C_0 = 1\, \text{fF}\) is assumed.
and neglecting any further losses this yields

\[
\frac{v_{\text{gate}}^2}{2R_d} = (1 - |\Gamma|) \frac{v_{\text{in}}^2}{2Z_0} \tag{4.22}
\]

\[
v_{\text{gate}}^2 = \frac{R_d}{Z_0} \frac{4R_{\text{eff}} Z_0}{(R_{\text{eff}} + Z_0)^2} v_{\text{in}}^2
\]

\[
= \frac{R_d}{Z_0} \frac{\beta}{(1 + \beta)^2} v_{\text{in}}^2 \tag{4.24}
\]

\[
= \frac{C_c^2}{(C_0 + C_c)^2} Q_{\text{load}}^2 v_{\text{in}}^2 \tag{4.25}
\]

with \( \beta = \frac{Z_0}{R_{\text{eff}}} = \frac{Q_{\text{int}}}{Q_{\text{ext}}} \). This relates the input power \( P_0 \) to \( v_{\text{gate}} \) as follows

\[
P_0 = \frac{v_{\text{in}}^2}{2Z_0} = \frac{(C_0 + C_c)^2}{2Z_0 C_c^2 Q_{\text{load}}^2} v_{\text{gate}}^2, \tag{4.26}
\]

where the maximum gate voltage below power broadening limits the maximum input power. Assuming operation below power broadening (\( v_{\text{gate}} = \frac{\Delta \alpha e}{2} \)) and substituting \( R_{\text{eff}} = \frac{L}{C_c R_d} (C_0 + C_c) \) and \( Q_{\text{int}} = (1 + \beta)Q_{\text{load}} \) further simplifications can be made

\[
P_0 = \frac{1}{2} (1 + \beta)^2 \frac{\omega_r}{Q_{\text{int}}} \frac{(C_0 + C_c) \Delta \alpha^2}{\alpha^2 e^2}, \tag{4.27}
\]

which reveals a scaling of \( P_0 \) with the inverse of \( Q_{\text{int}} \) and proportional to \( \omega_r \). Hence, a higher \( Q_{\text{int}} \) reduces the maximum input power. A low input power level is a key ingredient for further signal enhancement using a JPA as demonstrated in Section 4.3. Substituting 4.21 (with \( \Delta C_0 = \frac{\alpha^2 e^2}{2 \Delta \alpha} \)) and 4.27 into 4.7, the SNR for gate-based sensing of an inter-dot transition is obtained

\[
\text{SNR} = \frac{|\Delta \Gamma|^2 P_0}{P_n} = \frac{e^2}{8 P_n} \frac{\beta}{(1 + \beta)^2} \frac{Q_{\text{int}} \omega_r}{C_0 + C_c} \alpha^2, \tag{4.28}
\]

which is proportional to \( Q_{\text{int}} \) because \(|\Delta \Gamma|^2 \propto Q_{\text{int}}^2 \) and \( P_0 \propto Q_{\text{int}}^{-1} \).

**Minimum integration time** Given that \( P_n = k_B T_n / t_{\text{int}} \) with \( T_n \) being the system noise temperature (typically amplifier limited), the minimum integration time to achieve \( \text{SNR} = 1 \) is

\[
t_{\text{min}} = \frac{8 k_B}{e^2} \frac{(1 + \beta)^2}{\beta} \frac{C_0 + C_c}{Q_{\text{int}} \omega_r} \frac{T_n}{\alpha^2}. \tag{4.29}
\]
which summarises all possible ways to improve readout sensitivity in one simple expression. Paths to maximise sensitivity range from the resonator and microwave circuit level (by maximising $Q_{\text{int}}$, minimising total capacitance and $T_n$, and achieving good matching), to the device level (by maximising $\alpha$). Moreover, a shorter $t_{\text{min}}$ is obtained by increasing the resonance frequency $\omega_r$. Additionally, when increasing $\omega_r$ and maintaining the same loaded quality factor a larger readout bandwidth can be obtained. However, signal losses can be larger at higher frequency for an approach where the resonator is fabricated separately from the quantum device, due to high-frequency losses in the inter-connection between resonator and device. Moreover, the resonance frequency provides a lower limit for tunnelling rates that can be detected (see Section 2.3.4). Overall, operation close to 1 GHz or up to 2 GHz is expected to provide a good compromise. At matching Eq. (4.29) can also be expressed in terms of $C_c$ and $Z_0$: $t_{\text{min}} = \frac{32k_B}{e^2} C_c^2 Z_0 T_n$. However, such expression is less intuitive as it doesn’t illustrate the interplay between the resonator parameters such as $C_0$, $R_d$ and $Q_{\text{int}}$ as captured in Fig. 4.3 and Eq. (4.29), but valid as the value of $C_c$ required for matching reduces as $C_0$ is reduced and $Q_{\text{int}}$ is increased as shown in Fig. 4.3(d-e).

In capacitive gate-based sensing, as discussed in this chapter, the change in reflection coefficient $|\Delta \Gamma|$ is measured through a change in the phase of a signal at fixed frequency close to resonance. When $Q_{\text{load}} \Delta C/(C_0 + C_c)$ is increased from values smaller than 1, typical for approaches where the resonator is fabricated separately from the quantum device, to values close to or larger than 1, that can be achieved using on-chip resonators, the strategy to achieve maximum phase signal differs. The magnitude of the phase change can then typically be further increased by operating in the over-coupled regime (where an additional $180^\circ$ phase shift is obtained and $\Delta \phi = -2Q_{\text{load}} \frac{\Delta C}{C_0 + C_c}$) and there is an optimal $Q_{\text{load}}$ for a given fractional change in capacitance [237]. This can be derived in the picture of circuit quantum electrodynamics, which is summarised in Appendix B and provides a more general approach that yields Eq. (4.21) in the limit $Q_{\text{load}} \Delta C/(C_0 + C_c) \ll 1$.

### 4.2.2 Setup and implementation

In the previous section, a parallel resonant circuit design was motivated along with strategies to optimise the reflectometry signal. Following Eq. (4.29), implementation of a circuit with reduced parasitic capacitance $C_0$ and increased internal quality factor $Q_{\text{int}}$ is presented using an optimised PCB design and a low-loss (large $R_d$) superconducting NbN on sapphire spiral inductor. NbN inductors were developed
and fabricated by means of photolithography and reactive ion etching based on films deposited using DC sputtering. Depending on the number of turns, an inductance ranging from 30–600 nH is achieved. Implementation of a well-matched resonator of high quality factor operating in the 600–800 MHz band is presented for compatibility with a Josephson parametric amplifier.

Figure 4.4(a,b) show the PCB design consisting of multiple low frequency lines and one high-frequency line with coplanar waveguide carrying the radio-frequency readout signal. The waveguide is matched to 50 Ω minimised in length for reduced parasitic capacitance. Further reduction in $C_p$ is achieved by removing metal in selected areas of the bottom layer. The PCB allows back-gating of the QD chip and offers space for an inductor chip. An example NbN spiral inductor with 8 µm feature size and 26 turns is shown in Fig. 4.4(c) and the fully assembled PCB including components for filtering, inductor and QD chip is shown in Fig. 4.4(d). The gate voltage is delivered onto the device through the inductor.

As demonstrated using simulations in Fig. 4.3, selection of an appropriate coupling capacitor $C_c$ is important to achieve good matching and this depends on the parasitic capacitance and losses of the PCB and quantum device. The resonator is first characterised at 4 K without a quantum device. Figure 4.5(a) shows the reflection coefficient for different combinations of $C_c$ and $L$. For $C_c = 218$ fF and $L = 100$ nH, the resonator is within the desired frequency but strongly over-coupled and has a low $Q_{load}$. A custom made microstrip gap capacitor with $C_c \approx 40$ fF
4.2 Parallel high-\(Q\) resonator

![Fig 4.5: Resonator optimisation for 0.6–0.8 GHz. (a) Resonance at 4 K without QD device connected for different values of \(C_c\) and \(L\). (b) Schematic and image of the resonant circuit with inductor and QD chip shown. (c) Final resonance with the QD device connected at 30 mK.](image-url)
achieves a still over-coupled resonator with high $Q_{\text{load}}$ but outside the desired frequency range. Using $L = 170$ nH an under-coupled resonance in the desired range with acceptable $Q_{\text{load}}$ is formed. Next, a quantum device is connected as shown in Fig. 4.5(c) and an over-coupled resonance at 616 MHz with $Q_{\text{load}} = 790$ is achieved. From these measurements a contribution to the geometric parasitic capacitance of 235 fF from the PCB and inductor and 118 fF for the QD chip is obtained. $R_d$ and $C_c$ are calculated from $Q_{\text{load}}$ and $\beta = \frac{Z_0}{R_{\text{eff}}}$, see Eq. (2.40, 4.19, 4.12) as

$$R_d = \frac{Z_0 + R_{\text{eff}}\omega \omega LQ_{\text{load}}}{R_{\text{eff}}}$$  \hspace{1cm} (4.30) \\
$$C_c = \left( R_d R_{\text{eff}} \omega \omega \right)^{-1/2}$$  \hspace{1cm} (4.31)

and shown in Fig. 4.5.

### 4.2.3 Charge sensitivity

The performance of the gate-based sensor formed using a NbN inductor and parallel resonant circuit is demonstrated using QDs formed in a nanowire transistor of 30 nm channel width and gate length. In this device QDs form in the top most corners of the nanowire with a strong gate coupling $\alpha$ as shown in Fig. 4.6(a). The formation of one dot in each corner is typical for these device dimensions. Coulomb diamonds for both QDs are observed in the phase of the reflected signal as shown in Fig. 4.6(b) from which a gate coupling $\alpha \approx 0.9$ is estimated. The signal originating from each QD can be distinguished based on a significant difference in the $I$ channel as shown in Fig. 4.6(c).

To benchmark the gate-based sensor, the signal of an electron tunnelling from a QD to a reservoir is considered. The dot-to-reservoir transition (DRT) at $V_G = 0.53$ V and $V_{SD} = 0$ mV is selected, where one QD is mostly coupled to one of the reservoirs. The reflection coefficient of the resonator is shown in Fig. 4.6(d). Figure 4.6(e) shows that in addition to a shift in resonant frequency at the DRT, there is also a change in the matching and quality factor and the resonator changes from being over-coupled to perfectly matched, indicating additional dissipative components. A change of $\Delta C_d = 0.4$ fF and $\Delta R_d = 670$ k$\Omega$ is calculated, by fitting the model presented in the previous sections. At the DRT, $Q_{\text{load}} = 448$ is obtained. Next, a measurement of the charge sensitivity is performed using conventional methods [147], as explained in Section 3.2.5, to provide a device-specific benchmark on the performance of our gate-based sensor normalised to the gate charge. A modulation with calibrated charge equivalent of $\Delta q_{\text{rms}} = 148 \mu e$ and frequency of 2.3 kHz
4.2 Parallel high-$Q$ resonator

**Figure 4.6:** Dot-to-reservoir transition. (a) Schematic view of the nanowire device along the gate and source-drain direction. Formation of corner dots and dot-to-reservoir transition (DRT) is indicated. (b) Stability diagram of the device as a function of gate and source-drain voltage showing Coulomb diamond QD characteristics. (c) Signals originating from the two corner QDs are distinguished by a difference in the $I$ channel. (d) Shows the reflection coefficient as a function of $V_G$ where a shift in frequency is observed at the transition. (e) Close-up of the change of the resonance operating away and at the transition.
produces side-bands with a SNR = 25 dB as shown in Fig. 4.7(a), when operating at the steepest point of the DRT with optimal RF frequency $f_0 = 615.9$ MHz and power $P_0 = -119$ dBm as shown in Fig. 4.7(b,c). This translates into a charge sensitivity of $\delta q = \Delta q / (\sqrt{2BW}10^{\text{SNR}/20}) = 1.3 \mu e/\sqrt{\text{Hz}}$ with BW being the resolution bandwidth of the spectrum analyser. Such sensitivity represents an improvement of a factor of 30 compared to previous results obtained with a similar device using a surface mount series resonator [162], and is enabled by the increased $Q_{\text{load}}$ (by a factor of 25) of the resonator and reduced parasitic capacitance (by 40%). This result places the charge sensitivity of gate-based sensors to the same level as reported RF-SET sensitivities [154].

### 4.2.4 Video-mode readout

An essential requirement for any QD based spin qubit is tuning to the desired few electron regime. This typically involves optimisation of voltages on multiple gates, where the voltage on one gate can affect the voltage required at another gate. Consequently, multiple consecutive measurements can be required and the measurement time spent taking stability diagrams as a function of gate voltages can be a bottleneck. The improved sensitivity presented in the previous section allows for fast video-mode measurements. For this, ramp-waveforms are applied to the gates [238], rather than stepping of each gate, which removes any time spent on instrument communication during the measurement, such that the measurement time is sensitivity limited. This requires a fast digitizer (e.g. AlazarTech or Spectrum Instrumentation PCI-e digitizer card) with sufficient memory, which is able
4.2 Parallel high-Q resonator

Figure 4.8: Video-mode readout. (a) Waveforms for fast acquisition using a digitizer. (b) Stability diagram obtained within 700 ms. (c) Zero bias peak with Lorentzian fit accompanied by additional peaks.

to acquire a complete stability diagram in a single buffer after being triggered using a synchronisation pulse. The pulse sequence for such measurement is illustrated in Fig. 4.8(a), where a slow ramp is applied to the first gate while the second gate undergoes multiple ramp cycles at the same time. For a large amount of fast ramp cycles the voltage on the first gate can be treated as constant per cycle. This sequence can be realised using a waveform-generator with two output channels and one synchronisation channel. Both output channels are triggered simultaneously and the measurement time depends on the sampling rate (SR) and the number of desired measurement points ($N_{\text{meas}}$), which together determine the waveform frequencies $f_{G_i}$ as follows

\[ SR < \frac{1}{t_{\text{min}}} \]
\[ N_{\text{meas}} = N_{G1}N_{G2} \quad (4.32) \]
\[ f_{G1} = \frac{SR}{N_{G1}} \]
\[ f_{G2} = f_{G1}N_{G2} \quad (4.33) \]
\[ t_{\text{meas}} = \frac{N_{\text{meas}}}{SR} > N_{G1}N_{G2}t_{\text{min}} \quad (4.34) \]

where $N_{G_i}$ are the desired number of measurement points for each gate ramp. In such, the overall measurement time $t_{\text{meas}}$ is limited by the minimum measurement time $t_{\text{min}}$ which is reduced as the sensitivity increases. The SNR of the measurement can be increased using a smaller sampling rate or by averaging over multiple maps.

Figure 4.8(b) shows a fast map of 512x256 points acquired using a 4 kHz ramp waveform applied to $V_{SD}$ while $V_G$ is ramped at 7 Hz. The total measurement time is $\sim 700$ ms. Coulomb blockade is observed identical to Fig. 4.6(b) and due to
the increased sensitivity, even fine features such as excited states of the electron in
the QD, can be observed at this measurement speed. A line-cut at $V_{SD} \approx 0$ mV is
shown in Fig. 4.8(c) to illustrate the SNR. The zero-bias peak is accompanied by
additional smaller peaks which could be due to excited states or nearby charge cen-
tres. From the Lorentzian fit a tunnel coupling of $\frac{\Delta}{2\pi} = 25.6$ GHz $\gg f_0$ is obtained,
which translates into an expected capacitive shift of $\Delta C_d = 0.39$ fF consistent with
measurements presented in the previous section.

4.3 Josephson Parametric Amplifier

In the previous section, sensitivity improvements of gate-based readout based on an
optimized parallel resonant circuit with NbN spiral superconducting inductor were
presented. Amplifiers based on Josephson junctions have greatly improved readout
SNR in the field of superconducting circuits [192–197], which typically operate at
frequencies of several GHz and near the quantum limit of noise introduced by the
amplifier (or indeed below, for a single quadrature using squeezing) [182, 198–200,
202, 203]. Adopting such approaches in the measurement of QDs at RF/microwave
frequencies is expected to lead to corresponding improvements in SNR. This can
in principle be achieved at the operating frequencies of 4–8 GHz, that are typical
for Josephson-junction based amplifiers, as demonstrated using an InAs double
QD, Josephson parametric amplifier (JPA) and coplanar waveguide resonator [238],
however, lower frequency operation ($\lesssim 1$ GHz) becomes necessary\(^1\) for studying
lower QD tunnelling rates, at which exchange interaction is more easily controlled.
Moreover, at lower frequency resonator fabrication separate from the QD device
is facilitated, enabling readout of densely packed QD arrays. Suitable amplifiers
are available in such a frequency range, for example: a JPA operating at 600 MHz
with a noise temperature of $T_{JPA} = 105$ mK [219] or a SQUID amplifier chain with
$T_{SQUID} = 52$ mK at 538 MHz [216]. Building on such developments, readout of a
GaAs based quantum dot at 196 MHz with a noise temperature of 490 mK was
recently reported using a SQUID amplifier [239].

In the following sections, gate-based sensing of silicon QDs is combined with a
Josephson parametric amplifier that operates in the 600–800 MHz band to push
the bounds of SNR that can be achieved using this technique. The well-matched
and high-$Q$ resonator presented in the previous section is required to enable JPA

\(^1\)operating at a frequency $f > \Delta_c/4h$ comparable to the tunnel coupling $\Delta_c$ results in back-action
of the resonator onto the quantum dot device in form of fast voltage oscillations that manifest
as Landau-Zener transitions
enhanced readout as the typical dynamic range of a JPA is of the order of $-130$ dBm, making it unsuitable for the signal powers commonly used in previous reflectometry measurements ($-90$ to $-80$ dBm) [162]. Moreover, the high $Q$-factor enables a reduced RF signal power to be used while achieving the same gate voltage on the device as summarised in Eq. (4.27) and demonstrated in Fig. 4.7(c).

4.3.1 Setup and JPA tuning

Setup  A schematic of how the JPA is embedded into the cryogenic reflectometry setup is shown in Fig. 4.9. The setup consists of i) the cryogenic RF delivery and amplification chain including the JPA (pink background); ii) a lumped-element $LC$ resonator (green); and iii) the silicon quantum dot device (blue). In comparison to a regular radio-frequency reflectometry circuit, operation of the JPA requires multiple additional microwave components operating at milliKelvin temperature. First of all, there is an additional input line carrying the JPA pump signal at power $P_{JPA}$ and frequency $f_{JPA}$. This signal is coupled into the reflectometry setup via a directional coupler which guides the signal towards the JPA. A circulator directs the JPA pump and RF readout signal into the JPA (which operates in reflection) and further directs the amplified signal to the cryogenic amplifier for additional amplification. To achieve a noise temperature on the order of milliKelvin at the JPA and the quantum dot device, microwave input lines are attenuated at multiple stages and there is an isolator at the milliKelvin stage of the output line. Moreover, QD bias voltages and the low-frequency JPA bias are filtered. The JPA and circulators are fitted with cryoperm shields that withstand fields up to 1500 G.

The JPA is a low quality factor ($Q_{JPA} < 100$) superconducting resonator consisting of a SQUID loop array (6 SQUIDs) with tunable inductance shunted by a fixed capacitance as shown in Fig. 4.10(a,b). Fabricating a low-frequency JPA requires careful engineering of the capacitance and SQUIDs [176]. The main JPA characteristics that need to be considered are the resonant frequency, bandwidth and dynamic range. For a JPA compatible with RF reflectometry, a resonant frequency below 1 GHz is desired. Moreover, the resonance should have a sufficiently low quality factor to obtain a bandwidth of a few MHz and operate as a linear amplifier for input powers up to $-120$ dBm (power at which maximum sensitivity is achieved in Fig. 4.7(c)). All three parameters are determined by two physical parameters,
the overall capacitance $C$ and the critical junction current $I_0$ as follows [176]

$$\omega_{\text{JPA}} = \frac{1}{\sqrt{LC}}$$

$$Q_{\text{JPA}} = Z_0\omega_{\text{JPA}}C$$

$$L_j = \frac{\phi_0}{I_0}$$

$$P_{\text{max}} \propto I_0^2$$

where $L$ is the overall inductance. For a JPA operating at a few GHz typical values are $C = 5$ pF (overlap capacitor, and stray capacitance) and $L = 100$ pH (junction and stray inductance) [204]. In this low-frequency design, instead of a single SQUID an array of SQUIDs is selected to maintain sufficient dynamic range and bandwidth when reducing the resonant frequency. In such, the critical current is increased (with the number of SQUIDs) while the Josephson inductance remains unchanged.
Figure 4.10: JPA tuning. (a) The JPA device schematic, chip and packaging. (b) SEM micro-graph of the SQUID array that forms the JPA resonator. (c) The JPA resonance is tunable in frequency by passing a current $I_{bias}$ through a nearby coil. The JPA response as a function of pump power in the magnitude (d) and phase (e) of the reflected signal. At high power the non-linearity leads to a bi-stability.
Fast and sensitive gate-based RF-readout

Figure 4.11: JPA gain. (a) JPA gain as a function of pump power and (b) frequency at fixed $I_{bias}$ for a small additional signal at frequency $f$. (c) JPA gain profile when tuned to 754 MHz for different signal power. Gain is reduced with increasing power due to saturation. (d) JPA gain profile when tuned to 765 MHz. At higher frequency, further away from the bi-stability point larger bandwidth at reduced maximal gain is achieved.

JPA tuning  The JPA is tunable in frequency from 500–800 MHz, as shown in Fig. 4.10(c), by passing a current $I_{bias}$ through a coil that changes the flux through the nearby SQUIDs. At low drive power, the JPA behaves like a linear resonator and is only observed through a phase shift as shown in the magnitude and phase in Fig. 4.10(d,e). At high power, the non-linearity of the Josephson junctions manifests in a frequency shift of the JPA to lower frequency (parametric regime) until eventually the JPA reaches a bi-stable regime [176], where a dip in the magnitude signal is observed. To achieve gain at a desired frequency, the JPA is first tuned in frequency using $I_{bias}$ followed by identifying the parametric regime in a power scan as shown in Fig. 4.10(c-e). Parametric gain can be obtained for small input signals within the JPA bandwidth by pumping the JPA at resonant frequency (which is power dependent) in the parametric regime. An illustrative combination
4.3 Josephson Parametric Amplifier

**Figure 4.12: JPA operation principle.** (a) Schematic of the JPA transfer function. Small modulation in the JPA pump power lead to large modulations in the phase of the reflected signal. (b) Schematic illustration of the JPA operation principle in phase-preserving mode in the IQ plane. In the rotating frame of the pump signal (blue), a small input signal (at different frequency and in green) drives a circular modulation leading to double-sideband phase modulation. (c) Schematic illustration of the JPA operation principle in phase-sensitive mode. When pump (blue) and signal (green) are at the same frequency gain is only achieved when both are in-phase, represented by an increased vector difference between both signals at the output. (d) Experimentally obtained transfer function at different pump frequency. At the bi-stability there is an abrupt jump in the phase. At higher frequency the transfer function becomes less steep. Adapted from [178].
of JPA pump frequency and power within the parametric regime is indicated by a dotted line in Fig. 4.10(e). Fine tuning of both the pump power $P_{\text{JPA}}$ and frequency $f_{\text{JPA}}$ is typically necessary and shown in Fig. 4.11(a,b) respectively, where the gain $G$ for a small signal at power $-130$ dBm and frequency $f$ in addition to the pump is shown. The dynamic range and saturation of the JPA is demonstrated in Fig. 4.11(c), where saturation is observed at an input signal power $> -130$ dBm. Depending on the operation point within the parametric regime (close to or far away from bi-stability), there is a trade-off between amplification gain and bandwidth. Figure 4.11(d) shows an operation point at higher frequency (same $I_{\text{bias}}$), where a smaller gain at approximately twice the bandwidth is achieved. This can be understood when considering the operation principle of such JPA that is operated near a bi-stability and pumped at $f_{\text{JPA}}$ with $P_{\text{JPA}}$. The transfer function of the JPA is given by the steep change in reflected phase as a function of pump power as shown in Fig. 4.12(a). When biased at the steep point, changes in the pump amplitude due to a small signal lead to large changes in the reflected phase, such that the gain is determined by the gradient of the transfer function while the width in power sets the dynamic range. For small modulations of $P_{\text{JPA}}$, the response is linear. At the bi-stability, instead of a smooth edge in the phase, there is an abrupt jump and further away from the bi-stability (at higher pump frequency), the transfer function is less steep (and at lower power) as observed in line-cuts at different frequency shown in Fig. 4.12(d).

The amplification process can be further explored in the $IQ$ plane [178]. When the input signal is at a different frequency than the pump, the JPA operates in phase-preserving mode, and in the rotating frame of the pump frequency the small input signal appears as a rotating signal which introduces a sinusoidal modulation of the pump amplitude. Consequently, the pump signal at $\omega_p$ exhibits double-sideband phase modulation at frequency $\omega_p \pm \Delta \omega$ (signal and idler mode) with $\Delta \omega = \omega_p - \omega_s$ and direct gain is achieved both at the signal and idler frequency as shown in Fig. 4.12(b). When the small signal is at the same frequency as the pump, see Fig. 4.12(c), gain is only achieved for a signal in-phase with the pump, where a small signal modulates the pump power coherently and produces a large modulation in the output phase (as shown in Fig. 4.12(a)) with a vector difference that is much larger than between the input states (amplification is achieved). For a signal that is in quadrature with the pump, the vector sum of the small signal and pump is constant over time to first order such that the phase modulation of the pump signal is very small.
The characteristics of the JPA explored in this section demonstrate that enhancements of the RF readout signal with a gain on the order of 20 dB can be expected.

