DELFT UNIVERSITY OF TECHNOLOGY

MASTER THESIS

EXPLORING FREQUENCY RE-USE USING TRANSMON QUBITS IN A CQED ARCHITECTURE

Author:
Serwan Asaad

Supervisors:
Dr. Alessandro Bruno
Christian Dickel

June 9, 2015

Contents

This experiment has been published in Applied Physics Letters **106** 18 (2015).

Chapter 1

Introduction

Quantum circuits based on circuit quantum electrodynamics (cQED) have experience rapid development of the past few years. Especially transmon qubits, and variations thereof, have flourished up to the point where they are used to perform multi-qubit algorithms. Fedorov et al. [9] first demonstrated implementation of the Toffoli gate, a key ingredient for error detection schemes. The Grover's search algorithm, and the Deutsch-Josza algorithm have been demonstrated for two qubits by DiCarlo et al. [8]. Furthermore, the first quantum error detection algorithms have been implemented by Córcoles et al. [7], and by Ristè et al. [27].

There seems to be no inherent limit preventing a cQED-based quantum computing framework from scaling up to hundreds, possibly even thousands of qubits. One of the most promising quantum computing architecture is the surface code **TODO:** cite. It has the advantage of having a relatively high fault-tolerance threshold, and only requires nearest-neighbour qubit couplings.

That being said, even in an architecture such as the surface code, the challenges in scaling up are tremendous. Qubits are required to have long coherence times, high fidelity gates with a short duration, and must be arranged such that it is compatible with the architecture. Aside from on-chip challenges strong requirements are placed on supporting equipment. The chip must be cooled down to the milikelvin temperature range, and each of the qubits must be individually controlled.

My Master's project has been largely devoted to addressing some of the challenges accompanied with scaling up. To this end I have mainly been involved in two experiments, and so this thesis is separated into two parts.

During the first three months of my Master's project I was involved in the characterization of high quality factor coplanar waveguide resonators. These resonators achieve their high quality factors through a combination of deep-reactive ion etching and HMDS surface treatment. High quality resonators are necessary for long coherence time qubits, as the quality factor places an upper limit on the relaxation time T_1 of a coupled qubit through the Purcell effect. **TODO:** formula

In Chapter 2 the theory behind coplanar waveguide resonators is explained. The focus of this chapter is on the different loss mechanisms that reduce the resonator quality factor, and the different methods used to minimize these losses. The two methods making this experiment unique are the use of deep-reactive ion etching and HMDS surface treatment.

The experimental set-up, including the resonator chip, is described in Chapter 3. In Chapter 4 the results are discussed. We find that quality factors over 1 million can be attained in the single-photon regime.

The second and main part of my Master's project was spent on the Muxmon experiment. In this experiment two of the challenges in scaling up are addressed, namely frequency crowding and instrument scaling. A key ingredient in this experiment is the Duplexer, invented by Duije Deurloo. The Duplexer is a vector switch matrix having four input channels and two output channels. It can control the signal through each of the input-output combinations, thereby selectively directing qubit pulses. This allows for selective broadcasting, which enables the control of two qubits simultaneously, consequently reducing the amount of instruments required. These subjects are discussed in Chapter 6.

To test the concept of selective broadcasting the Muxmon chip was created. This chip consists of three qubits, two of which will be tuned to the same frequency, so that they can be controlled via the Duplexer. Chapter 7 is devoted to the measurements performed to characterize the Muxmon chip.

Since we want to know the best attainable qubit performance under these conditions the qubits must be accurately tuned. In Chapter 8 the calibrations used to tune the qubits are explained.

Measuring the performance of the qubits is done through randomized benchmarking. This is a method based on repeated application of unitary operations on the qubit, from which the gate performance can be extracted. Randomized benchmarking will be discussed in Chapter 9.

Part I

Deep-reactive ion etched resonators

This experiment has been published in Applied Physics Letters **106** 18 (2015).

Chapter 2

Theory

2.1 Coplanar waveguide

In the context of circuit QED, one of the most common types of resonators are coplanar waveguides (CPW). Coplanar waveguides consist of a long central conducting track, with on both sides a neighbouring grounded track. The conducting track is seperated from the grounded tracks by a fixed distance.