4.3.2 \textit{LC} resonator and JPA operation

In this section the JPA is tuned, using the procedures described in the previous section, such that gain is achieved at the frequency of the readout resonator. By including the JPA as an additional amplifier with gain $G_{\text{JPA}} \gg 1$ and lower noise temperature ($T_{\text{JPA}}$) at the beginning of the amplification chain, the effective noise temperature $T_{\text{noise}}$ is reduced:

$$T_{\text{noise}} = T_{\text{sys}} + T_{\text{JPA}} + \frac{T_{\text{HEMT}}}{G_{\text{JPA}}}.$$  \hspace{1cm} (4.39)

In the absence of the JPA, $T_{\text{noise}}$ is limited by the effective noise temperature $T_{\text{HEMT}}$ of the cryogenic high electron mobility transistor (HEMT) amplifier, typically a few Kelvin. For a JPA operating at $T = 10$ mK a minimum of $T_{\text{JPA}} = \frac{\hbar \omega}{2k_B} \coth \left( \frac{\hbar \omega}{2k_B T} \right) = 16.5$ mK is expected.

\textbf{JPA and \textit{LC} resonator} \hspace{1cm} Figure 4.13(a) shows the JPA response as a function of pump power when tuned such that the regime useful for parametric amplification coincides with the \textit{LC} resonator. The \textit{LC} resonator is formed by the parallel combination of a NbN spiral inductor $L = 170$ nH, parasitic capacitance and the quantum dot device ($C_p + C_d = 380$ fF), all coupled to the RF line via a coupling capacitor ($C_c = 37$ fF) as shown in Fig. 4.9. A resonance in the reflection coefficient $\Gamma = |\Gamma| \exp(i\phi)$ is observed at $f_{\text{RF}} = 1/2\pi \sqrt{L(C_c + C_p + C_d)} = 621.9$ MHz with a loaded quality factor of $Q_{\text{load}} = 966 \pm 8$, impedance $Z = \sqrt{L/(C_p + C_c + C_d)} = 650 \Omega$, return loss of 3 dB and phase shift $> 180^\circ$ (over-coupled) as shown in Fig. 4.13(b). When operating at a charge instability in the QD device, the resonator reaches perfect matching as previously observed in Fig. 4.6(e). The JPA is operated in phase-preserving mode, where there is an offset $\Delta f = f_{\text{JPA}} - f_{\text{RF}}$ between the JPA pump frequency ($f_{\text{JPA}}$) and $f_{\text{RF}}$, so power from the JPA pump is transferred onto $f_{\text{RF}}$ and $f_{\text{JPA}} + \Delta f$ (four-wave mixing) via double-sideband phase modulation. $\Delta f = 1$ MHz is selected to fall between the bandwidth of the resonator $\Delta f_{\text{RF}}^{3\text{dB}} = 0.65$ MHz and the JPA $\Delta f_{\text{JPA}}^{3\text{dB}} = 6$ MHz. This puts $f_{\text{JPA}}$ at the edge of the readout resonator to avoid power broadening due to leakage of the pump signal while maximizing gain at $f_{\text{RF}}$. When tuned and pumped, a gain of 17 dB is achieved at $f_{\text{RF}}$ as shown in Fig. 4.13(d). The decrease in gain near $f_{\text{RF}}$ is likely due to large
Figure 4.13: JPA and LC resonator. (a) JPA phase response as a function pump power and frequency. The regime of parametric amplification coincides with the readout resonator which are both highlighted. (b) Magnitude and phase of the reflection coefficient showing the readout resonator ($Q_{\text{load}} = 966 \pm 8$). (c) Gain profile of the JPA when tuned close to the readout resonator frequency (3-dB-bandwidth of 6 MHz). (d) JPA gain and estimated system noise temperature at $f_{\text{RF}}$ as a function of RF input power showing saturation at high input power (1-dB-compression at $-121 \, \text{dBm}$).
impedance variations of the resonator close to resonance and imperfect matching to 50Ω.

**Noise temperature**  Figure 4.13(c) shows the JPA gain and the effective noise temperature close to \( f_{RF} \) as a function of \( P_{RF} \). 1-dB-compression is observed at \(-121 \) dBm. Based on amplifier gain estimations (\( G_{HEMT} = 27 \pm 2 \) dB) an effective noise temperature \( T_{\text{noise}} = 2.5^{+1.4}_{-0.9} \) K with the JPA off (consistent with the cryogenic amplifier specifications) and a minimum noise temperature of \( T_{\text{noise}} = 200^{+110}_{-73} \) mK based on the SNR improvement with the JPA on is obtained. The effective noise temperature with the JPA on increases with increasing power due to saturation. There are multiple contributions to \( T_{\text{noise}} \), captured in Eq. (4.39). The contribution of the cryogenic amplifier \( \frac{T_{\text{HEMT}}}{G_{JPA}} = 50^{+28}_{-18} \) mK and \( T_{\text{JPA}} \) and \( T_{\text{sys}} \) which are estimated by comparing \( T_{\text{noise}} \) when operating the QD device away from or at a charge instability. \( T_{\text{sys}} \) can have contributions from the resonator circuit (\( T_{\text{circ}} \)) and the QD device (\( T_{\text{QD}} \)): \( T_{\text{sys}} = (1 - |\Gamma|^2)T_{\text{circ}} + kT_{\text{QD}} \) [149], where \( k \) is the fraction of the rf power dissipated in the device. Tunnelling between the QD and reservoir occurs adiabatically when the tunnel rate is larger than the RF drive and hence \( k = 0 \). Based on an increase in \( T_{\text{noise}} \) of \( 35^{+24}_{-13} \) mK when operating at a charge transition (where \(|\Gamma|\) decreases from 0.5 to 0), \( T_{\text{JPA}} = 47^{+35}_{-30} \) mK and \( T_{\text{circ}} = 142^{+94}_{-54} \) mK is estimated. \( T_{\text{circ}} \) relates to an electron temperature of dissipative elements in the resonant circuit and a JPA efficiency of 36% of the quantum limit (equivalent to \( \sim 1.5 \) photons) is compatible with previous results for operation close to a bifurcation point [197, 205, 206].

### 4.3.3 QD device

The improvements in the SNR of gate-based readout with the JPA are benchmarked using a QD-to-reservoir charge transition (DRT) and inter-dopant/dot charge transition (IDT) in a CMOS silicon nanowire field-effect transistor device with channel width and gate length of 30 nm similar to the previous device in Section 4.2.3. Figure 4.14(a) shows a schematic line-cut of the device along the gate. QDs form in the corners of the device and have a strong coupling to the gate \( \alpha_{\text{DRT}} = 0.86 \). Additionally, given the doping density, an average of 5 (phosphorus) donors are expected in the device channel [74] and one donor is observed in the device that is coupled to one of the reservoirs and both QDs. The donor is identified in Fig. 4.14(b) as a single line that couples more strongly to \( V_{BG} \) and anti-crosses with transitions of
both QDs. The IDT and DRT that are discussed further in the next sections are highlighted.

4.3.4 Charge sensitivity

To benchmark charge sensing with and without the JPA, the charge sensitivity at the DRT is obtained, in which a QD is primarily tunnel coupled to the drain reservoir as illustrated in a schematic line-cut of the device along the source-drain direction in Fig. 4.14(a). When operating at the DRT a capacitive shift of the resonance corresponding to $\Delta C_d = 0.5 \, \text{fF}$ is observed.

For measurements of the charge sensitivity a small sinusoidal modulation is applied to the top-gate of the device at frequency $f_{AC}$ and root-mean-square charge equivalent amplitude $\Delta q_{\text{rms}}$ and the SNR of sidebands appearing at $f_{\text{RF}} \pm f_{AC}$ is monitored (see Section 3.2.5 for details of the method). Fine-tuning of the gate-voltage offset is shown in Fig. 4.15(a) with and without the JPA. With the JPA on a SNR enhancement is observed and the point at which the maximum SNR is obtained and the overall width of the transition is identical. This indicates that there is no significant leakage of the strong pump signal into the readout resonator which could lead to broadening of the DRT. In this particular device, low-frequency noise close to the RF carrier is observed, as shown in Fig. 4.15(b), which resulted in a strong reduction in SNR at low $f_{AC}$ due to an elevated noise floor. The noise floor decreases with increasing $f_{AC}$ until a reduction of the modulation amplitude at

![Figure 4.14: Quantum dot device. (a) Schematic cross-section of the QD device along the gate, and the source and drain, showing two QDs in the top corners of the nanowire and a donor in the channel. Transitions at which measurements are performed are indicated. (b) Phase response of the device as a function of $V_G$ and $V_{BG}$ showing multiple lines originating from corner QDs and one line corresponding to the donor.](image)
Figure 4.15: Charge sensitivity. (a) Fine tuning of the gate voltage offset. Highest SNR corresponds to the steepest slope of the DRT. (b) Signal-to-noise ratio as a function of modulation frequency at $P_{RF} = -120$ dBm and low-frequency line filter function. (c) Power broadening of the selected transition with line-shape shown inset. The regime of significant power broadening is indicated by a dashed line. (d) SNR as a function of $P_{RF}$. The power at which JPA saturation (dotted) and significant power broadening occurs (dashed) indicated. (e) Charge sensitivity at $P_{RF} = -130$ dBm as a function of charge equivalent modulation amplitude.
$f_{AC} > 10 \text{ kHz}$ occurs due to line filtering. An operation frequency of $f_{AC} = 50 \text{ kHz}$ is selected to minimize the effect of charge noise on the charge sensitivity measurement and amplitude reduction due to filters on the line is calibrated: only 11% of the amplitude is delivered to the device. When operating at small RF power below power broadening $P_{RF} < -130 \text{ dBm}$ the low-frequency noise is further reduced. Figure 4.15(c) shows the full-width half maximum of the selected DRT (see inset) as a function of $P_{RF}$ to determine the maximum power before power broadening. Using a model of $\text{FWHM} = \gamma_0 \sqrt{1 + \frac{P_{RF}}{P_{\sqrt{2}}}}$ a natural line-width of $\gamma = 105 \mu\text{V}$ and $P_{\sqrt{2}} = -110 \text{ dBm}$ is extracted, where $P_{\sqrt{2}}$ is the power at which significant power broadening (factor of $\sqrt{2}$) occurs. Due to the high $Q_{\text{load}}$ of the resonator, only a small input power, compatible with the dynamic range and saturation of the JPA, is required and the RF disturbance at the device gate is calculated as $V_{\text{RF}}^{\text{pp}} = \frac{2C_c}{C_c + C_p + C_d} Q_{\text{load}} V_{\text{in}}^{\text{pp}}$, corresponding, for example, to $V_{\text{RF}}^{\text{pp}} = 13 \mu\text{V}$ for an input RF power of $-130 \text{ dBm}$.

Next, the SNR as a function of $P_{RF}$, with and without the JPA, is shown in Fig. 4.15(d). An improvement of up to 8 dB in SNR is observed with the JPA at low RF power. Irrespective of whether the JPA is used, for $P_{RF}$ between $-130$ and $-120 \text{ dBm}$ the SNR levels off as the DRT begins to become power broadened, and the SNR drops abruptly for powers above $-110 \text{ dBm}$. With the JPA on there is an additional decrease in SNR evident above $-120 \text{ dBm}$ as the JPA saturates. This shows that the JPA can either be used to increase the SNR beyond what could otherwise be achieved, and/or to provide the same SNR but at about 10 dB less RF power, with the corresponding reduction in the disturbance of the QD being measured, and its neighbours. Finally, Fig. 4.15(e) shows the optimal charge sensitivity, calculated as $\delta q = \Delta q_{\text{rms}} / (\sqrt{2} \cdot BW \times 10^{\text{SNR}/20})$, as a function of charge equivalent modulation amplitude (at $P_{RF} = -130 \text{ dBm}$), which shows that best sensitivity is achieved at small modulation. The charge sensitivity achieved with the JPA is $0.25 \mu\text{e}/\sqrt{\text{Hz}}$ compared to $0.5 \mu\text{e}/\sqrt{\text{Hz}}$ without the JPA, outperforming previous measurements using RF-SET [154] and gate-based approaches [236].

### 4.3.5 Video-mode improvements

The improvements in sensitivity when using a JPA presented in the previous section can be observed in fast video-mode measurements. Figure 4.16 shows measurements obtained within 2 s each at different power $P_{RF}$ and with the JPA either on or off. At $P_{RF} = -130 \text{ dBm}$ and with the JPA off the Coulomb diamond characteristic is barely visible as shown in Fig. 4.16(a). The signal improves significantly when
increasing the power to \( P_{\text{RF}} = -120 \text{ dBm} \) as shown in Fig. 4.16(b). A similar improvement is achieved when using \( P_{\text{RF}} = -130 \text{ dBm} \) but with the JPA on as shown in Fig. 4.16(c), which achieves a nearly identical SNR as demonstrated in the line-cuts as \( V_{\text{SD}} = 0 \text{ mV} \) below each figure. The fast maps at different RF power with and without the JPA highlight again, that while using the JPA leads to an increase in the SNR, it additionally allows operation at lower input power while achieving identical SNR reducing the overall disturbance of the RF signal onto the quantum device.

### 4.3.6 Inter-dot charge transition

In a coupled DQD, Pauli spin blockade allows for direct singlet-triplet readout using a gate-based sensor. This not only makes an architecture more compact, but also allows direct readout of spins which might have otherwise been far away from a readout reservoir and prior shuttling of the electron to a reservoir would be required. Moreover, as the singlet-triplet splitting is the basis for readout, the Zeeman energy is no longer required to be much larger than the thermal energy (which is a readout fidelity limiting factor in reservoir based readout), and readout at low external magnetic fields can be performed, which significantly reduces the requirements on microwave sources for spin manipulation. Consequently, singlet-
triplet readout alleviates multiple design constraints and features in several scalable silicon-based quantum computing architectures [50, 53, 133].

In this section, singlet-triplet based readout with and without the JPA is benchmarked using a donor-dot IDT. Figure 4.17(a) shows the IDT in the normalised phase response as a function of $V_G$ and $V_{BG}$, where the donor transition can be identified due to a stronger coupling to $V_{BG}$ (donor resides deeper in the channel, closer to the back-gate). When operating the back-gate, a reduction of resonator quality factor is typically observed as summarised in Appendix D. Figure 4.17(b) shows a reduction of the IDT signal with increasing magnetic field, which indicates that the IDT is of even parity [172], where singlet and triplet states are formed. As

![Figure 4.17: Inter-donor/dot charge transition.](image)

(a) Even-parity IDT between a donor and QD in the device. Effective electron occupation ($N_{dot}$, $N_{donor}$) indicated, up to an arbitrary offset. (b) Close-up of the IDT at different external field $B_z$. Signal slowly disappears with increasing field indicating even charge parity. (c) Normalized phase response along the detuning axis $\epsilon$ in (a) with and without the JPA for two different integration times (traces offset by 1.5 in $\Delta \phi$ for clarity). (d) SNR obtained from traces as shown in (b) as a function of integration time with JPA on and off. Linear extrapolation and SNR = 1 is indicated using dotted lines.
the magnetic field is increased the triplet $T_-$ state becomes the ground state. Due to the Pauli exclusion principle the triplet energy is linear as a function of detuning which results in zero charge susceptibility compared to the singlet which has a large curvature at zero detuning (see Fig. 2.5 and Eq. (2.49)). In Fig. 4.17(b) the IDT shifts towards higher $V_G$ with increasing field, which indicates that the $(2,0)$ charge configuration is found at positive detuning $\varepsilon$ where the $T_-(1,1)$ and $S(2,0)$ states cross. In this analysis a $(1,1)-(2,0)$ charge occupation is assigned to the IDT based on the fact that it is an even parity transition and no additional transitions are observed at lower voltage. However, there could be an arbitrary offset in the number of electrons. Additionally, from the shift to larger voltage (see Eq. (2.51)) with increasing field and the expected line-shape of the signal (see Eq. (2.49)) a gate coupling $\alpha = 0.36$ eV/V, tunnel coupling $\Delta_c = 20.9 \mu$eV and maximum capacitive shift 0.5 fF is calculated. Further measurements on the magnetic field response are shown in Appendix C.

### 4.3.6.1 SNR improvements

Figure 4.17(c) shows an example of the phase response across the IDT, along the detuning axis $\varepsilon$ shown in Fig. 4.17(a) for $\tau_{\text{int}} = 1$ $\mu$s and $\tau_{\text{int}} = 52$ $\mu$s with the JPA off and on for comparison. From such traces the power SNR is obtained as $\text{SNR} = \frac{\Delta \phi^2}{\sigma^2}$ where $\Delta \phi$ is the amplitude of the IDT signal and $\sigma$ the RMS amplitude of the noise. Figure 4.17(d) shows the IDT SNR as a function of integration time with the JPA on and off. An improvement in SNR of a factor of 7 is observed with the JPA on, consistent with the SNR improvement of 8 dB obtained in the charge sensitivity measurement. Using an extrapolation (dotted lines in Fig. 4.17(d)) a SNR of unity is reached at $t_{\text{off}}^{\text{min}} = 1.2$ $\mu$s and $t_{\text{on}}^{\text{min}} = 80$ ns with the JPA off and on respectively. However, the limited bandwidth of the resonator prohibits measurements faster than 1.5 $\mu$s ($\Delta f_{\text{RF}}^{3\text{dB}} = 0.65$ MHz). Additionally, multiple measurements of the SNR with the JPA on for $t_{\text{int}} < 10$ $\mu$s slightly deviate from the extrapolation, which could be due to noise introduced by the JPA pump signal which is operated only 1 MHz offset the RF signal.

Overall, measurements of the IDT with the JPA on and off show similar improvements in SNR as the previously analysed DRT. Combined with the high-quality resonator the SNR improvements with the JPA on allow singlet-triplet gate-based readout in the sub-microsecond time-scale.
4.3.6.2 Readout fidelity estimations

Ultimately, the single-shot readout fidelity at a given integration time, determines how fast and how well readout can be performed and combines limiting effects due to sensitivity of the sensor and coherence of the qubit as introduced in Section 4.1.

Based on the signal and noise levels, as shown in Fig. 4.17(c), the singlet and triplet readout probability densities, see Eq. (4.1), for many single-shot measurements are simulated, assuming a relaxation of the triplet within $T_1 = 4.5 \text{ ms}$ [235]. The simulated combined singlet and triplet probability distribution is shown in Fig. 4.18(a) as a function of integration time $t_{\text{int}}$ with the JPA off. While at low $t_{\text{int}}$ there is a large overlap between the singlet and triplet distribution that are separated by $\Delta \phi_{\text{norm}} = 1$, two distinct peaks form at large $t_{\text{int}}$ indicating high readout fidelity. From the probability densities the average readout infidelity with the JPA on and off is obtained, see Eq. (4.5) and is shown Fig. 4.18(b) where for each $t_{\text{int}}$ the optimal threshold is determined. $F_{\text{avg}} > 0.997$, indicated by a horizontal dotted line and corresponding to SNR > 5 in this setting, can be reached at an integration time of at least $t_{\text{int}}^{\text{off}} = 32 \mu s$ with the JPA off and $t_{\text{int}}^{\text{on}} = 1 \mu s$ with the JPA on. This demonstrates that the JPA enables high-fidelity readout that is much faster than the coherence time of electron spins in $^{28}\text{Si}$ ($T_2^* = 120 \mu s$ [91]).
4.4 Impact of external magnetic fields

Spin qubits in semiconductor devices are usually operated at fixed finite magnetic fields. Both the high-quality readout resonator and the JPA are made from superconducting circuits such that a response to a large external magnetic field is expected. While the $LC$ resonator is contained in the sample box right next to the QD device that is placed in the centre field $B_z$ of a cryogenic magnet, the JPA is mounted on the mixing chamber plate 40 cm away from the field centre where a maximum of 500 G is expected at $B_z = 6$ T.

Figure 4.19(a) shows that the field $B_z$ leads to a reduction of the readout resonator frequency due to a change in kinetic inductance. At first, up to 0.5 T, the resonance does not change significantly and then reduces by about 30 MHz from 0.5 T to 3 T. Similarly, the JPA resonance changes with applied field. The JPA is especially
susceptible to external fields as its resonance is tuned using the tunable inductance of SQUID loops. Any field penetrating into the loop tunes the resonance frequency. Additionally, any field noise strongly affects the JPA performance. Consequently, the JPA is enclosed in a cryoperm magnetic shield as shown in Fig. 4.9. The JPA resonance frequency merely changes (< 10 MHz) up to 2 T, as shown in Fig. 4.19(b), but above this field some of the external field penetrates into the shield resulting in a resonance frequency shift of about +100 MHz at 3 T. An obvious measure of JPA performance is the gain achieved in an external field $B_z$. For this the JPA is tuned to a gain of 28 dB at 644 MHz. In Fig. 4.19(c) the gain drops quickly as $B_z$ is increased and settles towards 15 dB at 0.2 T. While this is a significant change in gain, tuning a single parameter only, namely the flux bias $I_{\text{bias}}$, restores the gain of 28 dB and this was demonstrated for $B_z$ up to 3 T as shown in Fig. 4.19(d). This illustrates that up to 3 T, the performance of the JPA is not reduced and restoring the gain after a change in magnetic field could be easily automated by optimizing a single parameter. However, at some point, additional tuning might be necessary once the readout resonator frequency has shifted beyond the bandwidth of the JPA.

### 4.5 Conclusion

Results presented in this chapter demonstrate that a high-quality resonator based on a parallel circuit containing a NbN spiral inductor that is fabricated separately from the QD device combined with a JPA allows for fast RF gate-based readout of high-fidelity. Readout fidelity beyond the fault tolerant threshold within a measurement time faster than the coherence time $T_\ast$ is expected, as required to implement error correction codes and fast feedback in silicon-based quantum devices. Resonator fabrication separately from the QD supports readout of high density QD architectures and integration in a multi-layer quantum-classical chip could be achieved, which is further discussed in Chapter 6.

The sensitivity of the method is currently limited by the gain of the JPA, yielding a contribution of the cryogenic amplifier of at least 50 mK to the system noise temperature. Assuming a JPA gain of 23 dB or more, the contribution of the cryogenic amplifier would become negligible. The system noise performance could be further improved by operating the JPA in phase-sensitive mode, where the noise added by the JPA can be squeezed below the quantum limit. Changes in the circuit such as additional isolators between the JPA and readout resonator as well as additional line attenuation and filtering could be beneficial towards achieving larger gain, a lower system noise temperature and preventing
leakage of the JPA pump signal into the readout resonator. This can be significant when the pump falls within the \(LC\) resonator bandwidth, as shown in Fig. 4.20.

Additionally, the measurement speed in this implementation is, in principle, limited by the bandwidth of our high-\(Q\) readout resonator: increasing the coupling to the line or, preferentially, moving to a higher frequency of the resonator circuit (i.e. between 1 and 2 GHz) while maintaining high \(Q_{\text{load}}\) should allow sub-microsecond fault-tolerant gate-based spin readout.

Further development could reduce the footprint of the high \(Q\) resonators, to achieve an integrated and scalable readout architecture (see Chapter 6) with the potential of reduced circuit losses and parasitics. Using a travelling wave parametric amplifier (TWPA, see Section 2.4.3.2) with increased bandwidth, frequency multiplexing of multiple resonators could be achieved.
Chapter 5

Random-access readout

_In this chapter, random-access readout of QDs in Si nanowire transistor devices using gate-based reflectometry is presented. To begin, a CMOS single-electron memory cell, the building block of random-access readout, is demonstrated at milliKelvin temperature followed by sequential readout of two devices. An electron retention time on the order of one second is observed allowing for readout of many devices at the same frequency, one after another, before gate voltage refresh is required. Moreover, estimations of the dynamic power dissipation suggest that simultaneous operation of more than 10,000 cells is feasible._

Experiment have been performed at the Hitachi Cambridge Laboratory and the University of Cambridge using ‘Cambridge’ type PCBs under supervision of Dr M. F. Gonzalez-Zalba. The author acknowledges contributions from Dr A. Rossi for assistance in measurements and data analysis, V. N. Ciriano-Tejel for circuit simulations, and CEA-Leti for fabrication of nanowire transistor devices.

5.1 Motivation

In current solid-state quantum processors, signals are generated using general-purpose instruments at room temperature and delivered to the quantum processor at low temperatures. As the size of quantum processors continues to increase, the one-qubit-one-input approach will be unsustainable [34], especially when considering a large-scale fault-tolerant quantum computer that might ultimately require $10^8$ qubits to solve computationally demanding algorithms [136]. Efficiently delivering control and readout signals to increasingly more complex quantum circuits, while reducing the number of inputs per qubit, is a key challenge in developing a large-scale universal quantum computer.

Gate-based radio-frequency (RF) readout is a compact way for readout of qubits in silicon (see Section 2.3) and allows for high-density integration by making reser-
voirs and external sensors redundant. Frequency-domain multiplexing using multiple lumped-element circuits is a useful method to form multiple gate-based sensors that use the same line, however, the scalability of this approach is limited by the accessible readout bandwidth, where frequency crowding can quickly become a problem [240].

Therefore, random-access readout of multiple quantum dot devices, one after another, provides a solution to readout of a large number of devices using a single line and an overall quadratic reduction of input lines by adapting the concept of a memory cell with floating gate charge storage, routinely used in dynamic-random-access memory (DRAM) circuits, to gate-based RF readout. The key steps towards random-access readout of many devices are conditional high-sensitivity readout of at least two such memory cells, combined with sufficient floating gate charge storage when a memory cell is not selected for readout, which are presented in the following sections.

## 5.2 Conditional readout

In this section, a CMOS single-electron memory cell is characterised and conditional readout of high-sensitivity and charge retention is demonstrated — the basis for a dynamic access readout scheme.

### 5.2.1 Single electron memory cell

**Setup** In Fig. 5.1(a) a schematic of a CMOS single-electron memory cell (including SEM micrographs of individual transistor devices) and the $LC$ resonator circuit for readout is shown. The memory cell consists of a control field-effect-transistor (FET) and a quantum device which are both fabricated using the same silicon-on-insulator (SOI) CMOS processes and realised on the same chip (see Section 3.1.1). The configuration presented here resembles that of a single dynamic random-access memory (DRAM) cell in which the role of the charge storage capacitor is now played by the quantum device, realised in a nanowire transistor (60 nm wide channel, 50 nm gate) where single-electrons are trapped, while the ‘access FET’ in DRAM has the role realised by a micron wide channel transistor (‘control FET’) and the readout electronics is represented by the $LC$ resonator. The role of the FET is to retain a voltage at the quantum dot gate in the off state (charge storage) in order to keep the number of electrons in the dot constant, while enabling selective RF readout of the quantum device charge state in the on state, such that gate sensing can only
be performed when the FET is in the on state. The electrical connection between the FET and quantum device is made on-chip using a short bond wire (indicated by a cyan line), which has a significant contribution to the storage capacitance as demonstrated later on. The combined quantum-classical CMOS circuit has two primary inputs which are referred to as the word- and data- (bit) line in analogy to a multiplexer or memory device. The word-line is connected to the gate of the control FET, while the data-line signal passes through the channel of the control FET and is applied to the gate of the quantum device when the FET is in the on state. Source-drain transport through the quantum device can be measured directly and readout based on RF reflectometry can be performed (when the control FET gate voltage $V_{WL}$ is above threshold) by applying an RF modulation via the data-line (using an on-PCB bias tee) to an LC resonant circuit made from a surface mount inductor and the parasitic capacitance of the device $C_p$. The LC resonator response is amplified at multiple stages, followed by IQ-demodulation (see Section 3.2.4) from which the amplitude and phase of the reflected signal is obtained. The phase $\phi$ of the reflected signal is sensitive to small changes $\Delta C_d$ in the capacitance of the quantum device, associated, for example, with the tunneling of single-electrons: $\Delta \phi \approx -2Q_{load}\Delta C_d/C_T$ (see Section 2.3) with $Q_{load}$ being the loaded quality factor of the resonator and $C_T$ being the total capacitance of the circuit which includes the parasitic capacitance in parallel with the memory cell capacitance. The whole circuit is operated in a dilution refrigerator with a base temperature of 50 mK (Kelvinox K400).

**DC characterisation** To begin, the memory cell is characterised through transport measurements. In Fig. 5.1(b) the source-drain current $I_{SD}$ through the quantum device as a function of $V_{DL}$ and $V_{WL}$ at a source-drain bias of $V_{SD} = 1$ mV is shown. When the control FET is operated below threshold (the off state), the gate of the quantum device is isolated from the signal on the data-line. In this state of the circuit, the quantum device gate floats, allowing it to retain its charge over a timescale of a second, as demonstrated later on. For measurements where $V_{WL}$ is ramped slowly (as in Fig. 5.1(b)), the quantum device gate voltage tends to 0 V when the control FET is off. Once the control FET is operated well above threshold the turn-on of the quantum device transistor can be observed, while a transition region is also apparent where the control FET is still strongly resistive and a fraction of the data-line voltage reaches the quantum device gate. The quantum device threshold voltage is estimated as $V_{th}^Q = 0.63$ V (for $V_{WL} > 1.05$ V ), and the control FET threshold of $V_{th}^{FET} = 0.37$ V is obtained using $V_{th}^{FET} = V_{WL} - V_{DL}$ at $(V_{WL}, V_{DL}) = \ldots$
Figure 5.1: Setup and transport characterisation. (a) Memory cell (highlighted using a dashed box) and measurement circuit schematic, including SEM micrographs of the control FET and quantum device. Control and measurement signals are delivered to the quantum device via the channel of the control FET. (b) Transport through the quantum device as a function of $V_{DL}$ and $V_{WL}$ indicating the threshold voltage of the control FET and quantum device at $V_{BG} = 0$ V. (c) Turn-on characteristic of the quantum device as a function of $V_{BG}$ when the FET is biased well above threshold at $V_{WL} = 1.3$ V. (d) Illustration of the nanowire-based quantum device cross-section (along the gate and source-drain direction) under high back-gate bias and near-threshold top-gate bias, such that a single quantum dot forms. (e) Coulomb diamonds indicating a single quantum dot in the quantum device at $V_{BG} = 10$ V.
(1.02, 0.65) V and \( V_{BG} = 0 \) V. An additional tuning parameter for devices based on SOI technology is the back-gate voltage \( V_{BG} \) applied to the substrate which affects both the control FET and quantum device as they are realized on the same chip. In Fig. 5.1(c), \( I_{SD} \) as a function of \( V_{DL} \) and \( V_{BG} \) with the control FET biased at \( V_{WL} = 1.3 \) V (on state) is shown. As \( V_{BG} \) increases, \( V_{th}^{Q} \) is reduced and close to threshold Coulomb oscillations are observed, which look very regular and show a wider spacing at large \( V_{BG} \). A wide spacing could suggest formation of a single QD in the device, while many lines with a small separation could indicate formation of two QDs, one in each corner. Back-gate induced tuning between a single and double QD configuration is expected from simulations [223]. Formation of a single few-electron quantum dot at \( V_{BG} = 10 \) V under the gate of the quantum device is confirmed in Fig. 5.1(e), which shows clear Coulomb diamonds (at \( V_{WL} = 1.3 \) V). A first addition energy of about 16 meV is estimated, demonstrating strong confinement compatible with previous measurements [74, 222]. Figure 5.1(d) shows cross-sectional views along the gate and source-drain of the device at this operation voltage with a single QD well-coupled to both source and drain reservoirs. The back-gate is set to \( V_{BG} = 10 \) V in the following experiments to form a single QD memory cell where single electrons are stored and read out.

**RF characterisation**  Next, the \( LC \) resonant circuit is characterised by measuring the magnitude of the reflection coefficient \( |\Gamma| \) as a function of \( V_{WL} \) as shown in Fig. 5.2(a). Lowering of the resonance frequency is observed when the control FET is operated above threshold (\( V_{WL} > 0.63 \) V) due to the additional capacitance of the memory cell that appears in parallel to \( C_p \). Figure 5.2(b) shows the total capacitance \( C_T \) (assuming the nominal inductance \( L = 390 \) nH) and quality factor \( Q_{load} \) of the \( LC \) circuit obtained from Fig. 5.2(a), from which a contribution of the memory cell circuit to \( C_T \) of 105 fF is calculated. Additionally, there is a reduction of \( Q_{load} \) when the FET is in the on state. The quality factor and capacitance play an important role in the signal-to-noise ratio (SNR) of the phase response. Values of \( C_T \) and \( Q_{load} \) obtained for the memory cell are comparable to previous measurements [162], where direct readout is performed. To characterise the impact of FET on gate-based readout of the quantum device, the phase response of the resonant circuit is examined as a function of \( V_{DL} \) for multiple \( V_{WL} \) (see Fig. 5.2(c)), using an RF modulation at frequency \( f_{RF} = 313 \) MHz and power \( P_{RF} = -88 \) dBm. Starting with the control FET well above threshold (\( V_{WL} = 1.3 \) V), in the strong accumulation regime, three principal Coulomb peaks of single-electron transitions in the quantum device are observed when ramping \( V_{DL} \) (blue trace). The peaks remain
Figure 5.2: RF characterisation. (a) Magnitude of the reflection coefficient as a function of $V_{\text{WL}}$ (with $V_{\text{DL}} = 0.4 \text{ V}$ and $V_{\text{BG}} = 10 \text{ V}$). (b) Total resonator capacitance $C_T$ (with $L = 390 \text{ nH}$) and quality factor $Q_{\text{load}}$ as a function of $V_{\text{WL}}$. (c) Phase response at 313 MHz as a function of $V_{\text{DL}}$ for different $V_{\text{WL}}$ showing three Coulomb oscillations only when the control FET is operated above threshold. Traces are offset for clarity and features originating from charge transitions within the control FET itself are indicated as ($^\ast$). (d) Coulomb diamonds measured in the phase response ($V_{\text{BG}} = 10 \text{ V}$ and $V_{\text{WL}} = 1.3 \text{ V}$).
5.2 Conditional readout

initially visible as $V_{WL}$ is reduced, though a background signal begins to dominate as the control FET enters the weak inversion regime, where the FET gate capacitance strongly depends on the difference between $V_{WL}$ and $V_{DL}$ (which both determine the effective FET gate-source voltage). In this regime, the RF modulation and $V_{DL}$ ramp is picked up in the dispersive response of the resonator as an additional change in capacitance and in turn produces an additional phase shift. Eventually, when $V_{WL} < 0.5 \text{ V}$ the control FET is below threshold and the dispersive response vanishes (green trace). Additional features appear in the measurements (indicated by *) which depend on $V_{WL}$ and are attributed to single-electron tunnelling events in the control FET. These features become more apparent when operating the control FET close to threshold. Fig. 5.2(d) shows the phase response as a function of $V_{DL}$ and $V_{SD}$ (with the control FET well above threshold), showing Coulomb diamonds of the QD in the same voltage region as the transport measurements in Fig. 5.1(e). The correspondence between both sets of measurements shows that, in the strong accumulation regime, the FET channel has negligible impact on the RF readout.

5.2.2 Charge sensitivity

A measurement of the charge sensitivity provides a quantitative assessment of the impact of the FET circuit parasitic resistance and capacitance on gate-based read-out. To measure the charge sensitivity of the gate-based sensor with control FET, a small-amplitude signal of frequency $f_{AC} = 303 \text{ Hz}$ is applied onto the data-line (in addition to the drive at $f_{RF}$) and the SNR in dB of the side-bands appearing in the frequency spectrum at $f_{RF} \pm f_{AC}$ are monitored (see Section 3.2.5). The peak amplitude of the signal (0.2 mV) corresponds to a change of $\Delta q_{\text{rms}} = 0.00578e$ in the charge on the quantum dot, where $e$ is the charge of the electron. The side-band SNR is optimised by tuning the circuit parameters $V_{DL}$, $f_{RF}$ and $P_{RF}$ as shown in Fig. 5.3(a-c) respectively. The maximum in Fig. 5.3(a) at $V_{DL} = 0.525 \text{ V}$ corresponds to the point of maximum slope of a phase oscillation as shown in Fig. 5.2(c). The RF frequency dependence of the SNR, in Fig. 5.3(b), reveals a maximum at $f_{RF} = 313 \text{ MHz}$ and a 3 dB measurement bandwidth of 13 MHz which translates into a $Q_{\text{load}}$ of 24 in the on state of the control FET compatible with results shown in Fig. 5.2(b). The optimal value for the RF power $P_{RF}$ was found to be $-86 \text{ dBm}$. At these optimal operation points, a comparison of the power spectrum with the FET in the on ($V_{WL} = 1.2 \text{ V}$) and off ($V_{WL} = 0 \text{ V}$) state is shown in Fig. 5.3(d), which includes a fitting-model to obtain the side-band SNR. Finally, in Fig. 5.3(e) the SNR is shown as a function of $V_{WL}$ (with $V_{DL} = 0.525 \text{ V}$), where two plateaus
Figure 5.3: Charge sensitivity. Signal-to-noise-ratio (SNR) of a charge sensitivity measurement as a function of (a) data-line voltage $V_{DL}$, (b) carrier frequency $f_{RF}$, (c) carrier power $P_{RF}$. (d) Side-bands appearing around $f_{RF}$ when operating at parameter values of maximum SNR in (a-c). A comparison between control FET on and off and a fitting model to obtain the SNR is shown. (e) SNR as a function of $V_{WL}$. Regimes where the control FET is in the off (red squares), an intermediate (gray circles) and the on (green diamonds) state are identified. When not being swept, the following parameter values are used: $V_{WL} = 1.3$ V, $V_{DL} = 0.525$ V, $f_{RF} = 312$ MHz, $P_{RF} = -85$ dBm.
corresponding to the on and off state of the control FET can be identified. In the approximately linear transition between the plateaus, multiple scattered data points, attributed to transitions in the weak inversion regime of the FET, are observed (i.e. starred features in Fig. 5.2(c)).

Overall, using optimised operation parameters a SNR of 15.6 dB is obtained which translates into a charge sensitivity of $\delta q = \Delta q_{\text{rms}}/\left(\sqrt{2BW} \times 10^{\text{SNR}/20}\right) = 95 \, \mu\text{eHz}^{-1/2}$ for the chosen spectrum analyser bandwidth $BW = 50$ Hz. The charge sensitivity obtained in this experiment is lower than typical rf-QPC devices [156] and demonstrates more than a factor of 50 improvement compared to GaAs based gate sensors [168] and is only a factor of 2.5 higher than previously reported in a similar device without transistor circuit [162]. In this experiment the bandwidth of the charge sensitivity measurements was limited to 500 Hz due to low-pass filtering of the line used to deliver the sinusoidal signal $f_{\text{AC}}$. However, the bandwidth of gate-based reflectometry is limited by the $LC$ resonator bandwidth which is about 10 MHz when the resonator is coupled to the quantum device.

5.2.3 Electron retention time

For sequential readout of multiple quantum devices, one after another, the gate that is connected to the shared bias and readout circuit must be able to store charge. This charge storage is enabled by the control transistor which leaves the quantum device gate floating when operated in the off state. Such a deselected memory cell, should retain the charge long enough to allow subsequent readout of multiple other cells that are connected to the same bias and readout circuit. Charge retention measurements were performed in a different pair of devices, with nominally identical dimensions to the memory cell above.

To measure the electron retention time in this single memory cell, transport through the quantum device is monitored in a dynamic experiment, where the control FET is switched from the on to the off state while the data-line voltage remains constant. This pulse sequence is shown in Fig. 5.4(a) together with a simplified equivalent circuit of the memory cell. In this circuit, the cell capacitance, which has contributions from the quantum device gate capacitance $C_G$, the FET and the inter-connection, provides floating gate charge storage when the FET is off. Additionally, the FET resistance $R_{\text{FET}}$ forms a voltage divider for the data-line voltage $V_{\text{DL}}$ with the gate leakage resistance, $R_G$, that represents dielectric losses through the gate-oxide, which yields an effective voltage at the gate of $V_G = \frac{R_G}{R_{\text{FET}} + R_G} V_{\text{DL}}$. An estimate of $C_G$ is be obtained from the gate voltage spacing $\Delta V_{\text{DL}}$.
Figure 5.4: Retention time. (a) Equivalent circuit consisting of the variable control FET resistance $R_{\text{FET}}$, quantum device gate leakage $R_{G}$ and storage capacitance $C_{\text{cell}}$. (b) Voltage divider characteristic of the memory cell as a function of $V_{WL}$. (c) Demonstration of charge locking for different FET off states ($V_{WL}^i$). Slow leakage of quantum dot gate charge is observed. (d) Quantum device transfer characteristic with effective electron occupation indicated. (e) Coulomb diamonds obtained from a discharge measurements as a function of $V_{SD}$. 
between consecutive Coulomb blockade oscillations shown in Fig. 5.2(d). Using
\[ C_{G,n,n+1} = e/\Delta V_{DL}^{n,n+1}, \]
where \( n \) is the number of electrons in the dot with respect to an
unknown offset \( N \), \( C_{G}^{0,1} = 6.2\,\text{aF} \) and \( C_{G}^{1,2} = 7.0\,\text{aF} \) is obtained. This demonstrates
that the cell capacitance of \( C_{cell} = 105\,\text{fF} \) (extracted from Fig. 5.2(b)) is dominated
by the inter-connection and FET capacitance. In Fig. 5.4(b) the voltage divider
characteristic of the memory cell \( V_G/V_{DL} \) is shown which is obtained by tracking
the position of a Coulomb peak as a function of \( (V_{WL} - V_{DL}) \) in a measurement
as shown in Fig. 5.1(b). When \( V_{WL} < 0.5\,\text{V} \) the resistance of the control FET
channel becomes greater than the gate leakage in the quantum device. For very
large FET resistance (small or even negative \( V_{WL} \) ) \( V_G \) tends to zero. Figure 5.4(c)
shows the current through the quantum device \( I_{SD} \) (at \( V_{SD} = 5\,\text{mV} \) ) over time
with the FET being switched off at \( t = 0 \). For \( V_{WL}^{1} = 0.6\,\text{V} \) as the off state
voltage level, \( R_{FET} < R_G \) and the discharging of the gate capacitor occurs mainly
through the control FET channel. For a more resistive off state of the control
FET, as given by \( V_{WL}^{0} = 0.34\,\text{V} \), discharging of the gate capacitor occurs mainly
through gate leakage since \( R_{FET} > R_G \) and the steady-state voltage on the quantum
device gate \( V_G \) approaches zero. Comparing Fig. 5.4(c) with Fig. 5.4(d), which
shows a measurement of \( I_{SD} \) as a function of \( V_{SD} \), effective numbers for the electron
occupation can be assigned. Finally, Fig. 5.4(e) shows \( I_{SD} \) as a function of time
after switching the FET off for different \( V_{SD} \) at \( V_{WL} = 0.58\,\text{V} \). Electron numbers
are assigned and initially discharge happens fast while the last electron still remains
in the QD even after 5 s.

From the observed dynamics of the current in Fig. 5.4(c), the single-electron
retention time of the storage node is obtained through time lapses \( \Delta t^{n,n+1} \) between
successive Coulomb oscillations, obtaining \( \Delta t^{1,2} = 1450\,\text{ms} \) and \( \Delta t^{0,1} = 780\,\text{ms} \).
This retention time compares favourably to DRAM cells that operate at a refresh
time of 64 ms [241] and is well above the typical expected readout time of 100 ns
for gate-based reflectometry (see Section 4.3 or [56]) and the single qubit coherence
time of 28 ms in \( ^{28}\text{Si} \) substrates [104].

5.3 Sequential readout

In this section, sequential gate-based readout of two CMOS single-electron memory
cells using a single readout resonator and bias circuit is demonstrated. Furthermore,
guidelines for operation along with a discharge model of a deselected cell are given.
5.3.1 Frequency overlap

**Setup**  The sequential access circuit consisting of two CMOS single-electron memory cells (cell 1(2) in green(orange)) both connected to a single lumped-element RF resonator for readout and a single bias line is shown in Fig. 5.5(a). Each memory cell is made from two transistors referred to as $Q_i$ and $T_i$. $Q_i$ is a 60-nm-wide silicon nanowire transistor with a short gate length (25 and 30 nm for cell 1 and 2 respectively). Such devices are routinely used to trap single-electrons in QDs that form at the top most corners of the nanowire channel when operated in the sub-threshold regime at cryogenic temperatures [222]. Transistor $T_i$, the control FET, is a wider device with a channel width of 10 µm and gate length of 25 nm and 30 nm for cell 1 and 2 respectively. The four transistors are manufactured using fully-depleted silicon-on-insulator (FD-SOI) technology following standard CMOS processes (see Section 3.1.1). They are located on the same chip and are connected via bond wires as shown in Fig. 5.5(a) (indicated by blue wires).

The primary inputs of the circuit are given by the data and word-lines which are named in analogy with memory chips. Each cell has one word-line, with voltage $V_{WL_i}$, which connects to the gate of the control FET $T_i$ allowing control over the channel resistance. The data-line, with voltage $V_{DL}$, is shared among the two cells and allows control over the gate voltage on $Q_i$ conditional on the state of $T_i$. Additionally, a voltage applied to the silicon substrate, $V_{BG}$, acts as a back-gate. Switching $T_i$ to the on state while keeping all the remaining $T_j$ off allows for individual addressing of a single quantum device $Q_i$. Multiple devices can be addressed sequentially by timing the voltages on $T_i$ accordingly. A lumped-element $LC$ resonator is connected in parallel with the memory cells for readout using RF gate-based reflectometry. The natural frequency of the resonator $f_0$ is given by $f_0 = 1/2\pi \sqrt{L/C_T}$ where $C_T$ is the total capacitance of the system that includes, in particular, the state-dependent quantum or tunnelling capacitance of any quantum device (see Section 2.3) which is connected to the $LC$ circuit via the control FETs. The whole circuit is operated in a dilution refrigerator with a base temperature of 15 mK (Bluefors LD).

**RF characterisation**  To begin, the frequency dependence of the circuit’s reflection coefficient $S_{11}$ for the four possible states of the two control FETs is obtained and shown in Fig. 5.5(b). A dip in the reflection coefficient occurs when the resonator is driven at its natural frequency of oscillation. This frequency shifts towards lower values (by approximately 28 MHz) for each $T_i$ in the on state due to the additional
Figure 5.5: Setup and static RF characterisation. (a) Illustration of the sequential access circuit for gate-based RF readout including an optical microscope image that shows the wire-bonding of transistor devices to form two memory cells. A single high frequency line and readout resonator is connected to two cells (green and orange) consisting of one control FET, $T_i$, and quantum device, $Q_i$, per cell. $T_i$ enables selective readout of $Q_i$. Electrical connections made via bond wires are represented by blue lines. (b) RF response (in reflection $S_{11}$) of the circuit for different control FET states ($T_1$ - $T_2$). Spectra for addressing a single cell have been shifted down by 15 dB for clarity. (c) Enlarged view of the reflection coefficient in the on-off and off-on state configuration with a spectral overlap and resonance fits indicated (by dashed lines). (d) Phase response of $Q_1$ as a function of $V_{DL}$ for $V_{WL1} = 1.2$ V and $V_{WL2} = 0$ V. (e) Phase response of $Q_2$ as a function of $V_{DL}$ for $V_{WL1} = 0$ V and $V_{WL2} = 1.2$ V. The regions in grey highlight charge transitions that are selected for further measurements.
circuit capacitance introduced by an enabled cell. Most importantly for sequential

<table>
<thead>
<tr>
<th>FET state: T₁–T₂</th>
<th>( f_0 ) (MHz)</th>
<th>( \Delta f ) (MHz)</th>
<th>( Q_{\text{load}} )</th>
<th>( C_T ) (fF)</th>
<th>( C_{\text{cell}} ) (fF)</th>
<th>( C_{\text{FET}} ) (fF)</th>
<th>( C_S ) (fF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>off-off</td>
<td>645.9</td>
<td>0</td>
<td>96</td>
<td>740</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>on-off</td>
<td>618.3</td>
<td>27.3</td>
<td>39</td>
<td>808</td>
<td>68</td>
<td>4</td>
<td>64</td>
</tr>
<tr>
<td>off-on</td>
<td>616.7</td>
<td>29.2</td>
<td>40</td>
<td>812</td>
<td>72</td>
<td>5</td>
<td>67</td>
</tr>
<tr>
<td>on-on</td>
<td>593.0</td>
<td>52.9</td>
<td>28</td>
<td>878</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1: Two-cell spectral parameters. \( f_0 \) is the centre frequency, \( \Delta f \) is the resonant frequency shift, \( Q_{\text{load}} \) the loaded quality factor and \( C_T \) the total circuit capacitance obtained using a nominal inductance of 82 nH. \( C_{\text{cell}} \) is the capacitance added by a single cell being selected via a control FET which is composed of the FET channel capacitance \( C_{\text{FET}} \) and storage capacitance \( C_S \).

readout, a large spectral overlap of 13 MHz with 3 dB readout bandwidth is observed as shown in the enlarged view in Fig. 5.5(c), which shows the magnitude of the reflection coefficient \(|\Gamma|\) for the configurations of addressing one cell at a time. Spectral overlap is vital to dynamical multi-qubit readout as it means that both cells can be read using the same input frequency, while the degree of overlap determines the readout bandwidth of the architecture. Parameters of the LC resonator are extracted at each configuration (see Tab. 5.1) where estimations are made based on a nominal inductance of 82 nH and an expected geometric capacitance of the FET of

\[
C_{\text{FET}} = \frac{\varepsilon_r \varepsilon_0 (2H + W)L_g}{1.87 \cdot d_{\text{ox}}}, \tag{5.1}
\]

where \( H = 10 \) nm is the silicon wire height, \( W = 10 \) µm is the channel width, \( L_g = 30(25) \) nm is the gate length for \( T_1 \) (\( T_2 \)) and \( d_{\text{ox}} = 1.3 \) nm is the equivalent oxide thickness. The factor of 1.87 arises from the fact that the 2DEG barycentre is found 1.1 nm beneath the interface. The storage capacitance \( C_S \) has contributions from the quantum device gate \( C_G \) and the inter-connection, with the latter dominating.

In addition to the resonance frequency shift, a reduction in the loaded quality factor \( Q_{\text{load}} \) from a value of 96, when both \( T_1 \) or \( T_2 \) are in the off state, to a value of \(~ 40\), when either \( T_1 \) or \( T_2 \) are in the on state. An on state \( Q_{\text{load}} \) of 40 is comparable to previous experiments without control circuit [162].
**Single frequency readout** Based on the spectra shown in Fig. 5.5(c), a carrier frequency $f_{RF} = 615$ MHz is selected to probe the state of the quantum devices. Changes in the capacitance of the quantum device $\Delta C_G$, attributed to tunnelling of single electrons, are detected through changes in the reflected phase $\Delta \phi \approx -2Q_{load}\Delta C_G/C_T$ (see Section 2.3). In Fig. 5.5(d,e), peaks in the phase are observed as a function of $V_{DL}$ that corresponds to regions of charge instability in $Q_i$, where single electrons cyclically tunnel between the QDs in the channel and the source or drain electron reservoirs. For each measurement only one $T_i$ is set to the on state while the other is off and the same RF frequency is used to perform readout. In the next section, the effect of turning the control FETs $T_i$ off is studied by focusing on a particular region of the stability diagrams for both quantum devices, highlighted in grey in Fig. 5.5(d,e), with the aim to find optimal operation voltage levels for the control transistors.

### 5.3.2 Operation guidelines

For a dynamical random-access readout scheme, $T_i$ should fulfil several requirements: In the on state, $T_i$ should be sufficiently conductive to allow high-sensitivity gate-based readout of the selected quantum device. In the off state, $T_i$ should be sufficiently resistive to block the RF signal towards deselected cells and retain the charge on $Q_i$’s gate for the time operations are being performed in other cells. Guidelines on the control transistor operation voltage are developed to fulfil these requirements.