One end is usually capacitively coupled to a feedline and has an open end, while the other end can either be open or shorted. This determines whether the current has a node or an antinode at that end. In the case of a shorted end, current has an antinode at that end, resulting in a quarter wave resonator. This means that the wavelength of the fundamental mode fits a quarter times into the resonator. In the case of an open end, the current has a node at that end, resulting in a half wave resonator.

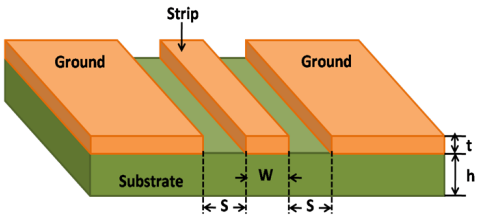


Figure 2.1: Schematic of a coplanar waveguide.

2.2 Quality factor

The quality of a resonator can be quantified through its quality factor. Generally speaking, the quality factor of a resonator determines the ratio between energy stored in a resonator and the energy leaking away from the resonator. For cQED resonators this corresponds to the rate at which photons dissipate from the resonator. A high quality factor corresponds to a low dissipation rate.

The quality factor can also be defined in two different ways [19, pp.23-24]:

$$Q = \omega_0 \tau_1 = \omega_0 / \Delta\omega \tag{2.1}$$

Here ω_0 is the resonance frequency of the resonator, and τ_1 is the decay time of the resonator.

The decay time is the time taken by a resonator to dissipate its energy to $1/e$ of its original energy.

Photons can dissipate from the resonator through the resonator's different loss channels. Each of these loss channels has a corresponding quality factor. One such loss channel is due to resonators in cQED being capacitively coupled to a feedline. The quality factor associated to this loss channel is known as the coupling quality factor Q_c. This coupling quality factor depends on the amount of capacitive coupling between the resonator and the feedline. It can therefore be engineered to have a certain value, depending on the amount of interaction wanted between resonator and feedline.

The other loss channels are usually unwanted, and therefore desired to be as low as possible. These individual channels are usually lumped together, resulting in a combined quality factor, known as the intrinsic quality factor Q_i.

The total quality factor of the resonator is known as the loaded quality factor Q_l. It is related to Q_c and Q_i through:

$$\frac{1}{Q_l} = \frac{1}{Q_c} + \frac{1}{Q_i} \tag{2.2}$$

From equation 2.2 it can be seen that if the difference between Q_c and Q_i is large, then the loaded quality factor Q_l will be approximately equal to the minimum of the two.

For a quarter wave resonator the amplitude of transmission has a minimum S_{21}^{min}, given by [19, p29]:

$$S_{21}^{min} = \frac{Q_c}{Q_c + Q_i} \tag{2.3}$$

With knowledge of the resonant frequency ω_0, the resonant width $\Delta\omega$, and the transmitted signal at resonance S_{21}^{min}, it is possible through equations 2.1 and 2.3 to determine both the coupling quality factor Q_c and the intrinsic quality factor Q_i. Note that as equation 2.3 depends on the ratio of the two quality factors,to get an accurate estimate of both quality factors, they should have a comparable value.

One reason why a high quality factor is important in the context of cQED is that a qubit can be coupled to a resonator. This qubit therefore experiences dissipation due to its coupling to the resonator, known as the Purcell effect. The result of dissipation is that when the qubit is in its excited state, it will relax to its ground state. The amount of relaxation due to the Purcell effect can be quantified through its relaxation time T_1^{Purcell}. The reason a high quality factor is important is because the Purcell relaxation time is proportional to the quality factor [11, p 22]. The Purcell relaxation time T_1^{Purcell} places an upper limit on the relaxation time T_1 of a qubit. If the qubit's relaxation time T_1 is close to this value, the qubit is said to be Purcell limited.

2.3 Losses

When a resonator is being driven at its resonance frequency, it is absorbing photons from the external source. When this external driving stops, the resonator slowly loses its photons through its different loss channels.

One loss channel has already been discussed in section 2.2, namely through the coupling to the feedline. This loss channel is not unwanted, as the amount of coupling to the feedline determines how fast the resonator and feedline can interact with each other. The other loss channels, however, are unwanted. They cause dissipation of energy, and hence information. Some of the main causes of loss will be discussed in this section.

2.3.1 Causes of loss

2.3.1.1 Two-level systems

Two-level systems (TLS) are systems which can be in a ground state or in an excited state. In some cases they can be useful. In fact a qubit itself is an example of a TLS. In other cases, however, TLS can also be a source of dissipation such as in the case of dielectric loss [18]. Study suggests that in cQED, most TLS reside in a thin oxide layer at the metal-substrate interface and the substrate-air interface [33].