**Transistor operation voltage** As a first step towards dynamically operating the circuit, suitable on and off state voltages for the control FET gate (i.e. the high, $V_{WLI}^H$, and low, $V_{WLI}^L$, signal levels) are identified. In Fig. 5.6(a,b), the phase of the reflected signal from the resonator as a function of $V_{DL}$ and $V_{WLI}$ is shown. Three regions can be identified: The on region for $V_{WLI} > 0.9$ V, where single electron tunnelling is observed, the off region for $V_{WLI} < 0.7$ V, where no transitions are observed and finally, for $0.7$ V < $V_{WLI}$ < $0.9$ V the forbidden region. In the latter, $T_i$ is in the depletion regime, where, due to the voltage-dependent gate capacitance of the control FET, the phase varies largely [242]. This region should be avoided when assigning voltage levels. To highlight the different response of the resonator in the digital on and off states, the phase change $\Delta \phi$ as a function of $V_{DL}$ for cell 1 and 2 is shown in Fig. 5.6(c,d), respectively, at $V_{WLI}^{L(H)} = 0.5(1.2)$ V. There is a close similarity between the operation voltage levels of both transistors $T_i$ for addressing
the quantum devices $Q_i$ at milliKelvin temperature. In a scaled up architecture, with increasing circuit complexity, reproducible electrical characteristics between cells will be essential.

### Discharge model

As already observed in the previous single-memory cell experiment (see Section 5.2.3), the voltage on the QD gate $V_G$ decays over time when the FET is switched to the off state. In order to determine an appropriate voltage refresh rate and provide operation guidelines to maximise retention time a discharge model is developed. A simplified equivalent circuit model of the memory cell is considered as shown in Fig. 5.7(a). It consist of the FET off state channel resistance $R_{FET}$, the gate leakage resistance $R_G$, and the cell capacitance $C_{cell}$. $R_G$ combines the FET and QD gate leakages and $C_{cell}$ is the parallel sum of the QD gate capacitance $C_G$ and the interconnect and storage capacitance $C_S$. The discharge of the cell is studied in a pulsed experiment by measuring the source-drain current $I_{SD}$ through the quantum device over time. As shown in Fig. 5.7(a), $V_{DL} = 0.68\ \text{V}$ is constant while $V_{WL}$ switches from the high level ($V_{WL}^H$) to a low level ($V_{WL}^L$) at $t = 0$. The pulse amplitude is set to 0.5 V and the pulse offset is varied ensuring
5.3 Sequential readout

Figure 5.7: Memory cell discharge model. (a) Equivalent circuit of a single memory cell and pulsing scheme for charge retention analysis. (b) Source-drain current $I_{SD}$ through the quantum device as a function of time after switching the control transistor to the off state $V_{WL} = 0$ V ($V_{SD} = 2$ mV, $V_{DL} = 0.68$ V). Single electron transitions are observed and peak positions are indicated by $\star$. The inset shows $I_{SD}$ as a function of $V_{DL}$ where the same transitions are observed. (c) Decay of the voltage on the QD gate $V_G$ as a function of time once the control transistor is switched off. Data-points (circles) are obtained from the peak positions ($\star$) in $I_{SD}$ and solid lines are fits to a double exponential function, see Eq. (5.2). (d) Top panel: Quasi-static gate voltage $V_{\text{final}}$ and time constant $\tau$ as a function of $V_{WL}^0$ obtained from the exponential decay fits. Bottom panel: $R_{\text{FET}}$ and $R_G$ extracted from $\tau$ and $V_{\text{final}}$. Dashed lines are guides to the eye.
that the transistor remains on in the high part of the pulse. An exemplary discharge measurement for $V_{\text{WL}} = 0.5\, \text{V}$ is shown in Fig. 5.7(b), where several single electron transitions (indicated by ⋆) can be observed in $I_{\text{SD}}$ over time. After 1.5 s the current appears to settle to a fixed value. The discharge data is compared with a measurement of the same single electron transitions of the device as a function of $V_{\text{DL}}$ in quasi-static conditions as shown in the inset of Fig. 5.7(b). By matching peaks in the decay over time to peaks as a function of $V_{\text{DL}}$ the dynamics of the voltage on the QD gate $V_G(t)$ is reproduced as shown in Fig. 5.7(c) for multiple values of $V_{\text{WL}}$. At $t = 0$, an initial fast decay is observed, possibly due to charge-injection and clock-feedthrough [243], followed by a slow decay that is attributed to the discharge of the cell. A fitting function that captures both the initial fast decay and the discharge

$$V_G(t) = V_{\text{final}} \left[ 1 + \frac{R_{\text{FET}}}{R_G} \exp \left( -\frac{t}{\tau} \right) \right] + V_{\text{fast}} \exp \left( -\frac{t}{\tau_{\text{fast}}} \right)$$  \hspace{1cm} (5.2)

is applied (lines in Fig. 5.7(c)) to extract $\tau$, $V_{\text{final}}$, $R_{\text{FET}}$, and $R_G$. In this simplified circuit model $\tau = \frac{C_{\text{cell}} R_G R_{\text{FET}}}{R_G + R_{\text{FET}}}$ is the circuit time constant and $V_{\text{final}} = \frac{V_{\text{DL}} R_G}{(R_{\text{FET}} + R_G)}$ is the equilibrium voltage at the gate of the QD at $t \rightarrow \infty$, and $V_{\text{fast}}$ and $\tau_{\text{fast}}$ are the amplitude and time constant of the initial fast decay. Since $\tau$ and $V_{\text{final}}$ depend on $R_{\text{FET}}$, and thus on the operation voltage level $V_{\text{WL}}$, the functional dependence of $\tau$ and $V_{\text{final}}$ on $V_{\text{WL}}$ is characterised to find the optimal voltage operation point that maximises the charge retention time. As expected and shown in Fig. 5.7(d), with increasing $V_{\text{WL}}$ ($R_{\text{FET}}$ decreases) $V_{\text{final}}$ becomes larger due to the voltage divider characteristic of the cell. For $\tau$, a reduction from 0.9 s to 0.2 s is observed in the same range of $V_{\text{WL}}$. Values of the resistance $R_{\text{FET}}$ and $R_G$ extracted from these measurements based on a cell capacitance $C_{\text{cell}} = 70\, \text{fF}$ (see Tab. 5.1), which are on the order of $10^{13}\, \Omega$. To summarise, the parameters extracted from the slow discharge model follow the expected trend and show a decrease/increase in $\tau(V_{\text{final}})$ as $V_{\text{WL}}$ is increased from 0.45 V to 0.55 V. Moreover, $R_G$ and $R_{\text{FET}}$ both show a $V_{\text{WL}}$ dependence and decrease by a factor of 3 and 5 respectively as $V_{\text{WL}}$ increases.

For the initial fast decay, a time constant on the order of one millisecond and an amplitude of $\approx 0.1\, \text{V}$ is obtained, which can be explained by charge injection and clock-feedthrough. There are limited data points available directly after $t = 0$, since oscillations in the discharge need to be matched to oscillations in the turn on characteristic, such that the extracted fast time constant should be treated as an upper bound. To estimate the contribution of charge injection, the amount of charge in the control FET channel, which escapes onto the quantum device gate when the charge
5.3 Sequential readout

in the FET is depleted, is calculated: \( \Delta V_{\text{CI}} = \frac{Q_{\text{channel}}}{2C_{\text{cell}}} = \frac{C_{\text{FET}}(V_{\text{HWL}} - V_{\text{DL}} - V_{\text{th}})}{2C_{\text{cell}}} \approx 0.02 \, \text{V} \). Here, \( C_{\text{FET}} \) is the gate capacitance of the FET and \( C_{\text{cell}} \) is the cell capacitance and the factor of 1/2 accounts for the assumption that the charge under the gate escapes equally to the source and drain. Clock-feedthrough can be estimated based on the capacitance between the control FET gate and source \( C_{\text{GS}} \). Due to unconventionally long gate spacers (25 nm), designed to create tunnel barriers to the source and drain, the contribution from direct overlap is assumed negligible. Fringing gate capacitance is likely to be the dominant parasitic contribution to clock-feedthrough. A fringing capacitance of 33% of the FET gate capacitance is observed for transistors of dimensions \( W = 0.15 \, \mu\text{m} \) and \( L_g = 60 \, \text{nm} \) [244]. Thus, for a transistor of dimensions \( W = 10 \, \mu\text{m} \) and \( L_g = 30 \, \text{nm} \) \( C_{\text{GS}} \approx C_{\text{FET}}/2 \) is estimated, which results in a voltage shift of \( \Delta V_{\text{CF}} = \frac{C_{\text{GS}}}{C_{\text{GS}} + C_{\text{cell}}} (V_{\text{HWL}} - V_{\text{WL}}) = 0.013 \, \text{V} \). Charge injection combined with clock-feedthrough yields an expected voltage shift of \( \Delta V \approx 0.033 \, \text{V} \). The amplitude of the fast decay observed in the experiment is more than twice as large as the combined estimate. The amount of injected charge in the experimental setting could be larger when more than half of the charge escapes towards the quantum device and could also depend on the specific on and off state word line voltages. Further cross-talk in the lines, PCB as well as the device could have an additional contribution to the fast decay. To minimise the effect of such cross-talk on the analysis of the slow decay, the pulse amplitude in the measurements presented in Fig. 5.7 is kept constant and the offset is varied to produce different low levels while keeping the high level above threshold. The voltage shift due to charge injection and clock-feedthrough can be reduced for smaller \( C_{\text{FET}} \) and larger \( C_{\text{cell}} \) and charge injection can be mitigated using a slow fall time of the clock. Moreover, charging effects can be compensated using NMOS and PMOS control FETs in parallel, where the geometry of both can be optimised such that a negative voltage jump introduced when switching off the NMOS FET is compensated by a positive voltage change due to switching of the PMOS FET. Typically, a PMOS transistor of larger width is required due to the lower mobility. Alternatively, a dummy transistor in series with the control FET, with approximately half the transistor width, operated with an opposite clock allows for compensation by storing the additional charge. Furthermore, additional cross-talk can be reduced by using coaxial lines to deliver control FET signals and careful PCB and chip design.

**Guidelines** After characterisation of transistor threshold levels and discharge, guidelines for operating the sequential access circuit with the aim to maximise charge retention are given. While initially it may seem beneficial to select a low
$V_{WL}^r$ level to maximise $\tau$, one needs to consider that the retention or refresh time is determined by the maximum tolerable gate voltage drop of the cell, $\delta V$. For an optimised circuit with negligible clock-feedthrough and charge injection the retention time is given by $t_r = R_G C_{cell} \ln [(1 - ar)^{-1}] / r$ where $a = \delta V / V_{DL}$ is the voltage drop ratio and $r = (R_{FET} + R_G) / R_{FET}$ the resistance ratio. $t(r)$ is a monotonically increasing function of $r$, given $R_G$ varies weakly with $r$, which means that $t_r$ is maximised by operating at large $V_{WL}^r$ while remaining in the off state regime, where RF readout of the deselected cell is blocked and readout of another selected cell is not disturbed. The minimum off state resistance that fulfils these requirements is estimated using an equivalent circuit of the two memory cells as shown in Fig. 5.8(a). From circuit simulations (see Appendix G and Section 6.3.2.3) with an equivalent circuit model as shown in Fig. 5.8(a) and experimentally extracted values, the impact of the off state resistance of the second cell $R_{FET2}$ on the sensitivity of the reflection coefficient to capacitive changes in the first cell $|\partial \Gamma / \partial C_{G1}|$ is estimated. The simulation in Fig. 5.8(b) shows that as long as $R_{FET2} > 10 \text{ M} \Omega$, the effect of a deselected cell on the readout of a selected cell is negligible. This is compatible with operating closely below the forbidden region shown in Fig. 5.6. Additionally, the fraction of the RF signal delivered to the deselected cell for a given $R_{FET2}$ is estimated. For $R_{FET2} > 10 \text{ M} \Omega$ the fraction of the RF signal delivered to the deselected cell is $< 1\%$ due to the low-pass filter formed by $R_{FET2}$ and $C_{S2}$ (assuming $C_{S2} = 64 \text{ fF}$ and an RF frequency at a few hundreds of MHz) such that the deselected cell is not disturbed.

**Figure 5.8: Effect of a deselected cell.** (a) Equivalent circuit used to perform simulation of the impact of the FET $R_{FET2}$ resistance of a deselected cell (2, in gray) on the sensitivity to changes in the gate capacitance $C_{G1}$ of a selected cell (1, in black). (b) Sensitivity $|\partial \Gamma / \partial C_{G1}|$ of readout of the selected cell as a function of the off state resistance of a second deselected cell. There is negligible impact up to a resistance of $10 \text{ M} \Omega$. 
Sequential readout

In the previous sections, guidelines on operation parameters for sequential readout have been determined using static and dynamic measurements of the two memory cell circuit. In this section sequential readout is demonstrated using $V^{L}_{WL} = 0.5$ V ($R_{FET} \approx 10^{13}\ \Omega$) and $V^{H}_{WL} = 1.2$ V to enhance the retention time in the off state while preserving good noise margins.

The pulsing scheme to dynamically read both memory cells is shown in Fig. 5.9, which shows the operation voltages and the phase response as a function of time for three sequential readout cycles. In the first half of the cycle, from 0 to 12 ms,

Finally, given a voltage drop ratio of 1%, a retention time of 20 ms is expected using the parameters extracted from the experiment, which is much larger than the coherence time $T^{*}_{2}$ of silicon-based electron spin qubits [91]. In this analysis, it is important to note that $V_{DL}$ is assumed to be constant, which is approximately what will happen when addressing multiple quantum devices with similar operating voltages.

**Figure 5.9: Sequential readout.** Pulse scheme for sequential readout and phase response of $Q_{1}$ and $Q_{2}$. $V_{DL}$ is ramped up and down while $V_{WL1}$ and $V_{WL2}$ are alternating between on and off states. Pulses are synchronised such that QD reservoir transitions from $Q_{1}$ are obtained when $V_{DL}$ is ramped up while $Q_{2}$ is measured when $V_{DL}$ is ramped down.
Figure 5.11: Sequential stability diagram. Stability diagram in the differential phase response of (a) $Q_1$ and (b) $Q_2$ as a function of $V_{DL}$ and $V_{BG}$. Both maps have been obtained sequentially as illustrated in Fig. 5.9 and by additionally stepping $V_{BG}$.

$T_1$ and $T_2$ are set to the digital on and off states respectively. Simultaneously, an analogue ramp signal is applied to the common data-line $V_{DL}$ (blue trace) which is connected to $Q_1$ in this half of the cycle.

Peaks are observed in the phase response due to single-electron transitions between a QD and a reservoir in $Q_1$. In the second half of the cycle, from 12 to 24 ms, the digital voltages on $T_1$ and $T_2$ are inverted such that transitions in $Q_2$ are now detected as the analogue signal ramps down the voltage on the data-line. The QD-to-reservoir transitions in the phase response are identical to those measured in a static experiment shown in Fig. 5.6(c,d). The RF frequency and amplitude is kept constant throughout the measurement and the change in phase $\Delta \phi$ rather than the absolute phase is presented due to a phase offset between the signal detected from $Q_1$ and $Q_2$ originating from a small difference in reflection coefficient between the cells (see Fig. 5.5(c) and Fig. 5.6(a,b)). A signal-to-noise ratio (SNR) of $10^5$ with 100 ms integration time is obtained. The maximum fre-
The frequency at which sequential readout can be performed in this implementation is 1 kHz. This is shown in Fig. 5.10 which shows sequential measurements as shown in Fig. 5.9, where $V_{DL}$ is ramped up and down while $V_{WL1}$ and $V_{WL2}$ are alternating between on and off states at the same frequency. Multiple traces represent measurement for increasing frequency and traces are offset for clarification. For measurements above 1 kHz a significant increase in the rise time of the pulses is observed, representing the frequency limit for sequential readout in this experiment due to low pass filtering of the lines used to deliver the control signals. By using coaxial lines for delivering $V_{WL1}$ and $V_{DL}$ the operational frequency could be increased until the limiting factor becomes excessive device heating caused by fast switching. Using this interleaved pulsing scheme for sweeping $V_{DL}$ combined with additional stepping of $V_{BG}$ after each cycle, charge stability maps of both $Q_1$ and $Q_2$ are obtained sequentially as shown in Fig. 5.11(a,b). The transitions observed in the measurement suggest formation of multiple QDs in both cells.

### 5.4 Power dissipation

Dynamic power dissipation is an important parameter to estimate when switching control FETs for (de)selection of devices in a random access architecture that operate in a dilution refrigerator. The dynamic power dissipated when switching one FET $T_i$ is $P = C_{FET} f_{op} \Delta V^2$, where $\Delta V = V_{WL}^H - V_{WL}^L$, $C_{FET}$ is the overall capacitance subject to the voltage swing $\Delta V$ and $f_{op}$ is the switching frequency. The maximum $f_{op}$ is given by the readout bandwidth of the architecture and $C_{FET}$ is estimated using geometric capacitance of $T_i$, see Eq. (5.1). From the experimental implementation $\Delta V = 0.7 \, \text{V}$, a readout bandwidth of 13 MHz and $C_{FET} = 4 \, \text{fF}$ is obtained. This yields a power dissipation of 25 nW/cell which allows operating 16,000 cells simultaneously at the maximum switching frequency when pulses are generated at milliKelvin. The power dissipation can be reduced by pushing the voltage levels closer to the forbidden region (e.g. $\Delta V = 0.4 \, \text{V}$), thus reducing the noise level margins, or using smaller control FET dimensions. If pulse generation is not be feasible at milliKelvin temperature based on the budget, pulses could be generated at a higher stage (e.g. 4 K) where higher cooling power is available, however this might lead to an increased capacitance that is subject to the voltage swing.
5.5 Conclusion

The concept of a single-electron memory cell has been introduced for conditional and sequential gate-based readout of multiple quantum device at the same frequency. High-sensitivity readout combined with single-electron retention times on the order of one second have been demonstrated. Based on this sequential readout of two cells up to 1 kHz, limited by the lines delivering the control signals, has been demonstrated at control FET voltage levels that enhance retention. Assuming a cell operation time of $1 \mu$s the demonstrated retention time would allow addressing of $10^6$ cells before the voltage on the initial node needs to be refreshed, and estimations of dynamic power dissipation suggest that of up to 16,000 control FETs can be switched at maximal readout bandwidth (13 MHz) without introducing heating of the quantum devices demonstrating the prospects of a random-access scheme for implementing a scalable readout architecture.
Chapter 6

Scalable architectures

Many challenges arise when scaling to a large number of qubits. Not only within the qubit fabrication and qubit architecture itself, but also within the classical control and readout system. In this Chapter, error-correction and integration with classical electronics is introduced. Furthermore, a surface code architecture using mediated exchange and routes towards a fully integrated and scalable readout architecture are presented.

The author acknowledges contributions from Z. Cai and Prof S. C. Benjamin for working out surface code operation and performing error threshold simulations of a surface code architecture based on multi-electron mediators and S. Patomäki and M. A. Fogarty for assistance in providing necessary physical details. Dr M. F. Gonzalez-Zalba for providing a large-scale readout architecture schematic. V. N. Ciriano-Tejel and T.-Y. Yang for low-temperature transistor characterisation and V. N. Ciriano-Tejel for transistor modelling and assistance in circuit simulations.

6.1 Motivation

Silicon-based qubit demonstrators are only at the few qubit level [36], however first complex circuits have been fabricated [37] and the small footprint of devices and compatibility with CMOS processes could allow for scaling to processors containing millions of devices on millimetre size chips. Quantum processors require 100s to 1000s of perfect (ideal) qubits in order to solve the most challenging problems [5]. However, to operate such processors extremely precise control of each qubit is required, as small error probabilities on each qubit rapidly accumulate into significant computational errors. This makes the implementation of error correction schemes inevitable, where repetition codes are one of the simplest solutions, while the surface code offers high error thresholds up to 1% in an attractive two-dimensional
layout with nearest-neighbour coupling only [136]. Reducing requirements on control precision in this way comes at the cost of encoding a logical qubit using multiple physical qubits and potentially more than $10^8$ physical qubits are required to perform non-trivial calculations.

This large amount of qubits brings extraordinary challenges in qubit addressing, control and readout and integrated electronics could provide a solution to these problems. Some of the challenges that face large-scale quantum computing resemble those that have already been solved for conventional computing and integrated electronics could allow for efficient signal routing and generation, data flow management, low-level feedback and high-level operations locally. Therefore, to relax wiring requirements and reduce the latency of solid-state quantum computers, the integration of conventional electronics with quantum devices at cryogenic temperatures could be a promising strategy [245, 246]. However, to apply this approach, it is vital to understand the behaviour of integrated circuits at cryogenic temperatures [247].

### 6.2 Error correction architectures

In this section, the surface code architecture and physical implementations using donor and QD based spin-qubits are introduced. Schemes compatible with a planar and nanowire fabrication approach are discussed and a surface code architecture based on multi-electron QD mediated exchange is presented.

#### 6.2.1 Surface code

The surface code describes an error-correction scheme that is based on a two-dimensional array of qubits with nearest neighbour interactions only [136]. These simple requirements combined with a large error threshold, that is close to 1\% for individual operations, make the surface code an attractive architecture for the implementation of a fault-tolerant quantum computer.

In the surface code, qubits are separated into data (open circles) and measurement (ancilla, filled circles) qubits as shown in Fig. 6.1(a). During a computation, the collective state of a group of data qubits is repeatedly checked using so called stabilizers. Two different types of stabilizers, $X$ and $Z$, are performed simultaneously and each data qubit is coupled to two $X$ and two $Z$ stabilizer from which the parity of the data qubits is obtained (in two different bases) by measuring the ancilla qubits. A negative outcome of a stabilizer (change in parity) infers an error
6.2 Error correction architectures

Figure 6.1: Surface code. (a) Two-dimensional array of qubits that implements a logical qubit in a surface code architecture with physical qubits separated into data (open circles) and measurement (filled circles) qubits. Measurement qubits perform either $Z$ (green) or $X$ stabilizer parity measurements on neighbouring data qubits. On the boundaries, measurement qubits only interact with a reduced number of data qubits. (b) Generation of an additional logical qubit by disabling a stabilizer measurement which leaves a defect. Logical operations are performed using a chain of physical operation that commute with the parity measurements and typically stretch from one boundary to another or along a boundary as indicated by $\hat{Z}_L$ and $\hat{X}_L$. (c) Circuit for implementation of $Z$ and $X$ stabilizers. Reproduced from [136].

happened to one of the qubits. Bit-flip ($X$), phase-flip ($Z$) error and combinations thereof can be detected (Pauli errors) and from the outcome of a round of stabilizer measurements (the syndrome) specific errors that occurred during the operation can be inferred. Errors in the surface code do not necessarily need to be corrected physically using imperfect gates. A more efficient approach is to keep track of errors and perform a change into an error-free reference frame. As single qubit gates are Pauli operations themselves, the stabilizer measurements also protect the surface code from intentional qubit manipulations. However, physical operations that stretch from one boundary of the lattice to the opposite commute with the stabilizer measurements and realise logical $\hat{Z}_L$ and $\hat{X}_L$ operation on the encoded qubit. Similarly, additional logical qubits can be encoded by creating new boundaries in the array by turning off some of the stabilizer measurements, which creates holes (defects) in the lattice as shown in Fig. 6.1(b). Moreover, defects can be moved around in the lattice which can be used to generate logical two-qubit gates using a braiding transformation or lattice surgery [248].
The key requirement to implements a surface code architecture are high fidelity single and two-qubit gates within the error threshold and an efficient implementation of stabilizer measurement combined with high-fidelity readout. A stabilizer is a parity measurement that involves one measurement and four data qubits as shown in Fig. 6.1(c). \( Z \) stabilizers are performed by initialising the measurement qubit in the ground state followed by CNOT operations with each data qubit and measurement in the \(|\pm\rangle\)-basis with outcome of either even or odd parity (\( \pm 1 \)). \( X \) stabilizers are performed using an identical circuit but transforming into the \(|\pm\rangle\)-basis by embedding the CNOT operations within Hadamard gates on the ancilla. An alternative implementation is to initialise the ancilla into the \(|+\rangle\) state and perform CZ operations rather than CNOT with each data qubit. Finally, the surface code requires a classical processor that keeps track of the stabilizer outcomes and analyses the results and offers fast feedback (low latency).

### 6.2.2 Proposed surface code implementations

In this section, silicon-based surface code architectures, that have been proposed in literature are introduced. In 1998, Loss-DiVincenzo and Kane presented two approaches realising a scalable architecture for spin-based quantum computation using QDs or donors, which sparked the field of silicon-based quantum computing. Both approaches are in principle compatible with extension into a two-dimensional array allowing for implementation of a surface code architecture.

**Loss-Divincenzo** The Loss-DiVincenzo architecture encodes qubits using single spins confined in QDs subject to an external field. Single qubits are manipulated via ESR (\( B_{ac} \)) and two-qubit gates are performed via the electron spin exchange interaction [16, 249, 250]. Single-qubit addressability can be achieved via local Stark shifts, \( g \) factor or magnetic field gradients. Figure 6.2(a) shows an illustrative device where QDs are formed in a heterostructure using top-gates. Local back-gates combined with a magnetised (or high-\( g \)) layer provide addressability for global ESR single qubit control. Two qubit control is achieved electrically, as illustrated by the right QD pair that is brought closer together to initiate exchange. Following subsequent variants (see Section 2.2), qubits could be defined using more than one spin, such that spin manipulation could be done all-electrically rather than using a global field.
Kane In the proposal by Kane [251], both the nuclear and electron spin of $^{31}\text{P}$ donors participate in the computation. A single qubit is formed using the nuclear spin $I = 1/2$ states which are split in an external magnetic field and manipulated using NMR techniques. The control frequency of each donor can be tuned electrically, based on the hyperfine interaction with the electron spin, using the green control gates directly above each donor as shown in Fig. 6.2(b). By transferring the spin information from the nuclear spin to the electron spin, a two-qubit gate based on the exchange interaction between neighbouring donor electron spins is performed. Exchange is controlled using red gates at the surface. A donor separation of about 20 nm is required to reach sufficient electron spin exchange coupling. Kane’s architecture makes use of the long nuclear spin coherence for single qubits combined with interactions of extended range between donor electron spins.