Resonators are surrounded by a large quantity of TLS, each of which has its own resonance frequency, depending on its energy landscape. When the resonance frequency of a TLS is close to that of the resonator, it can absorb a photon from the resonator, upon which it tunnels to an excited meta-stable state. TLS have a finite lifetime in their excited state, after which they decay back to their ground state and are then again able to absorb a photon. The rate at which a TLS absorbs a photon depends on the electric field surrounding the TLS.

In the low power, low temperature regime, TLS reside mostly in their ground state, and only occasionally tunnel to the excited state, upon absorption of a photon. It is theorized that, in this regime, TLS are the main source of dissipation for resonators [10]. At higher powers and/or temperatures, TLS will tunnel to an excited state at a higher rate. Due to their finite lifetime they become saturated at a certain point. Since the quality factor depends on the ratio between energy stored and energy dissipated, when the TLS are saturated the amount of dissipation is limited, while the energy stored in the resonator can still increase. Therefore, in the low power, low temperature regime, increasing either of the two parameters results in an increase in quality factor. At a certain point, however, further increasing either of the two will not improve the quality factor. This is due to other effects dominating the dissipation rate in these regimes.

2.3.1.2 Quasiparticles

Another source of dissipation for resonators is due to quasiparticles being present in the superconducting layer. When a Cooper-pair is broken up, Bogoliubov quasiparticles are formed [2, p16]. Once formed, the quasiparticles have a finite lifetime, depending on the temperature of the system. These quasiparticles can have either electron-like or hole-like properties. They are a source of dissipation for resonators, since they are non-superconducting and therefore cause the surface impedance to be slightly resistive [19, p18].

The breaking up of Cooper-pairs is due to excitations. These excitations can either be thermal, or due to photon absorption. Therefore an increase in temperature or an increase

in photon density will result in a higher density of quasiparticles. The quasiparticle density increases exponentially with increasing temperature [19, p44].

2.3.1.3 Radiation

A third source of dissipation is due to radiation from the resonator. This radiation is due to the spontaneous emission of photons.

The amount of dissipation due to radition is directly related to the geometry of the resonator through [28, 19]:

$$Q_{\text{rad}} = \alpha \left(\frac{L}{s+w} \right)^{n_r} \tag{2.4}$$

As shown in Figure 2.1, s is the distance between the conducting and grounded track, w is the width of the conducting track, and L is the length of the resonator. The parameter α depends on properties such as impedance and the dielectric constant of the substrate, and n_r depends on the shape of the resonator, and is equal to 2 in the case of a straight resonator. From Formula 2.4 it is clear that a decrease of the conducting track width or the distance between tracks leads to an increase in Q_{rad}. However, with a decrease of either of the two parameters, the field strength close to the resonator becomes higher. If the TLS are not saturated (i.e. low power and temperature), this will increase the amount of dissipation through TLS. Therefore it is not necessarily advantageous to minimize s and w.

Radiation loss becomes the dominant source of dissipation at high powers and/or temperatures, but otherwise usually is not the limiting factor. Since measurements relevant for quantum computing are usually operated at low power and temperature, this source of dissipation is usually less important than other sources, such as TLS dissipation.

2.3.1.4 Vortices

When a sample is cooled down to a superconducting state there may still be a small, but nonnegligible magnetic field present. The presence of a magnetic field can cause vortices to appear in superconducting materials. These vortices have a non-superconducting core. Current passing through superconductive materical exerts a Lorentz force on vortices. For a resonator being driven on resonance, this AC current results in the vortices near the resonator experiencing a dissipative oscillatory motion [25].

It is interesting to note that the presence of vortices does not necessarily lead to a lower internal quality factor. The influence of a vortex on a resonator depends on its location. As reported by Nsanzineza et al. [23], a vortex close to a current antinode of a resonator, can result in a significant loss of the quality factor. A vortex close to a current node, however, may even increase the quality factor of the resonator. They attribute this increase in quality factor to quasiparticles, which would otherwise lead to dissipation, being trapped in the vortex.