Planar Si-MOS technology seems the most natural way to implement extensions of the Kane and Loss-DiVincenzo architecture into a two-dimensional array. However, strong technological requirements are imposed, as both are based on direct exchange between individual QD or donor spins with a characteristic interaction range of $< 100$ nm and $< 20$ nm respectively. Challenges arise in integrating the number of metal gates typically required to achieve control and confinement as well as providing a sufficient amount of local electron reservoirs for initial tuning and reset of such a dense large-scale device. In a nanowire based device, electron confinement and QD formation is typically achieved using a smaller amount of metal gates due to the natural confinement at the edges of the nanowire, which could be beneficial for realising a large-scale architecture. Steps towards implementing surface code silicon-based architectures including designs that specifically suit nanowire or planar implementations, and mechanisms that aim to relax technological requirements are all discussed in the following paragraphs.
Linear arrays Linear arrays can serve as a test-bed for error-correction concepts as they allow implementation of a repetition code, where a logical qubit is encoded using multiple copies of a physical qubit. In such devices, the complexity of operating a logical qubit could be studied on a small scale which seems experimentally achievable in the near future as first linear arrays have been fabricated (see Fig. 6.6(a)) and tuning and operation of arrays of QDs is explored and automated [37, 38, 252–256]. Tuning of such devices consisting of many gates can be time consuming and cross-capacitances between close-by gates make it a complex problem, which is tackled using virtual gates and automated using image filtering, feature recognition and machine learning techniques. Figure 6.3(a) shows an implementation of a logical qubit made from 14 QDs intended for a planar architecture, however SOI nanowire based devices, as shown in Fig. 6.3(b,c), also provide an excellent basis for linear arrays when extended to more gates. In nanowire QDs strong gate-coupling is observed (see Chapter 4) which facilitates fast readout using gate-based sensors. In a QD array consisting of just five QDs, a three qubit repetition code with singlet-triplet ancilla for parity measurements could be demonstrated. 14 QDs provide a minimal design for the 9-qubit Shor code [257]. Details on the implementation can be found in Ref. [133].

Split-gate The nature of the corner dots that can be formed in nanowire-based structures (see Section 3.1.1) provides a simple way to form singlet-triplet qubits. However, when using a single gate to form both QDs which wraps around the nanowire, achieving control over both QDs is challenging. In a split-gate design [258] as shown in Fig. 6.3(c) and Fig. 6.4(a), each QD can be controlled using a separate gate voltage as demonstrated in Fig 6.4(b) that shows a reflectometry measurement as a function of $V_{G1}$ and $V_{G2}$. In this measurement features typical for a coupled
DQD are observed (c.f. Fig. 2.4) and each corner QD couples mostly to the gate under which it is being formed. A singlet-triplet split-gate architecture additionally allows separation of signals for control from signals for readout as shown in Fig. 6.4(a). Control and readout signals can have different bandwidth requirements. A QD gate that is intended for gate-based readout is connected to a resonator that typically operates at sub-1-GHz frequencies with a resonator bandwidth reaching from 15 MHz to less than 1 MHz depending on the resonator quality factor. The bandwidth of control pulses delivered via such a readout gate is limited to the resonator bandwidth. Faster pulses are affected by the rise time and result in resonator ringing. One way to increase the available control signal bandwidth is to form resonators that operate at higher frequency where a larger bandwidth can be achieved. Alternatively, in the split-gate architecture a separate control line can be designed to deliver control signals at any desired frequency.

Floating gate The nanowire and especially the split-gate architectures allows for implementation of a two-dimensional architecture compatible with the surface code based on floating gate mediated electrostatic interaction. Spin-spin coupling via floating gates has been proposed for spin qubits based on single spins and two spins (singlet-triplet) [259]. In the case of a single spin, sufficient spin orbit coupling (SOC) is required such that the coupling to electric fields is achieved, while a singlet-triplet qubit couples to electric fields directly. SOC is weak in silicon based...
Figure 6.5: Floating gate. (a) Two-dimensional architecture of single electron spins confined in nanowires (black). Interactions between spins are mediated using floating gates (white, dog-bone) that can be activated selectively. Reproduced from [259]. (b) Possible implementation of a floating gate architecture in a SOI nanowire design where QDs form in both corners of the nanowire and spin-spin interactions are realised using direct exchange within a nanowire and floating gates from one nanowire to another.

qubits but electric manipulation of electron spins, which similarly requires sufficient SOC, has been demonstrated in SOI nanowire QDs where SOC is enhanced due to reduced symmetry [260]. A floating gate two-dimensional architectures using single electron spins proposed in [259] is shown in Fig. 6.5(a). In this architecture, the electron can be brought close to or under specific dog-bone shape couplers that mediate electrostatic interactions between two electron spins. The corner dots that form in gated SOI nanowires allow formation of such an architecture, with cells consisting of four QDs each that are occupied by a single electron as shown in Fig. 6.5(b). Within a nanowire, interaction between QDs could be based on direct exchange (blue box) in contrast to an all floating gate architecture as shown in Fig. 6.5(a). The QDs could be accumulated using additional top and back-gates and the floating gates mediate the interaction from a cell in one nanowire to a cell in another nanowire (see green box). As a single electron has four different QD locations to choose from, interactions to any neighbouring cell can be selectively enabled or disabled. Shuttling of the electron between these positions might require additional gates or the ability to provide floating gates with gate voltages on-demand (not shown). Alternatively,
instead of using singly occupied cells, doubly occupied cells allow for floating gate coupling that does not require any SOC and gate-based readout based on Pauli spin blockade can be performed. In case of doubly-occupied cells, interactions within a nanowire could still be achieved using direct exchange or capacitively.

**Crossbar**

Based on the assumption that there is a sufficient homogeneity between individual qubits, a high degree of shared control lines can be introduced which in turn reduces the complexity of fabrication (quadratic reduction of control lines) [51, 52]. In contrast to latest generation complex multi-QD devices with multiple control gates per QD (plunger and barriers) as shown in Fig. 6.6(a), a crossbar architecture realises individual addressing using shared lines in a crossed geometry where typically a specific point in a grid is selected using a combination of row and column. An implementation of a crossbar architecture for QDs in silicon [52], where every second QD is occupied by an electron spin, is shown in Fig. 6.6(b,c). The architecture uses a three-layer gate design for shared plunger (QL), column (CL) and row (RL) barrier gates. In such a design, it is assumed that a single voltage for each shared gate can tune all QDs into the desired single electron regime and sufficient tunability of the tunnel barriers can be provided. The first barrier layer is made of a superconducting material and carries a current to provide addressability for control using global ESR. Two-qubit gates and readout are performed by shuttling the selected electrons into the same column and row respectively. Readout is performed dispersively (gate-based) using Pauli spin blockade (no reservoirs required). A similar crossbar architecture has been designed for donor nuclear spins in silicon [51], similar to Kane’s architecture, which is based on a two layer crossed gate structure with a qubit and SET layer in the centre as shown in Fig. 6.6(d). The top and bottom gates alternate between source/drain (of the SETs) and upper/lower gate for the donor qubits. Control is achieved using global ESR/NMR and individual qubits are selected using a combination of upper and lower gate to perform single and two-qubit operations as well as readout by electron loading/unloading from the nearby SET. This design also assumes high homogeneity between qubits, and while a $^{31}$P donor naturally provides identical qubits, the local environment and uncertainties in donor placement could be sources of inhomogeneity. STM-lithography could provide sufficient resolution and certainty for fabrication of such a device at the cost of being a so far slow and manual process.
Figure 6.6: Crossbar. (a) Nine QD linear array illustrating fabrication and control complexity when each gate voltage would be delivered in a one-gate-one-line approach. (b) QD based crossbar architecture where a single gate voltage is shared between multiple qubits. (c) Three dimensional illustration of the architecture in (b). (d) Crossbar architecture based on donors. Reproduced from [37, 51, 52].
Hybrid architectures

When donor spins and electron spins are combined in a single hybrid architecture, certain challenges due to inhomogeneities and qubit density compared to donor-only architectures can be overcome.

In an architecture using electron spins in QDs as data qubits and donor nuclear spins as measurement qubits [54], as shown in Fig. 6.7(a), requirements on donor placement precision are relaxed and inter-donor distance up to 1 µm could be realised. Shared CCD gates allow shuttling of data electron spins to the nuclear measurement spins to implement parallelised parity measurements by performing CNOTs on the donor/dot systems based on ESR/NMR transitions. Further control over individual data and measurement qubits is gained via local top- and back-gates and manipulation is performed using global ESR. Bismuth donors spins are envisioned in this architecture due to a strong hyperfine interaction. Readout based on SETs formed using electrodes close to the local back-gates is assumed (not shown).

By encoding a single qubit in an effective two qubit system of donor nuclear and electron spin and sharing the electron between the donor and QD that is formed above as shown in Fig. 6.7(b), the requirements on placement precision and spacing between donors can be relaxed [261, 262]. This so called flip-flop qubit is encoded in the $|\uparrow\downarrow\rangle$ and $|\downarrow\uparrow\rangle$ states when the electron is brought close to the donor (near the ionization point). In the idle state, the quantum information is stored in the nuclear spin and can be transferred onto the flip-flop states using ESR. Transitions between the flip-flop states (single qubit gates) are induced using resonant microwave electric fields (second-order process). The electric dipole of the nearly ionized donor affects nearby qubits that are not idle and realises long-range (up to hundreds of nanometres) interactions between flip-flop qubits. This electric dipole-dipole coupling could
be further mediated using floating gates or resonators as illustrated using turquoise structures in Fig. 6.7(b).

A conceptually similar but in the implementation very different architecture based on dipole-dipole interactions between two different species of donors can be constructed, where the data and measurement donors are arranged in identical grids on separate chips which are brought very close to each other and a mechanical circular movement realises parity measurements based on the dipole-dipole interaction [263].

**Long-range approaches** In addition to short and intermediate range coupling as discussed above, there are two approaches that can provide long-range coupling between spin qubits that are compatible with both nanowire and planar designs. These approaches involve either shuttling of electrons over large distances, or spin-photon coupling and are typically suggested as long-range interconnects between dense qubit modules rather than providing coupling between individual qubits [50].

Spin-photon coupling follows the approach of circuit-quantum-electrodynamics (CQED), which has been successfully implemented in the field of superconducting quantum devices. The basis for such spin-photon coupling is a strong interaction between the spin and photon in a microwave cavity as shown in Fig. 6.8(a). Achieving sufficient coupling requires very careful tuning and circuit design but a spin-photon interface has been demonstrated [265]. Additionally, strong photon coupling to the charge degree [266, 267] of freedom along with charge state readout [264, 268] of a QD has been demonstrated.

Electron shuttling allows to bring distant electron spins close together on-demand. Shuttling can be performed by a series of gates and appropriately timed pulses as shown in Fig. 6.8(b) or using a surface acoustic wave (SAW). Electron transfer using
SAWs has been demonstrated in GaAs [269]. An implementation for silicon spin qubits would require a piezo-electric material deposited on the surface. Isotopically purified silicon provides an excellent base-material for shuttling due to reduced electron spin decoherence via interaction with nuclear spins and shuttling via metal gates has been demonstrated [254, 270] and decoherence during the shuttling process is studied [271, 272].

Overall, a spin-photon interface and shuttling allows for long-range coupling but could impact the speed and fidelity of an architecture [273] and requires further study.

### 6.2.3 Surface code architecture based on multi-electron mediators

Multi-electron QDs, of a size larger than typical few electron QDs, can be operated as mediators for the exchange interaction between two spins. Such mediated exchange has been proposed between donor electron spins [274] and singlet-triplet qubits in DQDs [275] and mediated exchange has been explored in GaAs based devices [276–279]. In this section an electron spin based surface code architecture building on mediated exchange interaction is presented. The mediator QD reduces the requirements on the inter-QD spacing (up to 300 nm) which facilitates fabrication and fan-out of gate structures and provides space for local reservoirs which are not considered in many of the previously presented architectures. Finally, such an architecture could accommodate nano-magnets for fast electrical control.

#### 6.2.3.1 Multi-electron QD as mediator

Two singly occupied QDs in silicon with a diameter of 50 nm are considered. In such QDs a large charging energy of 10 THz and orbital and valley excited states on the order of 0.1–1 THz are expected [80, 91, 280] as shown in Fig. 6.9(a) (excited states are indicated by dashed lines). In between the two small QDs, a multi-electron QD with a size up to 50 nm \( \times \) 300 nm and even electron occupation is considered. A level spacing on the order of \( \Delta_M = 10 \text{ GHz} \) is expected as illustrated by the levels 1 and 2 in Fig. 6.9(a). Such level spacing is still sufficiently large compared to typically observed electron temperatures of 100 mK. Due to an even occupation number, illustrated by a fully occupied first level, the mediator has overall spin 0. In this configuration, the multi-electron QD allows for mediated exchange between
the left and right spin with an approximate interaction strength of $[274]$

$$J = -2 \left( \frac{t_{R2}^* t_{R1} t_{L1}^* t_{L2}}{\Delta_R \Delta_M \Delta_L} + \text{c.c.} \right)$$  \hspace{1cm} (6.1)$$

where $t_{xy}$ are the tunnelling rates from level $x$ to level $y$ and $\Delta_{R/M/L}$ is the energy difference between level 2 and R,1 and L respectively, as shown in Fig. 6.9(a). For tunnelling rates that are fixed and on the order of 1 GHz, the exchange interaction strength can be tuned by bringing levels L and R close to level 2 (i.e. $\Delta_{L/R} = 10$ GHz) resulting in a mediated exchange on the order of 1 MHz. The interaction is reduced by multiple orders of magnitude when levels are strongly detuned (e.g. five order of magnitude for $\Delta_{L/R} = 3$ THz).

### 6.2.3.2 Surface code architecture

Using such multi-electron QD mediators a surface code architecture can be implemented as shown in Fig. 6.9(b). In this architecture, the interaction between data qubits and measurement (ancilla) qubits is mediated through multi-electron QDs.
6.2 Error correction architectures

Data qubits are formed using a single QD while ancilla qubits are formed using a DQD for Pauli spin blockade readout that is detected using a gate-based dispersive sensor. The choice of DQD ancilla qubits with four mediators each allows for parallelisation where one ancilla interacts with two data qubits at the same time. To perform stabilizer measurements, the ancilla qubits are prepared in a singlet-state, which can be done by bringing the DQD in the (0,2) charge configuration. Parity measurements are performed using mediated exchange and errors occurring on the data qubits result in transformation of the DQD ancilla into the triplet state, for both $X$ and $Z$ stabilizers without any change of basis (no Hadamards required compared to Fig. 6.1(c)) [133]. Finally, global ESR does not affect the ancilla that is encoded in the singlet-triplet space and RF gate-based readout allows for fast and compact high-fidelity readout within microseconds (see Chapter 4).

The increased spacing between data and measurement qubits, enabled by the mediator, allows for local reservoirs, where one reservoir serves four mediators. These reservoirs facilitate initial tuning of a large-scale device, as reservoir electrons do not need to be shuttled for a long distance to initialise each data and ancilla dot with a single electron. Moreover, the reservoirs make the architecture robust against leakage errors. Leakage occurs when an electron ends up in a wrong dot, which is most likely to happen during exchange, where a qubit dot electron (data or ancilla) might end up occupying the mediator resulting in an effective charge configuration of (0,3,1) or (1,3,0) of the data, mediator and ancilla system. Double occupation of the qubit dots and leakages unrelated to mediated exchange are strongly suppressed by the large Coulomb repulsion in the dots. When brought back into the non-interacting state, any excess electron on the mediator relaxes back into the data dot on a charge relaxation timescale of $T_1$ which typically occurs within 10 ns given a tunnel coupling of 1 GHz [281] and thus the leakage error fixes itself. However, there is a chance that this relaxation process leaves the mediator in a wrong charge state which can be corrected locally using the reservoirs. This charge reset depends on the tunnel rate between the mediator and reservoir which could be tuned using a metal gate. A large tunnel rate during the idle mode of the mediator allows for reset of any charge errors, while tunnelling could be suppressed during mediated exchange to avoid unwanted coupling of the qubits to the reservoir.

The error threshold of a surface code architecture can be simulated classically by transforming all experimental noise into Pauli errors [282–286]. The mediated two-qubit exchange can be of $\sqrt{\text{SWAP}}$ or CPHASE gate character depending on a local Zeeman difference as discussed in Section 2.2.2. A CZ needs to be implemented to perform stabilizer checks using the DQD ancilla, which consists either of three...
single qubit and two two-qubit gates (\(\sqrt{\text{SWAP}}\)-based), or of two single qubit gates and a single 90° controlled phase gate (CPHASE-based). Using an experimentally motivated relation between single \(p_1\) and two \(p_2\) qubit error rates of \(p_1 = 0.1 p_2\), the surface code threshold on the mediated two qubit operation can be obtained, which is 0.86\% (CPHASE) or 0.77\% (\(\sqrt{\text{SWAP}}\)) [287]. Additionally, when fixing \(p_2 = 0.5\%\) a threshold on the leakage error \(p_{\text{leak}}\) can be obtained which is 0.28\% (CPHASE) or 0.23\% (\(\sqrt{\text{SWAP}}\)) and up to 0.66\% when assuming no gate errors (\(p_2 = 0\)) demonstrating the resilience to leakage errors by transformation into computational errors.

### 6.2.3.3 Device designs

Example designs for initial characterisation and implementation of such multi-electron QD mediators based on planar Si-MOS, as shown in Fig. 2.6(b) and discussed in Appendix A, have been developed. Figure 6.10(a) shows an initial design similar to current two-qubit device structures [104] with metal gates coloured by function. It consists of extended confinement gates providing confinement along one dimension with reservoirs at each end. Close to each reservoir there are multiple gates to form QDs which can serve as data or measurement qubits, and in between these qubit structures there are gates that allow formation of a mediator of variable size. In such device, mediated exchange for a variable size mediator could be studied. Readout could be performed using two SET structures or by connecting some of the gates to RF dispersive sensors. It is also possible to perform transport measurements through the qubit mediator structure from one reservoir to the other. In a next step, after having identified a suitable mediator size, the concept of connecting a single reservoir to the mediator only could be studied in a design as shown in Fig. 6.10(b). Such a design requires an additional gate layer (metallisation step) within fabrication as indicated by the colouring. With the first layer a metal gate for tuning of the tunnel coupling between the mediator and a reservoir is fabricated. In such a design, loading of electrons from the reservoir via the mediator and leakage error reset could be studied.

### 6.3 Integration

In this section, some requirements on integrated electronics for the implementation of a large-scale quantum computer are introduced and solutions to efficiently deliv-
6.3 Integration

**Figure 6.10: Mediator QD devices.** (a) Design of a variable size mediator dot (orange) that couples to two multi-QD structures (green). Reservoirs are indicated in blue and readout can be performed using SETs (red) or gate-based sensors. (b) Design of a fixed size mediator with two multi-QD structures and a single reservoir connected to the mediator. Colours indicate gates fabricated in the same step (layer).

...ering required control and measurement signals using integrated classical circuits are discussed. In particular, a scalable readout architecture is presented.

### 6.3.1 Electronics in a full-stack processor

**Full-stack** As introduced in Section 3.2, single qubit operation requires generation of stable voltages, arbitrary waveforms and microwaves. Additionally, readout is based on fast digitization of a small signal originating from the qubit that is amplified at multiple stages. For large-scale architectures the one-qubit-one-line approach, where each signal is delivered to the device at cryogenic temperature via a separate line using room-temperature electronics, is not viable [34]. Moreover, low-latency signal processing and fast-feedback is vital to implement error-correction schemes. Furthermore, to operate any quantum computer, a compiler and...
processing architecture is required that translates and performs an abstract quantum algorithm onto the qubit architecture. An illustration of a quantum computer from a high-level algorithm to low-level signal delivery is shown in Fig. 6.11. Many of these systems could operate at cryogenic temperature, partly alongside the qubits at a temperature below 200 mK, or at higher temperature stages. Integration of classical electronics alongside qubits (at the lowest level) is fundamentally limited not only by a limited cooling power at cryogenic temperature, but also the inter-qubit spacing which depends on the qubit architecture. (De)multiplexing circuits and some electronics for signal generation and readout are amongst the most promising candidates to be directly integrated with the qubits due to a similar footprint and moderate power consumption [50, 53, 289]. Other systems are likely to be implemented at higher temperatures where a larger cooling power is available but could be integrated with the qubits when these can be operated with high fidelity at temperatures of 1–4 K.

**Figure 6.12: Dense or modular.** (a) Multi-layer chip design for direct integration of classical control circuits with a dense qubit architecture. (b) Modular approach that allows for integration of small dense qubit arrays with control circuits and electronics on the same chip by using long-range couplers. (c) Summary of expected inter-qubit spacing for different inter-qubit coupling strategies. Reproduced from [50, 53].
**Qubit architecture limitations** Based on the inter-qubit spacing that can be achieved in silicon-based qubit architectures, as summarised in Fig. 6.12(c), two approaches for integration of qubits with conventional electronics have been suggested. Firstly, a completely dense architecture based on direct qubit coupling [53], which follows an approach of integration in a multi-layer silicon chip. An illustration of such a chip consisting of a two-dimensional qubit array and addressing circuit is shown in Fig. 6.12(a). Qubits ($Q_i$) are formed in single QDs and nearest neighbour $J$-coupling is proposed. In a transistor based circuit that relies on floating-gate charge storage, signals are efficiently delivered to every $Q$ and $J$ gate word- and bit-lines. Readout using dispersive gate-based sensors is envisioned. In such a multi-layer architecture, further integration with electronics for signal generation, digitisation, control and readout could be fabricated in additional layers as long as heating of the multi-layer chip is avoided. Materials and interconnects of low thermal conductivity (e.g. superconductors) could suppress heating of the qubits due to heat produced within the control circuits. A second approach is based on a modular design which makes use of long-range coupling mechanisms that allows for interconnection of multiple small but dense arrays (such as a crossbar architecture). The space between modules could allow for integration of multiplexing, digital-to-analogue converters (DAC), analogue-to-digital converters (ADC) and vector modulation circuits [50].

**Cryogenic electronics** Cryogenic electronics plays a key role in scalable quantum computing. Detailed requirements for the low-level control and readout signals are being developed [290]. Moreover, initial designs of cryogenic circuits for quantum computing, including footprint and power consumption estimations, have emerged and first and operation of circuits at 4 K has been demonstrated. In [288] a footprint of 180 $\mu$m $\times$ 180 $\mu$m and a power consumption of 190 $\mu$W per qubit is estimated for DC gate voltage and control-waveform generation based on 65 nm CMOS technology. Such circuit could be integrated directly with a small number of qubits, however, for a large number of qubits power dissipation becomes significant and DACs, ADCs, circuits for microwave signal and vector modulation are considered to operate at higher temperatures [289]. Transistors and first circuits have been tested at cryogenic temperature demonstrating operation down to 4 K including field-programmable gate arrays (FPGA) that are capable of being reconfigured for different tasks [247, 291]. CMOS circuits to realise fast DACs and circuits for microwave generation (phase-lock loop) sufficient for qubit operation have been identified with a power consumption of $\sim$ 40 mW [292–294]. When combined with
frequency multiplexing the power consumption can be reduced to 1 mW/qubit [290]. Moreover, CMOS based low-noise amplifiers have been demonstrated with a power consumption of 91 mW [291]. Such an amplifier has been combined with a directional coupler, further second stage amplifiers, an ADC and an FPGA to realise a complete cryogenic reflectometry readout system on a single PCB [295].

### 6.3.2 Scalable readout architecture

Dispersive gate-based reflectometry is a key method for readout of large-scale architectures. In Chapter 5, a random-access circuit for gate-based readout is demonstrated. In this section, a scalable architecture for gate-based readout is presented that combines random-access and frequency multiplexing techniques. Furthermore, using a transistor model the readout signal-to-noise ratio in this architecture is estimated as a function of control circuit footprint.

#### 6.3.2.1 Architecture

The architecture shown in Fig. 6.13 is designed for readout of a two-dimensional array distributed in rows \((i)\) and columns \((j)\) and is based on conventional word-line/data-line multiplexing [53]. A specific qubit in row \(i\) and column \(j\), \(Q_{ij}\), can be addressed by a word-line \(V_{WLj}\), that controls the voltage on the gate of transistor \(T_{ij}\), and by a data-line voltage \(V_{DLi}\), that controls the gate potential on the qubit. Additionally, each row is connected to a different \(LC\) resonator for readout, which is formed by the parasitic capacitance and inductor in parallel to each data-line. Distinct
resonant frequencies $f_i$ are achieved using different values for the inductance of each resonator $L_i$. Multiple resonators (columns) can be probed simultaneously using frequency multiplexing techniques [240] such that the whole array only requires a sole high-frequency line for readout which is connected to the data-lines via a bias-tee. The array can be operated dynamically in a random-access manner. When a specific word line $j'$ is activated qubits $Q_{ij'}$ can be read simultaneously for all $i$ or their gate voltages are refreshed using the specific data-line voltages $V_{DLi}$. Each qubit may be fabricated such that it is in close proximity to other qubits to perform two-qubit operations. This possibility is indicated by the circular connection at the source of each quantum device. The array can be extended in number of rows and columns limited by the spectral overlap of the data-line resonators and typical readout and retention time.

6.3.2.2 Transistor model

In Chapter 5, SOI nanowire transistors with a channel width $W = 10 \mu m$ were characterised and identified to be well suited as control FETs for the implementation of a random-access circuit for RF gate-based readout. When integrating these control FETs in a dense scheme as shown in Fig. 6.12(a) the control circuit per qubit should not exceed an area defined by the inter-qubit spacing. To simulate the impact on readout sensitivity and retention time when scaling the control FET of $W = 10 \mu m$ to a transistor of smaller dimensions, multiple nanowire FETs of different width and gate length have been characterised at 4 K.