2.3.2 Minimizing losses

2.3.2.1 Surface treatment

Previous research has determined that for resonators, TLS are mostly present at the surfaces [10]. These oxides may reside at the interface between metal and dielectric, or between the dielectric and vacuum, or possibly between metal and vacuum (depending on the type of metal used). One explanation for TLS being present is the presence of an amorphous oxide layer at the interfaces. These oxides may act as TLS. During deposition of the metal on the dielectric, this oxide layer can become trapped between the two interfaces. For a silicon dielectric, this oxide layer can be removed by shortly treating the sample with hydrophluoric acid. This process is also known as an 'HF dip'.

Aside from the HF dip, additional surface treatment can be applied. For the resonators measured in this report, before depositing the metal on the substrate, an additional exposure to hexamethyldisilazane (HMDS) was applied. The reason for this additional step is that there is a lattice mismatch between the metal and substrate. The intermediate layer of HMDS can possibly mediate this lattice mismatch. See [5] for more information.

2.3.2.2 Infrared shielding

Aside from thermal excitation, quasiparticles are also formed from the absorption of photons. High-frequency photons (UV-range or higher) are usually not a significant contribution, as they are easily absorbed by materials, well before they reach the inner layers of the fridge. Lower-frequency photons, such as in the infrared range, however, can penetrate through the fridge to the sample. These infrared photons can cause the excititation of quasiparticles. By using infrared shielding, such as a coating film inside the fridge, the amount of infrared radiation reaching the sample can be lowered.

2.3.2.3 Deep-reactive ion etching

Another technique applied to the resonators studied in this report is deep-reactive ion etching (DRIE), which is a type of Bosch process [5]. In this technique two alternating steps are performed:

1. An etching step in which an SF_6 plasma is used to etch the substrate layer.
2. A passivation step in which C_4H_8 is released. The gas forms a protective layer on the substrate, except for the direction in which the etching plasma is accelerated. The result is that the sidewalls are protected from the etching process

Using DRIE, nearly vertical sidewalls can be created for the substrate. The result is that the substate-air interface is removed from the regions between the CPW tracks, which are the regions where the electric field strength is high. As the dissipation due to TLS depends on the electric field strength, it is expected that DRIE will result in a lower TLS dissipation rate at this interface.

2.3.2.4 Magnetic shielding and vortex trapping

Vortices are created when a magnetic field is present. One method to lower the amount of vortices is to use proper magnetic shielding around the sample. Furthermore, using nonmagnetic materials also result in lower amounts of vortices being present.

Even when using these methods to counter the presence of magnetic fields, there may still be a small amount of vortices present in the sample, which may lead to dissipation. To counter their movement a grid-like structure can be added in the superconducting material, effectively pinning the vortices.

Chapter 3

Experimental set-up

The fridge used in this experiment is a dilution refrigerator, made by Leiden Cryogenics. The refrigerator has a base temperature of ~15 mK. An input signal was generated using a Rohde & Schwarz ZVM vector network analyzer, connected to an Aeroflex 8310 step attenuator, which has an attenuation range of 120 dB. The signal out of the fridge was measured using the same vector network analyzer.

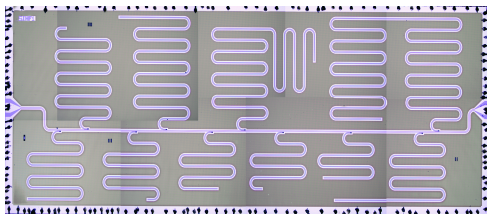


Figure 3.1: Optical microscopy image of the sample measured in this report. The sample consists of ten quarter wave resonators, with frequencies ranging between 1–11 GHz, connected to a central feedline. The resonators are made using NbTiN on a Si substrate. The sample is treated with HMDS and DRIE.

Using this set-up, quarter wave resonators, fabricated by Alessandro Bruno, were measured in a frequency range between 1–9 GHz. The sample is shown in Figure 3.1. The resonators were made using NbTiN on a silicon substrate. The advantage of NbTiN is that the metal atoms are bound to nitrogen, thereby inhibiting bond formation with oxides. In a way to minimize losses, all resonators were treated with HMDS and deep-reactive ion etching.

By driving a signal through the feedline, the resulting transmitted signal S_{21} can be measured. At or close to the resonance frequency of a resonator, the resonator will interact strongly with the feed line, resulting in a reduction in transmission.

Unless stated otherwise, all measurements were performed with the fridge at base temperature (~15 mK).