A typical cryogenic measurement of the resistance and capacitance of a nanowire transistor of dimensions $W = 10 \mu m$ and $L_g = 40 \text{nm}$ is shown in Fig. 6.14(a) as a function of word-line voltage, indicating a threshold of $V_{WL} \approx 0.5 \text{V}$. Identical measurements of transistors of different width and gate length were performed to obtain a model for the transistor resistance and to confirm a geometric capacitance model:

$$R_{\text{FET}} = a \cdot \frac{L_g}{W} + b \cdot \frac{1}{W}$$

$$C_{\text{FET}} = \frac{\varepsilon_r \varepsilon_0 (2H + W)L_g}{1.87 \cdot d_{\text{ox}}}.$$  (6.2)

Resistance measurements and fits of the model are shown in Fig. 6.14(b) with model parameters $a = 2503 \pm 80 \Omega$, $b = 603 \pm 26 \text{k}\Omega \text{nm}$. The capacitance model is confirmed at the example of a $W = 10 \mu m$ wide transistor where $H = 10 \text{nm}$ is the nanowire thickness and the factor of 1.87 accounts for the dark space where the barycentre of
the 2DEG is formed about 1.1 nm beneath the oxide-silicon interface. This cryogenic model of SOI nanowire transistors is useful, not only for estimating the impact of transistor scaling on gate-based readout in a random-access architecture, but could also be adapted to integrated circuit design software allowing for circuit simulation and optimisation of 4 K electronics.

### 6.3.2.3 Circuit simulations

An integrated design of a random-access readout architecture requires careful analysis of the relevant circuit parameters and their effect on chip footprint and readout SNR. To simulate performance of the architecture for a given footprint, the equivalent circuit of a single cell based on the discharge model in Fig. 5.7(a) is expanded into the RF domain. The model is shown in Fig. 6.15(a) and consists of a readout resonator (inductance $L$, capacitance $C_p$ and resistive losses $R_p$), the control FET (channel resistance $R_{FET}$ and gate capacitance $C_{FET}$, split equally between source and drain), a charge storage capacitor $C_S$ and the quantum device with state dependent gate capacitance $C_G$ and resistance $R_G$, respectively. Simulations presented in this section are performed by numerically calculating the reflection coefficient $\Gamma$ and the absolute change $|\Delta \Gamma|$ for a given capacitive change $\Delta C_G$ for the circuit.
Figure 6.15: Memory cell readout simulations. (a) Equivalent circuit for readout of a single memory cell in a random-access architecture, consisting of readout resonator, control FET, storage capacitance and quantum device. (b) Magnitude of the reflection coefficient obtained using circuit simulation with the FET off (blue), FET on (orange) and additional tunnelling in the quantum device when operating at a IDT (green).

shown in Fig. 6.15(a). From $|\Delta \Gamma|$ the SNR is obtained which additionally depends on the input power $P_{RF}$ and noise power $P_N$.

$$|\Delta \Gamma| = \left| \frac{\partial \Gamma}{\partial C_G} \Delta C_G \right|$$  \hspace{1cm} (6.4)

$$\text{SNR} = |\Delta \Gamma|^2 \frac{P_{RF}}{P_N}$$  \hspace{1cm} (6.5)

For the simulations, improved resonator parameters are selected compared to the experimental implementation presented in Chapter 5, involving NbN inductors with higher quality factors of $Q_{\text{load}} = 400$–900, as presented in Chapter 4 and Ref. [236]. The selected resonator parameters for the simulations are $C_c = 90 \text{ fF}$, $C_p = 480 \text{ fF}$, $R_p = 800 \text{ k}\Omega$ and $L = 405 \text{ nH}$ which yields a resonator operating at $\approx 300 \text{ MHz}$ with a quality factor of 400. The transistor model, see Eq. (6.3), provides values for the FET resistance and capacitance based on a selected FET width $W$ and gate length $L_g$. A storage capacitance of $C_S = 90 \text{ fF}$ and a transistor with $W = 10 \text{ \mu m}$ and $L_g = 30 \text{ nm}$ are initially considered, representing the experimental setup in
Chapter 5. For the singlet-triplet state dependent device capacitance $C_G$, a signal originating from an inter-dot transition (IDT) with a tunnel coupling $\Delta_e = 20 \, \mu\text{eV}$ and lever arm $\alpha = 0.5$ is considered, where $C_G$ changes by $\Delta C_G = 1 \, \text{fF}$ from the geometrical value of $\approx 10 \, \text{aF}$ when tunnelling is allowed \[70\]. Based on the calculations of $R_G$ in Section 5.3.2, $R_G$ is assumed to be much greater than the RF impedance of the gate capacitance and is treated as infinite. Using these circuit values, the magnitude of the simulated reflection coefficient is shown in Fig. 6.15(b) to verify that the model reproduces the features observed in experiments. When the FET is off, a resonance with similar frequency and quality factor to the experiment in \[236\] is observed. When the FET is on, there is a shift in frequency and a change in quality factor similar to the experiment shown in Fig. 5.5(b). When the FET is on and the device is operated at an IDT, there is a small additional shift which is the state-dependent gate-based signal to be detected.

The SNR depends on multiple circuit components but those that affect the physical dimensions of the cell most significantly are $W$ and $C_S$. In the following simulations the SNR as a function of $W$ and $C_S$ is obtained. Throughout the simulations the gate length is fixed to a small value of $L_g = 20 \, \text{nm}$, which keeps the FET resistance at a minimum, see Fig. 6.14(b), for a given $W$ and additionally reduces undesired effects such as charge-injection and clock feed-through when switching the FET between the on and off state (see Section 5.3.2). For every combination of $W$ and $C_S$ a coupling capacitor $C_c$ that achieves optimal matching between the resonator and the line is selected. The value of $C_c$ for perfect matching is shown in Fig. 6.16(a) which increases both with increasing $C_S$ and $W$ from 90 fF up to 400 fF. Additionally, the maximum input power below broadening (for the selected transition of $\Delta_e = 20 \, \mu\text{eV}$ and $\alpha = 0.5$) is recalculated for each point, because any reduction in quality factor of the resonator or a change in the $R_{\text{FET}}C_S$-filter function results in a reduced RF amplitude at the QD device. The obtained $Q_{\text{load}}$ ranges from 768 at low $C_S$ and large $W$ to 35 at large $C_S$ and small $W$. The resulting optimal RF operation power is shown in Fig. 6.16(b) with values ranging from $P_{\text{RF}} = -124 \, \text{dBm}$ to $P_{\text{RF}} = -92 \, \text{dBm}$. Finally, the SNR as a function of $W$ and $C_S$ is shown in Fig. 6.16(c) for an integration time of 4 µs (much shorter than the coherence time of electron spins in $^{28}\text{Si}$, $T^*_2 \approx 100 \, \mu\text{s}$), a noise temperature of 4 K and optimised RF power and coupling capacitor at each data point. Additionally, the expected retention time $\tau$ and thermal voltage noise $V_{\text{RMS}}$ for a given storage capacitance $C_S$ is shown in Fig. 6.16(d). The SNR decreases as $W$ decreases (on state $R_{\text{FET}}$ increases), which can be compensated by decreasing $C_S$, but only at the cost of reducing the time constant, which determines the voltage retention time in
Figure 6.16: SNR simulations. (a) Value of the coupling capacitance $C_c$ to achieve perfect matching and (b) maximum power $P_{RF}$ below power broadening as a function of transistor width $W$ and storage capacitance $C_S$. (c) Signal-to-noise ratio as a function of $W$ and $C_S$. (d) Retention time $\tau$ and voltage noise $V_{RMS}$ as a function of $C_S$. 
the off state (see Section 5.3.2), and increasing the thermal noise $V_{\text{RMS}}$. Contour lines indicated SNRs of 6 and 1 and a black star represents the experimental configuration presented in Chapter 5. For a qubit pitch of 100 nm, a set of parameters ($W = 100$ nm, $C_S = 25$ fF) with SNR $> 1$ can be identified that seem feasible to be fabricated within an approximate $100 \text{nm} \times 100 \text{nm}$ footprint, commensurate with a QD pitch in a dense array (see the following section for a footprint estimation). For these parameters, thermal noise increases towards $5 \mu$V, comparable to the precision and noise of common low noise voltage sources ($1–10 \mu$V), and the $RC$ time constant decreases to $0.1s$, which is sufficient for a regular refresh of gate voltages. If longer retention time and better voltage stability (drift and noise) at the same SNR and readout bandwidth are desired (or indeed even higher SNR and readout bandwidth), then the control circuitry requires a larger transistor or capacitor footprint; unless further advances in low on state resistance transistors or compact storage capacitors are made. Alternatively, the SNR can be improved by increasing the frequency of operation, $Q_{\text{load}}$ of the resonator, or using quantum-limited amplification [56, 236, 238]. Moreover, requirements on critical dimensions could be relaxed depending on the architectural implementation as shown in Fig. 6.17(c).

### 6.3.2.4 Footprint estimations

In balancing these various requirements to optimise for $C_S$, it is important to consider the capacitor footprint. In DRAM, a storage capacitance of $C_S = 10–25$ fF is required to achieve a refresh time in the range of milliseconds [241]. DRAM cells have been continuously scaled down while maintaining a total footprint of $6F^2$ by using trench or stacked capacitors with exotic high-$k$ dielectrics and large capacitor aspect ratios, where $F$ is the minimum feature size currently reaching sub 10-nm [296].

Figure 6.17 shows a multi-layer chip considered for the integration estimations, as proposed in [53], with one layer containing quantum devices $Q_{ij}$, a second layer for the control circuit including FETs $T_{ij}$ and further layers acting as insulating spacers and containing inter-connections between both (not shown for clarity). Fabrication of both active layers could be done monolithically where interconnects are realised using metal lines in an oxide spacer, or on separate dies and using three-dimensional interconnect and packaging strategies which are part of recent developments for improving performance of conventional processors [297] and are regularly used in DRAM chips. The quantum layer shows four quantum dots with a pitch of 100 nm. For clarity, the control layer shows the circuit of a single cell.
only, consisting of the drain, gate and source electrode of the control FET, a storage capacitor represented by a cuboid of large aspect ratio and multiple interconnects. Fabrication of the control circuit at the density shown in Fig. 6.17 seems feasible at the 10 nm or 14 nm technology node and using DRAM 1x nm technology considering characteristic parameters summarised in Tab. 6.1. A footprint of $\sim 12F^2$ is reserved for the capacitor (where $F$ is the minimum feature size) with an aspect ratio (AR) of up to 100. Such footprint and AR should allow for fabrication of a 25 fF capacitor when comparing with the overall cell footprint of $6F^2$ in current DRAM technology (which includes capacitor, transistor and bit-line), where a capacitance up to 12 fF is achieved. To accommodate larger storage capacitance (e.g. to achieve longer retention with reduced SNR at same transistor width) or if additional dummy transistors for mitigation of charge injection are desired a second control layer can be considered. For a quantum device pitch below 100 nm advances in low on state transistors and compact storage capacitors are required in order to maintain the same SNR and voltage stability. Additionally, for a QD pitch approaching 50 nm, further scaling towards smaller gate pitch (which is 54 nm in 10 nm technology) might be necessary.
Finally, the required footprint of on-chip spiral inductors should be considered for a fully integrated and scalable architecture. Rather than using NbN based spirals, TiN is a good candidate for achieving high quality inductors at small footprint in a CMOS process which is already found in recent gate stacks (see Section 3.1). When reducing dimensions, kinetic inductance can dominate over geometric inductance in such superconducting films. In 30-nm-thick TiN films, a kinetic inductance of 22.2 pH/sq is observed [299]. Based on this, 133 pH/sq could be achieved in a 5-nm film. A 5-nm-thick, 50-nm-wide and 19µm-long wire then yields an inductance of 50 nH, compatible with gate-based RF reflectometry readout at ~ 1 GHz (assuming $C_p = 0.5 \text{ pF}$). While such an inductor still requires a substantial footprint, a large-scale architecture that employs random-access as shown in Fig. 6.13, significantly reduces the number of inductors required for readout limited by the refresh time of the architecture. If a resonator is shared amongst 100 devices using random-access, a footprint of 10µm × 10µm is available for each resonator given a QD pitch of 100 nm and assuming that every QD connects to a resonator for readout, which might not be the case in a surface code architecture where only the measurement (ancilla) qubits are readout. The resonators could be fabricated separately from the qubits in an additional higher level layer within a multi-layer chip.
6.4 Conclusion

Different large-scale fault-tolerant architectures for spin based quantum computing have been presented. A new architecture has been introduced and analysed that uses multi-electron QDs for mediated exchange between qubits. An inter-qubit spacing up to 300 nm is expected providing space for fan-out of gate structures and local reservoirs. These reservoirs make the architecture resilient to leakage errors and a surface code threshold up to 0.86% has been determined.

Furthermore, concepts for integrating classical electronics with spin qubits have been introduced and a scalable RF readout architecture combining random-access and frequency multiplexing has been presented. Circuit simulations and estimations of the control circuit footprint indicate that such architecture could be directly integrated with spin based qubits at an inter-qubit spacing of 100 nm or more.
Chapter 7

Conclusion and outlook

In this thesis, a scalable approach to fast and high-fidelity readout of silicon-based qubits has been presented. In this chapter, the main achievements are summarised and suggestions for future research directions emerging from these results are identified.

7.1 Achievements

7.1.1 CMOS foundry-based quantum devices

Most silicon-based quantum devices are still fabricated in university clean-rooms using processes which are incompatible with industrial fabrication at semiconductor foundries. In this thesis, qubit design adaptations to semiconductor foundry friendly fabrication following CMOS processes at a 300 mm wafer scale have been presented using silicon-on-insulator (SOI) technology developed CEA-Leti as shown in Section 3.1. Formation of single and double quantum dot and hybrid donor-dot systems are achieved, as shown in Chapter 4 & 5, including observations of Pauli spin blockade in a donor-dot system. Moreover, high stability and reproducibility is observed and devices have reduced complexity at the cost of less tunability by engineering some parameters in fabrication. Furthermore, highly sensitive gate-based readout of QDs formed in SOI nanowire devices is demonstrated. Finally, CMOS compatible fabrication allows for integration with low-temperature electronics, as explored in Chapter 5 and 6. To conclude, quantum devices realised within a scalable CMOS compatible framework provide a viable platform towards high-fidelity and reproducible qubits. Multiple requirements for implementation of a qubit have been fulfilled and further efforts promise to be fruitful in the future.
7.1.2 High sensitivity gate-based sensing

Methods to significantly improve radio-frequency (RF) gate-based reflectometry techniques to provide fast, sensitive, non-disturbing and compact quantum state readout of silicon devices have been presented in Chapter 4. Firstly, using an optimised resonant circuit design, which includes a NbN superconducting spiral inductor, well-matched lumped-element resonators with a loaded quality factor up to $\sim 1000$ have been demonstrated, enabling improvements of the charge sensitivity, measured using a SOI nanowire device, by a factor close to 30. Secondly, by combining such a high-quality resonator with a Josephson parametric amplifier (JPA) further improvement of the charge sensitivity down to $0.25 \mu\text{e}/\sqrt{\text{Hz}}$ is achieved, outperforming established radio-frequency SET sensors. Moreover, when using the JPA the disturbance on qubits during readout is reduced as the same sensitivity is obtained at lower RF signal power. Finally, measurements of the signal-to-noise ratio of a donor-dot hybrid singlet-triplet system with and without the JPA indicate that single-shot spin readout at a fidelity of 99.7% within 1 µs and 32 µs respectively could be achieved based on the presented sensor improvements. To conclude, the results demonstrate that fast and compact spin readout of high-fidelity can be achieved using a resonant circuit and JPA that is fabricated separately from the quantum devices allowing for high-density qubit architectures.

7.1.3 Scalable quantum computing architectures

When scaling silicon quantum devices to the large number of qubits required to solve the most challenging computational problems in a fault-tolerant quantum computing architecture, challenges arise not only within fabrication of the quantum devices themselves but also within the classical control and readout system. In Chapter 5 sequential readout of two quantum devices, one after another, using an integrated transistor circuit has been demonstrated, providing the basis for a reduction of readout circuit complexity in an integrated quantum-classical processor. Gate-based readout via control transistors at a sensitivity comparable to direct readout is demonstrated, when the transistor in the on state. Moreover, floating gate single-electron charge storage on the order of 1 s is achieved when the control transistor is in the off state, allowing readout of up to 1 million devices before gate voltages need to be refreshed. Based on an experimentally obtained control transistor model, it is demonstrated that the required control circuit per qubit could be realised within a 100 nm x 100 nm footprint allowing for direct integration of the control circuit with a dense two-dimensional qubit architecture in
a multi-layer chip, while maintaining a SNR > 1 at 4µs integration time combined with sufficient charge storage. Moreover, a readout architecture combining random-access and frequency multiplexing has been introduced for efficient readout of a two-dimensional array of a large number of qubits. Finally, a fault-tolerant architecture based on mediated inter-qubit exchange via elongated quantum dots has been presented. Such architecture provides an increased footprint available for control electronics per qubit of up to $300\,\text{nm} \times 300\,\text{nm}$ and addresses challenges related to leakage errors by integrating local reservoirs. To conclude, these results provide first steps towards implementing a fully-integrated fault-tolerant quantum computing architecture.

7.2 Future research directions

7.2.1 CMOS devices

Based on results presented in this thesis, improvements on device designs and fabrication techniques can be identified.

In SOI nanowire based devices highly stable few electron QDs have been demonstrated. However, due to a single gate design tunability was limited. Using a top-gate that is split in the middle, as introduced in Section 6.2.2, individual control over each corner QD could be achieved. Additional gates between the split-gate and the source-drain reservoirs and an additional top-gate above the split-gate could allow for individual tuning of reservoir and inter-dot tunnelling rates. The increased tunability should allow formation of a number of different qubit implementations ranging from single spins, to singlet-triplet qubits and multi-spin qubits as shown in Fig. 7.1. Coherent control and single-shot readout could be demonstrated in such devices by combining the improved tunability with the high-fidelity readout demonstrated in this thesis. Moreover, multi-qubit structures could be fabricated within the same nanowire or in separate nanowires coupled via a floating gate, as introduced in Section 6.2.2, allowing to study exchange-based or capacitive two qubit gate implementations in a nanowire architecture.
Finally, qubit performance could be strongly improved when adapting techniques for the realisation of $^{28}$Si rich layers, such as growth by means of chemical vapour deposition, to industrial scale wafers and large-scale fabrication processes.

### 7.2.2 Gate-based readout

The implementation of high-sensitivity gate-based readout presented in this thesis has been limited by the gain of the JPA. Gain improvements could be achieved by increasing the isolation between the readout resonator and the JPA. Furthermore, noise added by the JPA could be avoided (reduced below the quantum limit) by operating in phase-sensitive mode. Using these improvements, the mechanisms ultimately limiting sensitivity of a gate-based dispersive sensor, such as Sisyphus noise or Johnson noise, could be explored further. Additionally, the measurement speed in this implementation was limited by the bandwidth of the high-$Q$ readout resonator. Moving to a higher frequency of the resonant circuit (i.e. between 1 and 2 GHz) while maintaining a high loaded $Q$ should allow for sub-microsecond gate-based spin readout. Furthermore, identifying an appropriate coupling capacitance for achieving a well-matched resonance is an essential part in optimising sensitivity. The capacitance required for matching is smaller than with what is achieved with commercial SMD capacitors. Multiple appropriate coupling capacitors could be fabricated on-chip along with the inductor. Combined with inductors of different inductance, frequency multiplexed readout using the same line could be achieved. Alternatively, instead of capacitively coupling the RF signal into the resonant circuit, a transmission line could be fabricated on the inductor chip and inductive coupling to multiple inductors could be achieved using the same line with the potential of increasing sensitivity further due to a reduced resonant circuit capacitance. Moreover, using a wide-band travelling wave parametric amplifier a frequency multiplexed approach could operate at the quantum limit of noise. Finally, the resonator footprint could be reduced and fabricated using a CMOS compatible metal by shrinking the characteristic dimensions and exploring TiN as a base material for inductor fabrication as discussed in Section 6.3.2.4.

### 7.2.3 Integration

The transistor model and guidelines developed in Section 6.3.2 could be made compatible with integrated-circuit design tools for fabrication of optimised and integrated random-access circuits for efficient reflectometry readout of many quantum devices. However, further work is necessary to integrate additional electronics for
7.2 Future research directions

signal generation, detection and processing. While conventional electronic circuits can be operated at cryogenic temperature, gaining further understanding of the behaviour of individual circuit elements at cryogenic temperature is still necessary, especially with respect to power consumption of more complex circuits. The limited cooling budget split over different temperature stages when operating in a dilution refrigerator is the main constraint and challenge for transitioning from general-purpose room-temperature electronics towards special-purpose integrated cryogenic electronics.
Appendices
Appendix A

Study of RF-SETs in Si-MOS technology

In this appendix, measurements on first-generation planar Si-MOS devices fabricated using CMOS compatible processes are presented. A radio-frequency single-electron transistor is formed and estimations on yield and fabrication problems are reported. These results exemplify non-scalable spin qubit readout based on an external charge sensor.

Measurements have been performed under supervision of Prof F. Kuemmeth at the Center for Quantum Devices and Station Q at the Niels Bohr Institute at the University of Copenhagen. The author acknowledges contributions from F. Ansaloni and Dr A. Chatterjee in jointly carrying out the experimental work and imec for device fabrication.

Fabrication

In planar Si-MOS devices, QDs are formed at the silicon/oxide interface using metal gate structures at the surface (without etching of the silicon layer) similar to planar transistors. Figure A.1 shows an illustration of a Si-MOS QD device [44] including a summary of processing steps developed for university clean-room fabrication [300].

To begin, a field oxide is grown on a bare (high-resistivity) silicon wafer which is typically done by controlled oxidation in a furnace (SiO$_2$). Using micro-fabrication patterning steps, involving spinning of a photo-resist, writing of a pattern using optical lithography followed by etching, regions for ion-implantation are defined. Ion implantation is the basis to form electron reservoirs and ohmic contacts. Via such reservoirs, transport measurements can be performed and single electrons can be loaded onto a QD device. After ion-implantation, the contaminated oxide is removed followed by micro-patterning of the active device area, where the field oxide is removed and a thin oxide (typically 5–10 nm SiO$_2$) of ultra high-purity is deposited. A final micro-fabrication patterning, metal deposition and lift-off
Figure A.1: Planar Si-MOS device fabrication. Shows simplified microfabrication steps to form electron reservoirs and a high-quality thin oxide. On the thin-oxide multi-layer gate metal structures are fabricated using nano-fabrication to form SETs and QDs. Adapted from [44, 300].
step provides metal contacts for bonding after which the chip is ready for nanofabrication of QD gate structures.

Development of a technique for fabrication of thin and dense gate structures has been a key milestone for reliable QD formation in planar Si-MOS devices [301]. To begin, a PMMA resist is spun and a first set of gate structures is patterned using electron-beam patterning (typical separation and width: < 50 nm) followed by deposition of a 30-nm-thick aluminium (Al) film. In a lift-off process the resist and any Al on top is removed while the Al gate pattern remains. The set of Al gates is oxidised on a hotplate, where a thin AlO$_x$ oxide layer is formed, providing electrical isolation and allowing for fabrication of subsequent and overlapping gates. Additional structures are fabricated in the same process but typically the Al layer thickness is increased by $\approx 20$ nm for each additional layer to avoid breaking of gate structures at points of overlap. Such multi-layer gate fabrication allows formation of SET and dense one-dimensional QD arrays. An SET consists of an overlapping top-gate for accumulation and barriers for depletion, while a QD structure consists of multiple slightly overlapping barrier and accumulation gates. Exemplary devices fabricated using this technique can be found in Refs. [64, 91, 300, 301].

Many of these processes are incompatible with industrial scale CMOS fabrication technology, such as electron-beam patterning, deposition of aluminium films and metal lift-off. In this appendix measurements on a first-generation of ‘foundry friendly’ planar 300 mm wafer scale Si-MOS QD devices fabricated at imec are presented. Devices are based on established designs [91] and engineers at imec have adapted processes presented in Fig. A.1 to be compatible within a large scale production line. For this, gate structures are fabricated using CMOS compatible TiN and dry-etch patterning combined with high-quality gate oxide and high-k dielectrics. To allow flexibility in the QD structure designs and to reduce processing time masks are still electron-beam defined. A scanning electron microscope (SEM) micrograph of such a foundry fabricated device is shown in Fig. A.2. The gate
structures are buried underneath ~ 200 nm of oxide while the ESR line is fabricated alongside with bonding pads on the surface.

Finally, QDs can also be formed in a similar way using a Si/SiGe heterostructure rather than a bare Si wafer, where electron accumulation occurs within a Si well buried ~ 40 nm underneath the surface [37, 108, 302]. A global top-gate is used for accumulation combined with gate structures for confinement and depletion.

**Device and setup**

Figure A.3(a) shows a chip containing multiple devices (two are bonded for characterisation) on a ‘Copenhagen’ board. A close-up view of the bonding pads of a single device is shown in Fig. A.3(b). Bonding pads follows a standard imec layout of 60 µm × 80 µm and pads that connect to implanted regions, acting as electron reservoirs, are highlighted (with suffix “imp”). At the top of each device, a coplanar wave guide, that is continuously reduced in size and undergoes a transition into a coplanar strip, is fabricated to deliver high-frequency signals for spin manipulation.
based on electron spin resonance (ESR). See the next section for details on wave
guide design and simulations. These metal structures sit on top of a ∼ 200 nm-thick
layer of oxide. Connections to the device that sits underneath the oxide at the sili-
con wafer surface are made using through-oxide-vias as shown in Fig. A.3(c), where
metal structures on the silicon surface can be identified. Different colours within
the gate structures are observed due to a three step (layer) metal deposition process.
Figure A.3(d) shows the device design, where structures are coloured by function,
with the ESR line at the top (gray), single-electron transistor (SET) device on the
left (red), multi-QD structure in the centre (black and green) and electron reservoir
(blue) on the right. The SET is formed using a top-gate that induces a channel
between the top and bottom reservoir (STtop/STbot) and using barriers (B1/B2)
that constrict the channel to form a tunnel coupled island. QDs are formed using
the confinement gate (CG, typically biased at 0 V) and accumulation gates G1 to
G4, and electrons are loaded using the reservoir gate GR, that extends an implanted
region ending close to G4. Using the device highlighted in Fig. A.3(a) formation
of a radio-frequency (RF) SET is demonstrated in Section A, where a resonator is
connected to the top-reservoir of the SET as shown in Fig. A.3(d). Additionally,
fast lines to deliver pulses at a frequency up to 1 GHz are connected to G1 and G3.