Because the relevant regime where resonators interact with qubits is the single-photon regime, a very weak signal must must be applied to determine its properties in that regime. At these low powers noise becomes a relevant issue. In Appendix A the noise of the system is characterized.

Chapter 4

Results and discussion

4.1 Resonator measurement

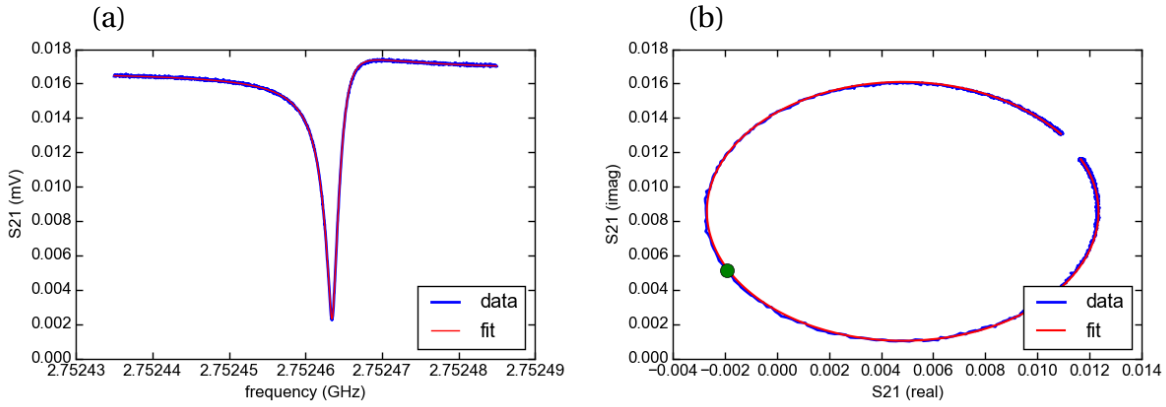


Figure 4.1: Forward transmission S_{21} spectrum of a resonator around 2.75 GHz. Panel (a) shows the amplitude of S_{21}, along with a fit (red). Panel (b) shows the path of S_{21} in the complex path, along with a fit (red). The green dot indicates the resonance frequency of the resonator. Measurement was performed at 15 mK at an input power of −123 dBm corresponding to $\sim 5 \times 10^4$ photons.

In Figure 4.1 the transmission S_{21} of a resonator is shown. Figure 4.1(a) shows the transmitted voltage $|S_{21}|$ of the resonator as a function of frequency. As one can see, the resonator has a shape similar to a Lorentzian dip. One interesting point is that the Lorentzian exhibits an asymmetry, which is often attributed to reflections in the feedline [11, p192]. This could be caused by impedance mismatching.

4.2 Power dependence

To be able to study the behaviour of the resonators, measurements were performed for several powers. Using proper calibrations for the attenuation down to the sample, the power can be

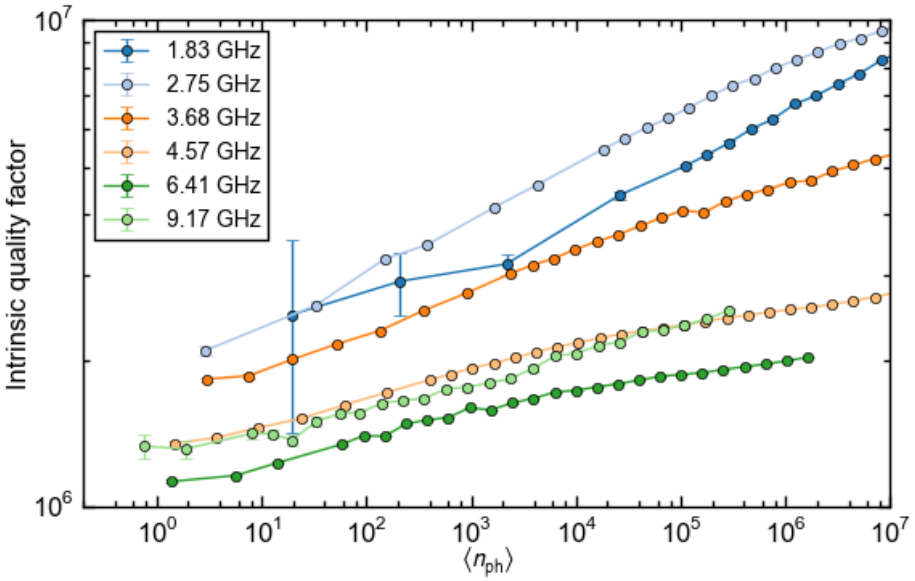


Figure 4.2: Intrinsic quality factor of resonators as a function of mean number of photons present in the resonator. Measurements were performed at 15 mK.

converted to the input power at the sample. This value can then be converted to the mean number of photons present in the resonator [5]. The results are shown in Figure 4.2.