**Transmission line modelling**

The microwave transmission line should deliver a large magnetic field at the de-
vice while minimising the electric field which could otherwise affect highly sensitive
charge sensing techniques. Designs and simulations follow the approach presented
by Dehollain *et al.* [303] of transmission line modelling using CST Microwave stu-
dio, a three-dimensional microwave engineering software. Two different designs are
considered. To begin, as shown in Fig. A.4, a coplanar waveguide (CPW) struc-
ture that is continuously narrowed down and shorted at the end. This structure
was designed by Dr A. Chatterjee for post-processing of foundry-fabricated silicon
nanowire devices. CPW structures have proven to be able to deliver sufficient mag-
etic fields for spin control [304] and offer operation at a wide range of frequencies
as no further conversion in the transmission line is performed. A CPW structure
was chosen due to space constraints and a very dense pad layout at the surface of
such foundry fabricated devices. Secondly, a structure that undergoes a CPW to
coplanar strip-line (CPS) as shown in Fig. A.4(b) is considered. A shorted CPS has
the advantage that all the current travels through a single thin wire (rather than
two), providing an increased magnetic field at the spin location. A CPW-to-CPS
Figure A.4: Transmission line. (a) Design of a shorted coplanar waveguide (CPW). (b) Design of a CPW with coplanar strip-line (CPS) transition. (c) Modelling of the CPW-CPS structure within CST Microwave studio, including on-PCB CPW and bond-wires. (d) Illustration of the fine-mesh in the area near the shorted end of the CPS (thin wire). (e) Simulation results of reflection coefficient as a function of frequency for the CPW and CPW-CPS structure. (f) Simulation results of the expected magnetic field 200 nm or 1 µm away from the centre of the thin wire. (g) Expected magnetic field at the chip surface.
transition could lead to additional stray reflections. To avoid reflections, transmission lines should be matched to 50 Ω for a given substrate thickness and material by selecting appropriate conductor and gap width which can be obtained from a transmission line calculator such as TX-Line. By following CPW-to-CPS design guidelines [305, 306], additional reflections at the transition can be minimised.

Figure A.4(c,d) shows the modelling of the CPW-to-CPS design in CST microwave studio. To obtain a realistic estimate of the performance, the model includes an on-PCB CPW and connections made from the PCB to the on-chip structure via bond-wires. The CPW-to-CPS structure is modelled to sit on top of a silicon substrate of 700 µm and the PCB dielectric is Rogers 4003. Bond-wires within the CPW-to-CPS structure, from one ground plane to the other, are part of CPW-to-CPS transition guidelines. Fine meshing is important to obtain accurate results at good spacial resolution as CST uses a finite element solver. To minimise simulation time, fine meshing is defined locally at points of interest. Figure A.4(d) shows the meshing close to the CPS short. Simulations of the reflection, $S_{11}$, (see Fig. A.4(e)) show that both the CPW and CPW-CPS structure are well matched over a large range of frequencies. At $f > 20$ GHz the CPW-CPS performs slightly worse and shows a resonance feature compared to the CPW. Nevertheless, due to the increased current, the CPW-CPS structure outperforms the CPW in terms of the magnetic field delivered at the device over all frequencies as shown in Fig A.4(f), where the field at two different locations, 200 nm and 1 µm away from the centre of the thin wire, is shown. Using the CPW-CPS structure an increased magnetic field is observed. Moreover, due to the overall increased field sufficiently large fields can be delivered to a location further away from the wire (e.g. nearly identical field at 1 µm and 200 nm distance when comparing between CPW and CPW-CPS). Figure A.4(g) shows the magnetic field at the chip surface at 20 GHz. All estimations of the magnetic field are based on an input power of 0 dBm.

To conclude, a design with CPW-to-CPS transition delivers stronger fields at spin locations nearby compared to a simple CPW. This is particularly important for foundry friendly fabricated designs presented in appendix, where the transmission line is fabricated on top of a 200 nm oxide layer which results in an increased distance to the spins, such that A CPW-to-CPS design was selected by imec for fabrication.
SET transport characteristics

In this section, transport characteristics and statistics of multiple SETs at room-temperature (RT), 4 K and milliKelvin temperature are shown.

Room-temperature and 4 K

Prior to bonding of a device, as shown in Fig. A.3(a), multiple devices were tested at RT and 4 K in a needle probe station and 4 K system to obtain typical device characteristics and statistics. For device testing, a small source-drain bias of 10 mV is applied between the SET top and bottom reservoir and transport through the SET device is measured ($I_{\text{SET}}$). Due to a limited number of probes in the probe station, unused gates remained floating during these measurement. At first, turn on of the SET channel is characterised by applying a voltage $V_{\text{TG}}$ to the SET top-gate. Following this, the channel is depleted using either of the barrier gates, by applying a voltage $V_{\text{B1/B2}}$, while keeping $V_{\text{TG}}$ at a high level where transport through the channel has been observed.

A typical measurement of a device at room-temperature is shown in Fig. A.5(a), from which a turn-on voltage $V_{\text{TG}}^{\text{on}}$ and depletion voltages $V_{\text{B1/B2}}^{\text{off}}$ can be obtained. This example measurement shows symmetrical behaviour for both barriers, which deplete the channel at a voltage just above 0 V while a turn-on is observed around 2 V. At 4 K, as shown in Fig. A.5(b), the turn-on threshold generally shifts to larger voltage. Similarly, both barriers deplete the channel at larger values of $V_{\text{B1/B2}}$. In this example both barriers are not perfectly symmetrical but oscillations are observed in both measurements, possibly due to striction into the regime of a one-dimensional conductance channel or unintentional QDs or traps in the barrier region. Figure A.5(c-f) show statistics of $V_{\text{TG}}^{\text{on}}$, $V_{\text{B1/B2}}^{\text{off}}$ and the channel resistance $R_{\text{SET}}$ (at $V_{\text{TG}} = 3$ V) of 18 devices. For all parameters a spread around averages value of $V_{\text{TG}}^{\text{on}} = 1.8$ V, $V_{\text{B1/B2}}^{\text{off}} = 0.17(0.21)\ V$ and $R_{\text{SET}} = 700\ \text{k}\Omega$ is observed with a few outliers.

In addition to testing formation of an SET at RT and 4 K, individual gate leakage and leakage tests from one gate to another were performed. These revealed systematic leakage between both barriers (B1, B2) and gates G2 and G4. These four gates have been fabricated in the same deposition step as shown in Fig. A.6(a), which shows the device design with structures fabricated in the same deposition step appearing in the same colour. Most of the devices tested show significant leakage between all four of the gates (whole layer) indicating a systematic fabrication problem (possibly insufficient metal etching), while a small number of device shows
Figure A.5: Transport characteristic. (a) SET current $I_{SET}$ as a function of top-gate and barrier voltages showing turn-on and depletion characteristic of one SET device at room-temperature. (b) Turn-on and depletion of one SET at 4 K. (c-f) Statistics obtained from measurements of 18 devices on the turn-on $V_{on,TG}$ and depletion voltages $V_{off,TGB1/B2}$ and SET channel resistance $R_{SET}$ (at $V_{TG} = 3$ V). A bias of $V_{SD} = 10$ mV is typically applied to observe current through the SET.
leakage between three gates (no barrier B1 leakage) or even no measurable leakage between all gates at RT as shown in Fig. A.6(b).

**MilliKelvin temperature**

Two devices have been bonded and tested at milliKelvin temperature and transport characteristics are shown in Fig. A.7 (at a SET source-drain bias of 1 mV).

SET transport ($I_{SET}$) as a function of both barriers for device 1 and 2 is shown in Fig. A.7(a,b) respectively. For both devices some horizontal and vertical lines are observed due to one-dimensional transport through the barriers or unintentional QDs forming under the barriers. Additionally, there are diagonal lines indicating Coulomb blockade due to formation of an island in between both barriers. In device 1 features of Coulomb blockade are observed over a large range of barrier gate voltage, while device 2 only shows Coulomb blockade in a small region. Both devices show significant leakages between gates in metal layer 2 at milliKelvin temperature. An example of barrier B2 leakage is shown in the bottom panels of Fig. A.7(a,b) at a fixed barrier B1 voltage of $V_{B1} = 1.5$ V. Due to barrier-to-barrier leakage, both barriers are kept at the same voltage for all following measurements. For device 1, this voltage is $V_{B1,B2} = 1.7$ V and for device 2 $V_{B1,B2} = 1.8$ V.

Transport through each device as a function of source-drain and top-gate voltage, as shown in Fig. A.7(c,d), reveals Coulomb diamond characteristics. For device 1 diamonds do not appear clearly, which could indicate transport through multiple
Figure A.7: MilliKelvin characterisation. (a,b) Top panels: SET current as a function barrier gate voltage at $V_{TG} = 3$ V for device 1 and 2 respectively. Bottom panels: Gate leakage observed at barrier B2 during these measurements. (c,d) Top panels: SET current as a function of source-drain bias and top-gate showing Coulomb diamonds in device 1 and 2 respectively. Bottom panels: SET current as a function of top-gate at $V_{SD} = 1$ mV showing Coulomb blockade oscillations.
QDs. Device 2 shows diamonds more clearly and a lever-arm of $\alpha \approx 0.15$ and charging energy of $E_C \approx 3$ meV is estimated. In the bottom of Fig. A.7(c,d) Coulomb oscillations over a large range of $V_{TG}$ at $V_{SD} = 1$ mV are shown.

**Radio-frequency SET**

In this section, radio-frequency operation of device 1 is shown using a resonator that is formed by a 1200 nH surface mount inductor connected to the top-reservoir of the SET and the parasitic capacitance in parallel. For device 2 the resonator is connected to the bottom reservoir and no RF response is observed even though similar matching of the resonator to the device is obtained. The bottom reservoir is much further away from the device, which results in a more resistive channel between the resonator and the SET device. Combined with an increased parasitic capacitance a smaller $RC$ time constant is obtained and the RF signal could be strongly attenuated before it reaches the device.

For device 1, a strong response of the resonator to changes in the SET top-gate are observed as shown in Fig. A.8. At low $V_{TG}$, the resonator is badly matched due to the large resistance of the SET as shown in the magnitude of the reflection coefficient in Fig. A.8(a). Close to $V_{TG} = 2.5$ V the resonator goes through perfect matching and the resonance settles at $f_r = 153.7$ MHz when a complete SET channel is formed. From the shift of the resonance, a parasitic capacitance of 0.87 pF with the SET channel off, and 0.89 pF with the SET channel on is obtained. From the coupling coefficient $\beta = 1.77$ an SET resistance of 47 kΩ is obtained at $V_{TG} = 2.9$ V (see Eq. (2.36, 2.40)). From transport measurements shown in the previous section, Coulomb oscillations are expected at $V_{TG} > 2.9$ V, which should lead to a sudden changes of the SET resistance. In Fig. A.8(b), which shows the magnitude of the reflection coefficient as a function of $V_{TG}$ and frequency, these changes in resistance are observed as a periodic modulation of the resonance. When performing homodyne detection at a fixed frequency, these oscillations observed in the offset corrected voltage $\Delta V_{RF}$ clearly coincide with the Coulomb oscillations observed in transport, as shown in Fig. A.8(c), demonstrating radio-frequency readout of the SET.

**Charge sensing**

After characterisation of the SET in transport and at radio-frequency, charge sensing of QDs forming under gates G1 and G3 is attempted in device 1. Forming of QDs
is complicated due to leakage between gates B1, B2, G2 and G4 which are all connected to the same voltage source of 1.7 V to form the SET.

To begin, transport through the SET as a function of \( V_{G1} \) and \( V_{TG} \) is obtained to estimate the capacitive coupling of a voltage on G1 to the SET as shown in Fig. A.9(a). A lever-arm of \( \alpha_{G1} = 0.4 \) is obtained from the slope of the transitions. In a similar measurement a lever-arm of \( \alpha_{G3} = 0.32 \) is obtained for G3. Lever-arm measurements are the only method to test integrity of the QD structure and values obtained are consistent with nanometre scale metal gates. For charge sensing, the SET is biased at the maximum point of transconductance of one of the Coulomb oscillations in a regime of dense Coulomb oscillations as shown in Fig. A.9(b). The reservoir gate RG is set to 3 V to allow loading of electrons from the reservoir into the QD structure. Electrons are loaded via gate G4 and G2 which reside at 1.7 V. Consequently, a trapping potential for electrons under G1 or G3 is expected to form at \( V_{G1/G3} > 1.7 \) V. A measurement of the differential current through the SET \( \frac{dI_{SET}}{dV_{G3}} \) as a function of \( V_{G1} \) and \( V_{G3} \) is shown in Fig. A.9(c). Multiple
wide structures are observed in the background that couple both to G1 and G3 and are attributed to the SET moving through multiple Coulomb oscillations due to non-zero gate lever-arms $\alpha_{G1/G3}$ (no SET top-gate compensation is performed). Additionally, very regular features strongly coupled to $V_{G3}$ and weakly coupled to $V_{G1}$ are observed at $V_{G3} > 1.7 \, V$ which are attributed to formation of a QD under G3 and represent single electron loading events. Assuming a lever-arm of $\alpha = 0.15$ for $V_{G3}$ to the island underneath, similar to what is measured for the SET top-gate, the separation of the regular lines indicates a charging energy of $E_C \approx 5 \, \text{meV}$ compatible with what is measured in the SET device. Larger charging energy could be achieved using slightly negative voltages on the confinement gate CG. Further analysis of this device was hindered due to lack of reproducibility as a result of stability problems which are discussed further in the next section.

**Device stability**

In this initial batch of devices, formation and charge sensing of QDs was complicated due to hysteresis and instabilities. To study whether the leakage observed between gates of metal layer 2 could have a significant contribution to such instabilities, an additional device that does not show any gate leakage at RT and milliKelvin was characterised.
Even though, no measurable leakage was observed in this device, voltages applied to G2 and G4 (of metal layer 2) showed a much stronger effect on the SET stability than voltages applied to G1 and G3, which shows that even when there is no measurable leakage, fabrication problems within metal layer 2 still affect device stability. Additionally, even when QD gates are grounded the SET suffers from hysteresis as shown in Fig. A.10, which shows an exemplary sequential measurements of the SET in a device with non-measurable gate leakage using high-stability voltage sources and voltages dividers for further precision and stability.

**Conclusion**

In this appendix, Si-MOS planar devices have been designed and adapted to industrial processes at imec. Leakage between multiple metal gates related to an imperfect fabrication process has been identified. Nevertheless, formation of SETs at milliKelvin temperature has been demonstrated and rf-sensing is achieved using a resonator that connects to the top SET reservoir. Using such SET, formation of a QD in the QD device structure has been sensed. However, stability problems have prevented further study of SET sensitivity and sensing of QDs in these devices.

In a future fabrication run, additional etching of gate layer 2 could prevent gate leakages. Exploring of different thin and gate oxides could provide ways to reduce instabilities related to charge traps in the oxides. Finally, gate voltage compensation techniques (e.g. capacitance matrix based [38]) could provide a path to improved SET sensitivity.

SET readout techniques demonstrated in this appendix exemplify non-scalable readout of spin qubits in silicon. The main results of this thesis focus on improving scalable gate-based readout techniques at the example of silicon nanowire QD devices, where QDs of high stability and low hysteresis are formed. These gate-based techniques could be applied to devices presented in this appendix, or other planar Si-MOS or Si/SiGe devices, for achieving scalable readout.
Appendix B

Small signal limit of dispersive circuit QED

Throughout this thesis a small signal analysis ($Q_{\text{load}} \Delta C / C \ll 1$) is performed, that typically applies to resonators that are fabricated separately from the qubit, where moderate resonator quality factors are achieved and a significant parasitic capacitance is observed and $|\Delta \Gamma| \approx |\partial \Gamma / \partial C \Delta C|$. In this appendix a more general formalism is introduced which is derived from circuit quantum electrodynamics (circuit QED) and reproduced from work that has been performed in the context of superconducting qubits [237, 307].

Circuit QED has been implemented using on-chip superconducting transmission line resonators coupled to superconducting qubits [308–310]. Recently, these approaches have been adapted to QD based qubits in GaAs and Si based systems. Using high-impedance resonators, for enhanced coupling to single charges in the QD, strong coupling [265–267] capable of providing long range qubit interactions and dispersive readout [265, 268] has been demonstrated. In circuit QED the interaction between light and matter is studied, where a single photon in a cavity is coupled to an artificial atom (qubit). When the qubit frequency (difference between the qubit ground and excited state) is resonant with the cavity, radiative decay of the qubit into the cavity is strongly enhanced. When the coupling rate $g$ between the qubit and the cavity is larger than the cavity linewidth $\kappa_r = \omega_r / Q_{\text{load}}$ (determined by the photon lifetime) and the qubit linewidth $\gamma$ (determined by relaxation and dephasing rates), $g > \kappa_r, \gamma$, strong coupling is achieved where coherent exchange between the qubit and a single photon in the cavity is realised. In the dispersive limit of circuit QED, where the qubit is coupled off resonantly to the cavity, qubit coherence is preserved and a dispersive measurement is achieved based on a frequency shift of the resonator depending on the qubit state. The following analysis focuses on the dispersive regime where qubit readout is achieved.
A resonator and qubit parallel to a transmission line coupled through a coupling capacitor as shown in Fig. 4.2 is considered. The Hamiltonian of a qubit \((H_q)\) coupled to a linear resonator \((H_r)\) with interaction term \((H_I)\) is given by (Jaynes–Cummings model) \([237]\)

\[
H = H_r + H_q + H_I = \hbar \omega_r a^\dagger a - \frac{\hbar \omega_q}{2} \sigma_z + \hbar g (\sigma_+ a + \sigma_- a^\dagger),
\]

(B.1)

with \(\sigma_y = i(\sigma_+ - \sigma_-)\), \(\omega_r\) being the resonator frequency, \(\omega_q\) the qubit frequency and \(g\) being the coupling strength. In the dispersive limit \(|\Delta| = |\omega_q - \omega_r| \gg g\), such that a transformation and expansion around \(g/\Delta\) can be performed which simplifies the Hamiltonian to \([237]\)

\[
H = \left(\hbar \omega_r + \frac{-g^2}{\Delta} \sigma_z\right) a^\dagger a - \frac{\hbar \omega_q}{2} \sigma_z,
\]

(B.2)

where a dispersive shift of the resonator frequency dependent on the qubit state becomes apparent with an overall difference of the resonator frequency between the two different qubit states of \(2g^2/\Delta\). Additionally, in the dispersive limit the...
cavity enhanced Purcell relaxation rate \( \Gamma_{\text{Purcell}} = \kappa_r \frac{g^2}{(\kappa_r/2)^2 + \Delta^2} \) [311] becomes small, preserving qubit coherence. The qubit dependent frequency shift is detected by measuring the amplitude and phase of a probe signal \( \omega_r \) that is scattered at the resonator and sits in between the two possible frequencies \( \omega_r \pm g^2/\Delta \). To obtain the corresponding two possible values for \( \Gamma \), the scattering of the probe signal is expanded for small changes \( \delta \omega = \omega - \omega_r \) in the resonant frequency \( \delta \omega/\omega_r \ll 1 \) [237]

\[
\Gamma = \frac{\Gamma_{\text{min}} + 2iQ_{\text{load}}\delta \omega/\omega_r}{1 + 2iQ_{\text{load}}\delta \omega/\omega_r}, \tag{B.3}
\]

which describes a circle with diameter \( D = 1 - \Gamma_{\text{min}} \) in the complex plane as shown in Fig. B.1(a). Consequently, \( \Gamma(\delta \omega = \pm g^2/\Delta) \) should sit at diametrically opposite points on the circle and the separation \( |\Delta \Gamma| \) is maximised when the north and south pole of the circle are reached, which occurs when \( \delta \omega/\omega_r = \pm 1/2Q_{\text{load}} \) [237] such that maximum visibility is achieved when

\[
\frac{g^2}{\Delta} = \frac{\omega_r}{2Q_{\text{load}}} = \frac{\kappa_r}{2}. \tag{B.4}
\]

This illustrates that for a given frequency shift \( g^2/\Delta \) there is an optimal resonator bandwidth \( \kappa_r \). Moreover, the minimum of the reflection coefficient \( \Gamma_{\text{min}} \) determines the circle diameter and in such the maximum value of \( |\Delta \Gamma| \). \( \Gamma_{\text{min}} \) can take values from \([-1, 1]\) and reaches zero for perfect matching. The largest circle diameter is obtained when the resonator is strongly over-coupled and \( \Gamma \to -1 \).

Finally, from this analysis, the small signal limit given in Eq. (4.21) can be derived by expressing the change in \( \Gamma \), which is purely imaginary, in terms of capacitive changes using \( \delta \omega/\omega_r = -\Delta C/2C \) and expanding for small \( Q_{\text{load}}\Delta C/C \ll 1 \) [237]

\[
|\Delta \Gamma| = \text{Im} \Gamma = \frac{2Q_{\text{load}}\delta \omega/\omega_r(1 - \Gamma_{\text{min}})}{1 + (2Q_{\text{load}}\delta \omega/\omega_r)^2} = \frac{Q_{\text{load}}\Delta C/C(1 - \Gamma_{\text{min}})}{1 + (Q_{\text{load}}\Delta C/C)^2} \tag{B.5}
\approx (1 - \Gamma_{\text{min}})Q_{\text{load}}\Delta C/C = (1 - \Gamma_{\text{min}})Q_{\text{int}}\Delta C/C \tag{B.6}
\]

\[
= \frac{\beta}{(1 + \beta)^2}Q_{\text{int}}\Delta C/C \tag{B.7}
\]

The difference between the large and small signal limit is illustrated in Fig. B.1(b,c). In the small signal case of moderate \( Q_{\text{load}} \) and significant \( C \) (which is dominated by parasitics) the largest phase shift \( \Delta \phi \) is obtained by operating close to matching and using a large internal quality factor which determines the slope of the phase shift close to \( f_r \). The loaded quality factor could be increased up to a limit given by the
desired readout bandwidth. In contrast to this, in the large signal limit there is an optimal loaded quality factor and readout bandwidth for a given coupling strength and detuning and the signal is maximised by operating in the strongly over-coupled regime.
Appendix C

Even parity inter-donor/dot charge transition

In this appendix, further measurements of the inter-donor/dot transition (IDT) discussed in Section 4.3.6 are presented including extraction of gate coupling $\alpha$ and tunnel coupling $\Delta_c$.

Even parity of the transition is confirmed by measuring charge stability diagrams in the readout resonator phase response at increasing $B_z$ field as shown in Fig. C.1. With no external field applied the IDT is clearly visible as highlighted in the line-cut beneath the stability diagram. At a field larger than 0.4 T the IDT disappears, while reservoir transitions remain visible, and then re-appears at a field of 3 T. Such magnetic field dependence matches the energy diagram of a doubly-occupied DQD of even parity, where singlet and triplets states are formed, as shown in Fig. 2.5(c). As the external field is increased, the triplet $T_{-}(1,1)$ state becomes the ground state. Due to the Pauli exclusion principle the triplet energy is linear as a function of detuning which results in zero charge susceptibility and no phase response. At large $B_z$ the signal is recovered due to the $T_{-}(1,1)$ and $T_{-}(0,2)$ anti-crossing, whose energy is reduced for increasing field. Figure C.1(b) shows a finer scan of the IDT with magnetic field, where the IDT shifts to larger gate voltage with increasing field before disappearing. The shift to larger voltages indicates that the (2,0) charge configuration is found at positive detuning (as it tracks the $T_{-}(1,1)$ and $S(2,0)$ state crossing). This is consistent with a shift to lower voltage of the dopant transition at high field observed in a broader scan indicating predominant tunnelling of spin down electrons. Consequently, the donor transitions from the $D^+$ (ionised donor) to the $D^0$ (neutral donor) state, when an electron is transferred from the QD. The donor $D^+ \rightarrow D^0$ transition appears much broader than the QD transitions indicating a larger tunnelling rate to a reservoir such that the dopant could be located between the QD and one reservoir. In this analysis (1,1)-(2,0) charge occupations have been
Figure C.1: Even parity transition. (a) Charge stability diagram at increasing magnetic field from 0 T to 3 T with line-cuts at the inter-donor/dot transition (IDT). (b) IDT as a function of $B_z$. (c) Shift in gate voltage of the IDT $\Delta V_G$ with the magnetic field with linear fit (dashed line). (d) IDT as a function of calibrated energy detuning $\varepsilon$ with fit (dashed line) to obtain the tunnel coupling.
assigned to the IDT based on the fact that it is an even parity transition and that no additional lines at lower voltage have been observed. However, there could be an arbitrary offset in the number of electrons. Further parameters can be extracted from the shift of the IDT. The IDT gate coupling $\alpha = 0.36 \text{eV/V}$ is obtained from a linear fit as shown in Fig. C.1(c).

$$\alpha V_G = \frac{\Delta_c^2 - (2g\mu_B B_z)^2}{4g\mu_B B_z} \approx -g\mu_B B_z$$

(C.1)

Additionally, fitting of the IDT line-shape at zero field following Eq. (2.49) and shown in Fig. C.1(d) yields a tunnel coupling of $\Delta_c = 20.9 \mu\text{eV}$ and maximum capacitive shift of 0.5 fF.

Finally, there is some evidence that the QD is interacting with a phosphorus donor at this IDT. In Fig. 4.14(b) one additional transition at large top-gate voltage close to 0.55 V appears broader than the surrounding transitions. This signal can be attributed to the $D^0 \rightarrow D^{-}$ transition of the phosphorus donor. The difference in slope compared to the donor transition at lower voltage could be explained by the increased Bohr radius of the negatively charged donor leading to stronger top-gate coupling. From the separation in voltage between both donor transitions and given a top-gate lever-arm of $\alpha_P = \alpha_{\text{DRRT}} - \alpha_{\text{IDT}} \approx 0.5$ a charging energy of 37 meV is obtained, which is close to the expected binding energy of phosphorus in bulk of 44 meV and consistent with previous measurements in nano-devices where a slightly reduced charging energy is observed due to additional capacitive coupling to surrounding electrodes [312].
Appendix D

Back-gate operation with high-quality resonators

In silicon-on-insulator (SOI) devices as introduced in Section 3.1.1 and discussed in Chapter 4–6, a voltage applied to the silicon substrate acts as a global back-gate. This global back-gate typically allows tuning between a single and double quantum dot configuration and tuning of dot-reservoir and inter-dot tunnelling rates. To operate the back-gate, a blue LED is turned on to enable charge redistribution in the otherwise frozen-out substrate. When performing RF reflectometry using low quality resonators $< 50$, as demonstrated in Chapter 5 and in previous experiments [162], operation of the back-gate and LED only has a weak effect on the resonator quality factor. In this appendix, the effect of LED and back-gate operation on a high-quality factor resonator is discussed.

For high-quality resonators, as demonstrated in Chapter 4, there is a significant reduction in quality factor when operating the back-gate. Figure D.1(a) shows the loaded quality factor of the device demonstrated in Section 4.3 as a function of $V_{BG}$ applied to the back-gate with the LED on. The maximum quality factor is $< 800$, which is significantly smaller than 966 demonstrated in Fig 4.13. Consequently, operating the LED reduces the

![Fig. D.1: Back-gate operation. (a) Loaded quality factor of the readout resonator as a function of back-gate voltage. (b) Resonant frequency as a function of back-gate voltage.](image)
loaded quality factor and in such the readout sensitivity, see Eq. (4.7), which can be explained by charge re-distribution. For $V_{BG}$ between 0 V and 4 V, the quality factor remains approximately constant. Similarly, the resonant frequency shown in Fig. D.1(b) only changes very little in the same range. At smaller and larger $V_{BG}$ the quality factor drops very quickly below 100. This strong change in quality factor can be explained by a large change in capacitance evident in a shift of the resonant frequency of \(~5\text{ MHz}\) to lower frequency in Fig. D.1(b), possibly due to charge re-distribution or formation of electron gases at the silicon and buried oxide interface.
Appendix E

Silicon qubit summary and benchmarks

<table>
<thead>
<tr>
<th>spin qubit</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$Q_{\text{load}}$</th>
<th>$f$ GHz</th>
<th>$\tau_{\text{int}}$</th>
<th>$T_1$</th>
<th>$F_{\text{avg}}$</th>
<th>resonator</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOS [235]</td>
<td>0.1</td>
<td>2</td>
<td>38</td>
<td>0.266</td>
<td>2 ms</td>
<td>4.5 ms</td>
<td>73%</td>
<td>SMD</td>
</tr>
<tr>
<td>$^{31}\text{P}$ [313]</td>
<td>&lt; 0.3</td>
<td>2</td>
<td>300</td>
<td>0.34</td>
<td>300 $\mu$s</td>
<td>2 ms</td>
<td>83%</td>
<td>NbTiN spiral</td>
</tr>
<tr>
<td>SiGe [56]</td>
<td>&lt; 0.3</td>
<td>1.2</td>
<td>2600</td>
<td>5.7</td>
<td>6 $\mu$s</td>
<td>160 $\mu$s</td>
<td>98.4%</td>
<td>$\lambda/2$ NbTiN</td>
</tr>
</tbody>
</table>

Table E.1: Gate-based readout. Average readout fidelity $F_{\text{avg}}$ with integration time $\tau_{\text{int}}$ and related blockade time $T_1$ for different resonator and qubit implementations in natSi. Readout is performed either directly using Pauli spin blockade. Resonators are either based on surface mount (SMD) or superconducting components and are realised off-chip with the exception of the $\lambda/2$ NbTiN which is fabricated alongside QDs. Gate to quantum dot coupling $\alpha$, resonator coupling coefficient $\beta$, quality factor $Q_{\text{load}}$ and frequency $f$ are given.
Table E.2: Qubit readout using charge sensors. Average readout fidelity $F_{\text{avg}}$ with integration time $\tau_{\text{int}}$ and related blockade time $T_1$ for various qubit implementations. Spin readout is performed via spin-to-charge conversion either using spin-dependent tunnelling (SDT) or Pauli spin blockade (PSB) and is detected using a quantum point contact (QPC), single electron transistor (SET), sensor quantum dot (SQD) or tunnel coupled ancilla dot.
<table>
<thead>
<tr>
<th>spin qubit</th>
<th>$T_2^*$ (µs)</th>
<th>$T_2$ (ms)</th>
<th>$F_{c1}^q$ (%)</th>
<th>$f_{c1}^q$ (MHz)</th>
<th>$F_{c2}^q$ (%)</th>
<th>$f_{c2}^q$ (MHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P ($^{28}\text{Si, ESR}$)</td>
<td>270</td>
<td>560</td>
<td>99.6</td>
<td>&lt; 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P ($^{\text{nat}\text{Si,}}$ $J(\varepsilon)$) [42]</td>
<td>1.75 s</td>
<td>35.6 s</td>
<td>99.99</td>
<td>90 (gate)</td>
<td>1250</td>
<td></td>
</tr>
<tr>
<td>P$_{\text{nmuc}}$ ($^{28}\text{Si, NMR}$) [39]</td>
<td>120</td>
<td>28</td>
<td>99.6</td>
<td>&lt; 1</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>QD MOS ($^{28}\text{Si, ESR, }\delta E_Z$) [91, 104]</td>
<td>24</td>
<td>0.29</td>
<td>99.3</td>
<td>85 (Bell)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>QD MOS ($^{28}\text{Si, ESR}$) [44]</td>
<td>99.96</td>
<td>&lt; 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QD MOS ($^{28}\text{Si, ESR}$) [40]</td>
<td>2</td>
<td>0.08</td>
<td>98.6</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QD MOS ($^{28}\text{Si, ESR, }1.4$ K) [32]</td>
<td>20</td>
<td>3.1</td>
<td>99.9</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QD Si/SiGe ($^{28}\text{Si, EDSR}$) [41]</td>
<td>0.9</td>
<td>0.044</td>
<td>99</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QD Si/SiGe ($^{\text{nat}\text{Si, EDSR}}$) [99]</td>
<td>1</td>
<td>0.019</td>
<td>98</td>
<td>2</td>
<td>89 (Bell)</td>
<td>20</td>
</tr>
<tr>
<td>QD Si/SiGe ($^{\text{nat}\text{Si, EDSR}}$) [36]</td>
<td>1.4</td>
<td>0.08</td>
<td>99.7</td>
<td>78 (Bell)</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

Table E.3: Qubit coherence and control. Summary of dephasing ($T_2^*$), decoherence ($T_{2,\text{decoh}}$) times, single qubit control fidelity ($F_{c1}^q$) and frequency ($f_{c1}^q$), and two qubit control fidelity ($F_{c2}^q$), either as gate or Bell state fidelity, and frequency ($f_{c2}^q$) achieved using silicon qubit devices at milliKelvin temperature unless stated otherwise.
Appendix F

Notes on experimental procedures

Cleaving  Foundry-fabricated devices typically come on a chip (die) with dimensions on the order of 10 mm and contain many devices (ranging from 50 to a few hundreds). While large chips facilitate handling, cutting a die into smaller chips allows for experiments of similar devices in multiple setups at the same time. Furthermore, small chips have a reduced parasitic device capacitance when performing reflectometry measurements, which is desired for maximising readout sensitivity. Step by step small pieces of a die can be cut off using a diamond scriber and cleaving of the chip by aligning the cut with the edge of a glass slide (or a different object that has a small edge) and by applying a small amount of force. To create a deep cut with the diamond scriber, the diamond tip is passed over the same line two to three times. The force to break the chip across an edge is applied using both thumbs at either end of the chip, while covering the chip with a clean-room wipe that is folded multiple times. If devices are very sensitive or contain very fine structures at the surface, spinning of photoresist prior the cleaving process is suggested for additional protection. Manual cleaving is preferred over automatic cutting using a wafer dicing saw, as devices have been found non-functional after dicing in multiple cases.

Device identification  A room-temperature or even cryogenic probe-station is the ideal tool for testing many devices in a short amount of time. Experiments that can be performed in a probe-station are limited due to a small number of probes (4–6) and noise performance. Gate leakage and turn-on tests to identify promising devices are suited for such a limited setup. When operated at 4 K features of Coulomb blockade can even be observed, given that the charging energy is sufficiently large.

Typical dimensions of silicon nanowire transistor devices for quantum dot (QD) experiments range from $W = 30–60$ nm for the channel width and $L_g = 20–80$ nm for the gate length. At dimensions such as $W = 60$ nm and $L_g > 30$ nm QDs can
Figure F.1: Grounding. (a) Procedure to have the PCB grounded during bonding. (b) Procedure to keep the device grounded when transferring from the bonder to a measurements setup.

typically be tuned to a single QD configuration and show very regular Coulomb diamonds. At smaller dimensions, QDs show more disorder and can show strong inter-dot coupling and tuning into the single QD regime might not be possible.

**Bonding and grounding** Once a die is cut into small pieces, the chip can be transferred onto a PCB for bonding. Bonding of foundry-fabricated devices can be challenging due to a high-density bond-pad layout designed for automated testing and automated bonding. Bond-pad dimensions are approximately $70 \, \mu m \times 70 \, \mu m$, which is compatible with standard manual bonders which typically yield a bond size close to $60 \, \mu m$ (using a $25 \, \mu m$ thick wire). However, the pitch between pads can be as small as $10 \, \mu m$ in one direction which easily leads to shorts between two pads when a bond is not perfectly placed in the centre of a bond-pad. To reduce the chance of a short, a wire of $17.5 \, \mu m$ thickness (part no.: S-W-ALSI-0007/09) and a bonding wedge that produces bonds at a size of only $38 \, \mu m$ (part no.: MP-4WFV1-1515-W5H-M00) is chosen for bonding of a very dense pad layout. As the wire is thinner and the wedge has a smaller area, the bonding force, ultrasonic time and power for achieving reliable bonds should be reduced to a setting between 1 and 1.5.
In the foundry-fabricated devices presented in this thesis, no protection diodes are in place such that electrostatic discharge (ESD) can easily damage devices and measures for protection during device handling need to be taken into account. Particularly when bonding, where the electric connection from the device to the PCB is made, there is a high chance to break a device due to ESD. Therefore, grounding of the PCB is essential. This is achieved using a grounding plug that shorts all pads on the PCB together and ends in a crocodile clip for connection to the bonding stage, such that the bonding tip and PCB are at the same potential as shown in Fig. F.1(a). After bonding, the PCB is transferred into a metal box for safe transport and grounding as shown in Fig. F.1(b). Prior to transfer, the metal box can be brought onto the same potential as the PCB using an additional cable with crocodile clips at each end (not shown).

**Device testing** Once a device is bonded, it is transferred onto a measurement system which includes making an electrical connection to the device for measurements. Every part of the measurement systems should be grounded, which can be checked using a multimeter. Some parts of cryogenic systems might be isolated on purpose to prevent heat transfer. Such floating sections can be grounded using crocodile clip cables to avoid any ESD when transferring the PCB onto the system. Once the PCB is attached, the grounding plug can be removed and the electrical connection to the measurement lines can be made. Grounding of the measurement lines should be ensured, either using switches on a breakout-box or an additional grounding plug.

Devices can be tested for basic functionality at room-temperature. For this a small bias (1 mV) can be applied to the source and a current can be measured either using a lock-in amplifier or using a current pre-amplifier and multimeter via the drain. A source-meter can be used to supply a gate voltage and measure any leakage current. To begin, leakage can be tested by increasing the gate voltage in small steps of 10 mV. When no leakage is observed the voltage can be stepped in larger steps of up to 0.1 V. For silicon nanowire transistors a turn-on should be observed between 0.2–0.6 V and the gate voltage should not exceed 1.5 V to avoid any damage. For Si-MOS planar devices presented in Appendix A turn on is typically observed between 1.5–2 V and the gate voltage could be supplied to all required gates or only a single gate while others are kept floating. Gate voltages should not exceed 4 V to avoid damage. When operating a Si-MOS device for the first time, a turn on might not appear immediately and only after holding the gate voltage above threshold for up to 10 s, especially when some gates are floating.
Reflectometry  To perform reflectometry measurements, first the readout resonance needs to be identified. For this, transmission from the reflectometry drive line to the readout line is measured using a vector-network-analyser (VNA) (see Fig. 3.3). When the resonance is of high quality factor, identification of the resonance can be quick and simple. However, if the resonance is of low quality factor and especially if circulators are present on the readout line, identification of the resonance can be difficult. In this case, comparing spectra when operating at a gate voltage below and above threshold could produce a small change in the resonance which could be observed when using the maths-divide or calibration function of the VNA. Operation of the LED and back-gate (if available) can be helpful as the additional capacitance typically leads to significant and visible resonance frequency shift (without division or calibration) as a function of back-gate voltage. The VNA measurement additionally provides an indication of the relative power level of the drive signal at the readout line output which is attenuated, reflected by the resonator and amplified. Alternatively, a spectrum analyser can be used to determine the absolute power level of the RF drive at the readout line output for a given drive input power.

Once the resonance is identified, a reflectometry setup as shown in Fig. 3.6 can be used for taking first reflectometry measurements. The RF source should be operated at a frequency close to resonance of the readout resonator. If a source without a separate LO port (such as the SMB100A) is used, an output power close to 0 dBm should be selected and the room-temperature attenuator connected to the drive line should be chosen such that an RF power on the order of $-80$ dBm or $-110$ dBm, for a low or high quality resonator respectively, is delivered at the device at milliKelvin temperature (typically ranging from 10–40 dB). If a source with separate LO output is used, no directional coupler and attenuator is required at room-temperature and the LO output can be directly connected to the LO input of the IQ mixer and the power level delivered to the drive line can be set on the source independently (output power). The amplifiers between the readout line output and IQ mixer RF input should be chosen such that the drive signal power level at the RF input is on the order of 0 dBm, which means that the signal power at the LO and RF port is of the same magnitude (typically 60 dB amplification).

To be able to observe a reflectometry signal even when parameters are not optimised, it is important to include additional amplifiers at the $I$ and $Q$ outputs of the IQ mixer combined with low-pass filtering. To begin, a low cut-off frequency, such as 1 kHz, and a low frequency of the ramp/triangular waveform delivered to a gate of the device, such as 83 Hz, should be selected. At this initial stage, it is often
useful to use an oscilloscope rather than a digitizer to have direct feedback while manually tuning parameters. The oscilloscope should be selected to trigger from the SYNC output of the waveform generator and AC coupling is chosen to be very sensitive to small changes in the $I$ and $Q$ signal due to single-electron transitions in the device. Moreover, averaging over a few hundreds of traces should provide sufficient signal-to-noise ratio. Next, the offset and amplitude can be increased such that the ramp falls within a range where Coulomb blockade has been observed in transport. Features should appear at the same voltage in the $I$ and $Q$ signal. If features only appear in one channel, the RF frequency could be tuned such that a signal appears in both channels. If no features are appearing, a different range, starting at smaller or larger voltage with a smaller or larger amplitude, could be selected for the gate voltage. Moreover, slight variations in RF frequency and power or additional averaging could make signals more visible. Once a signal is found, all parameters can be optimised to maximise signal-to-noise ratio at the desired measurement bandwidth, while keeping the RF power below power broadening. Finally, to verify that the features represent Coulomb blockade oscillations, a stability diagram should be recorded.
Appendix G

Code

The code presented in this appendix can be found on my personal github repository: https://github.com/simonschaal

Resonant circuit simulations

The lcapy python package is a great tool for analysing electrical circuits analytically and numerically. Based on lcapy, sympy and numpy the RFCircuit class has been developed to facilitate radio-frequency circuit simulations. An example calculation of the reflection coefficient for different values of coupling capacitance in a parallel resonator circuit is shown below.

```python
from rfcircuit import RFCircuit
from lcapy import *
import numpy as np
from matplotlib import pyplot as plt

mycirc = RFCircuit()
# simple syntax to create a circuit: + for series, / for parallel
mycirc.set_circuit(C('Cc') + ( R('Rd') | C('C0') | L('L')))

# analytical sympy expressions for circuit parameters can be obtained
mycirc.Z # impedance
mycirc.Gamma # reflection coefficient
mycirc.wres # resonant frequency
```
By assigning physical values to the circuit parameters numerical simulations can be performed.

```python
# circuit can be visualised
mycirc.draw_circuit()

# fixed simulation parameters
args = {mycirc.args['L'] : 405e-9, mycirc.args['C_0'] : 480e-15,
        mycirc.args['R_d'] : 800e3}
gamma = mycirc.get_Gamma(args)  # returns gamma(Cc, w)

# obtain gamma for a range of Cc and f
f = np.linspace(200e6, 400e6, 1000)
cc = np.linspace(1, 500, 501)*1e-15
data = gamma(cc[:, np.newaxis], 2*np.pi*f)

# show gamma for selected values of Cc
plt.figure()
plt.xlabel('$f$ (MHz)')
plt.ylabel('$\Gamma$')
plt.plot(f*1e-6, np.abs(data)[::50,:].transpose())
```
Fidelity simulations

A numerical model for singlet-triplet readout fidelity as introduced in [234] has been implemented. The singlet and triplet probability density, readout fidelity and optimal threshold can be calculated. Exemplary code that can be used to generate Fig. 4.1 is shown below.

```python
from fidelity_helpers import *
import numpy as np
import matplotlib.pyplot as plt

# given measurement signal and noise
signal=1.
noise=0.2

# signal voltage range
V=np.linspace(-2*signal, 2*signal, 10000)

# probability density for 100us integration time and 500us relaxation time
nS, nT = probST(V, signal, noise, 100e-6, t1=500e-6)

# find optimal threshold, -0.075
Vt=optimizeVt(-0.5,0.5,100, signal, noise, 100e-6, t1=500e-6)

# generate plot
```
```python
plt.plot(V, nS, label='$P_S$')
plt.plot(V, nT, label='$P_T$')
plt.plot(V, nS+nT, label='$P_S + P_T$', ls='--')
plt.ylabel('$P$')
plt.xlabel('$\Delta\phi_{\text{norm}}$')
plt.legend(frameon=False)

# visibility as a function of integration time

tmin=1e-6#us
signal=1
tau=np.logspace(-7, -1, 200)
trelax=10e-3
noise=np.sqrt(tmin/tau)

fidel = np.array([fidelity(0, 1, np.sqrt(tmin/elem), elem, t1=trelax) for elem in tau])

plt.figure()
plt.semilogx(tau*1e6, -1+np.sum(fidel, axis=1), label='1us, 10ms')
plt.ylabel('$V$')
plt.xlabel('$t_{\text{int}}$ (s)')
```

**QCodes**

All measurements were taken using the Python-based data acquisition framework developed by the Copenhagen, Delft, Sydney and Microsoft quantum computing consortium to perform nano-electronic device experiments. The QCodes framework handles instrument communication, data-saving and visualisation on a high and efficient level through python libraries such as *numpy, pandas, sqlite, matplotlib* and *pyqtgraph* facilitating quantum dot experiments. 

[https://github.com/QCoDeS/Qcodes](https://github.com/QCoDeS/Qcodes)

**QDev-wrappers**

The *QDev-wrappers* are a selection of functions that make repetitive tasks in an experiment more efficient. The *station-configurator* allows loading of instruments
in a single-line command using settings that are pre-defined in YAML file. Moreover, the *doNd* functions allow for simple N-dimensional sweeps in a single-line command. https://github.com/qdev-dk/qdev-wrappers

**Measurement notebooks**

Using *QCodes*, measurements can be performed in a python shell by executing commands step by step, by writing code in a file and running an experiment in one go or using a *jupyter* notebook. Notebooks have the advantage that measurement code can be accompanied by comments and graphs in a chronological order. However, notebooks can sometimes promote bad programming practices such as duplication of cells, rather than writing generalised functions or classes for repeatedly used code. Moreover, code in notebooks is hard to test and it is not possible to perform code versioning or merging. Nevertheless, notebooks are a great way for exploring a device and to perform simple measurements. More complicated experiments should be developed within regular python files for proper versioning and testing which could afterwards still be executed from a notebook.

**Useful measurement code snippets**

**Monitor** The *QCodes* monitor provides an overview over selected parameters and their current values in a separate webpage. The monitor runs in a separate python shell and the snippet below shows how to run a separate shell in the background in a jupyter notebook cell using the `%script` notebook magic command. Once an instance of the monitor class is created, where parameters that should be monitored are passed within the constructor, the monitor web-page should be filled with current parameter values.

```
%%script cmd --bg
python -m qcodes.monitor.monitor

monitor = qcodes.Monitor(param1, param2, param3, ...)
```

**Stepping** Large jumps in voltage might be harmful for a nano-scale device. The step and *inter_delay* attribute of a parameter allows to define a maximum step and an effective ramp rate.
```python
vsd.step = 0.001
vsd.inter_delay = 0.3
```

**Exemplary yaml file**  When using the *station-configurator* of the *QDev-wrappers*, both parameter monitoring and stepping can be defined within the *YAML* file. Individual setting of the attributes and creating an instance of the monitor class is unnecessary.

```yaml
instruments:
  VNA:
    driver: qcodes.instrument_drivers.Keysight.Keysight_E5071C
type: Keysight_E5071C
address: TCPiP0::169.254.71.72::inst0::INSTR
enable_forced_reconnect: true
parameters:
  timeout: {initial_value: 1000}
  start: {initial_value: 500000000, monitor: true}
  stop: {initial_value: 900000000, monitor: true}
  keith:
    driver: qcodes.instrument_drivers.tektronix.Keithley_2400
type: Keithley_2400
init:
  address: GPIB0::26::INSTR
  terminator: "\n"
  enable_forced_reconnect: true
parameters:
  volt:
    monitor: true
    step: 0.1
    inter_delay: 0.3
```

Instruments can then be loaded via the *station-configurator* and simple measurements can be performed using the *doNd* convenience functions.
STATION = qc.Station()
SC = StationConfigurator('setup.yaml', station = STATION)

VNA = SC.load_VNA()
doid(VNA.power, -40, -10, 31, 0.001, VNA.trace)

Custom parameters  Often it can be useful to define a custom virtual parameter as a function of physical parameters. This can be easily done using the qcodes.Parameter class as shown at the example of a symmetric source-drain bias voltage below. To every parameter a validator can be assigned that allows only specific input values.

```python
def set_Vsd(volt):
source.offset(volt/2)
drain.offset(-volt/2)

vsd = qcodes.Parameter('Vsd', label='Vsd', unit='V',
set_cmd=set_Vsd, vals=qcodes.utils.validators.Numbers(-50e-3,
50e-3))

vsd(1e-3)
```

Widgets  Widgets within a jupyter notebook provide a simple way to manually change parameters by entering numbers into text field or using sliders. Below an example for creating a floating number text widget to manually tune the pump signal frequency of a Josephson parametric amplifier is shown. More widgets can be found at:

```python
import ipywidgets as widgets

pumpfreq=widgets.FloatText(
    value=pump.frequency(),
```
Video mode  While *matplotlib* is a great python library to produce production-quality graphs, generating and updating graphs is slow. *Pyqtgraph* is a scientific plotting graphics library that is embedded within the *Qt* framework. Generating graphs is not as simple as using *matplotlib* but *pyqtgraphs* can be embedded into any *Qt* application allowing to create graphical interfaces for measurements with very fast rendering. *QCodes* offers a *QtPlot* class for generation of simple *pyqtgraphs* based on *qcodes* datasets. Below code to realise video-mode measurements using the *QtPlot* class is presented at the example of fast reflectometry measurements. The code can be adapted to any other two-dimensional measurement and could be extended to realise a complete graphical measurement interface.

To begin, the fast reflectometry measurement is setup and a *QtPlot* is generated from an initial reflectometry measurement.

```python
# Uses station-configurator
alazar=SC.load_Alazar()
awg=SC.load_AWG()
reflectometry=SC.load_AlazarFastReflectometry()

reflectometry.sample_rate(2e6)
# amplifier is already AC coupled
reflectometry.input_coupling('DC')
reflectometry.input_range(0.8)
reflectometry.impedance(1000000)

# setup AWG voltage space for fast and slow ramp
reflectometry.y_start(0.53)
reflectometry.y_end(0.54)
reflectometry.y_npts(256)
```
reflectometry.x_start(-6e-3)
reflectometry.x_end(6e-3)
reflectometry.x_npts(256)

# enter I and Q DC offset for accurate calculation of phase in AC coupling mode
reflectometry.I_DC(6.4)
reflectometry.Q_DC(5.3)

reflectometry.awg=awg
# ramp to setpoint for x and y, and calculate frequency from sample rate and npts
reflectometry.setup_AWG()
reflectometry.acquisition.setup_sweep(buffers_per_acquisition=10) # 10 averages

# create plot from one measurement
# alternatively use:
# plot, dataset = do0d(reflectometry.acquisition)
# which will create additional graphs for mag and phase
plot = qcodes.QtPlot()
dataset=qcodes.Measure(reflectometry.acquisition)
plot.add(dataset.Alazar_I)
plot.add(dataset.Alazar_Q)
Next, within a while-loop new measurement data is acquired and the selected graphs are updated. A waiting time between individual updates can be defined. To stop live updates the loop needs to be interrupted.

```python
import time
wait = 0  # seconds

# select I and Q array to plot, choose 2,3 for mag and phase
selection=[0,1]

# loop until interrupted
while True:
    # returns tuple of I,Q,mag,phase data acquired in less than a second
    # data is not saved on the harddrive
    data=reflectometry.acquisition()
    # update graphs
    for i in range(len(selection)):
        po=plot.traces[i]['plot_object']
        po['image'].setImage(data[selection[i]].transpose(),
                             levels=po['hist'].getLevels())
    time.sleep(wait)
```
Bibliography