As can be seen, the internal quality factor Q_i of all resonators decrease with decreasing photon number. One explanation for this phenomenon is that the dissipation dissipation is mainly due to TLS. Since measurements were performed at ~15 mK, the TLS are not saturated since the rate of thermal excitation is low. As discussed in section 2.3.1.1, the relative loss due to TLS is highest at low power, in the regime where they are not saturated. Therefore the fact that the internal quality factor Q_i rises with the mean number of photons present in the resonator can be attributed to a larger amount of TLS being satured. This would suggest that, even with HMDS and DRIE treatment of the sample, at low power and temperature, the internal quality factor is still limited by TLS being present.

The mean photon number of a resonator is inversely proportional to the square of frequency [5], so for resonators with a low frequency a lower input power is required than with a high frequency. At high photon numbers this is not a concern, as the transmitted signal is high enough to be accurately measured in a short period of time. For the single-photon powers, however, which is the region of interest for quantum computation, acquiring enough signal took up to five hours for the lowest frequencies. The reason that for the resonator with a resonance frequency at 1.83 GHz has large error bars at low powers can be partly attributed to this, but as we will see in section A.4, the main reason is that its frequency lies outside the bandwidth of the amplifiers and circulators of the set-up, resulting in a large amount of additional noise.

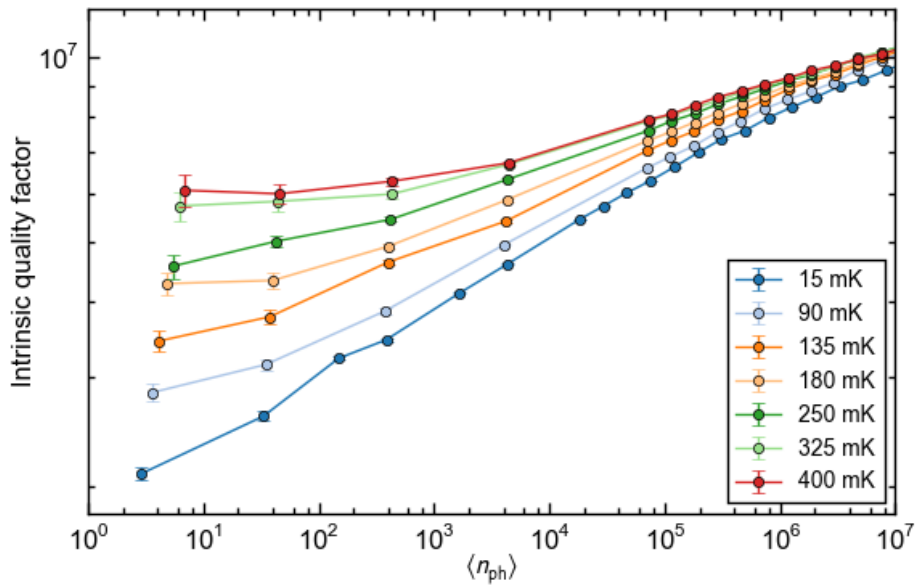


Figure 4.3: Intrinsic quality factor versus photon number for temperatures ranging from 15 mK up to 400 mK. All measurements were performed for a resonator with resonance frequency f_0 =2.75 GHz.

4.3 TEMPERATURE DEPENDENCE

Aside from power, some of the dissipation channels also depend on the temperature of the system. To be able to study the effect of temperature on resonators, the resonator with frequency 2.75 GHz has been studied as a function of power for several temperatures ranging from 15 mK up to 400 mK. The reason for choosing this resonator is that it has the highest internal quality factor of all the resonators measured, and so any change in quality factor would be most clearly visible.

The results are shown in Figure 4.3. As can be clearly seen, the quality factor increases with increasing temperature. This is likely due to the fact that TLS are thermally excited for a larger percentage of time. Therefore, the relative energy dissipation with respect to total energy in the resonator will be lower, resulting in an increase in quality factor.

Another interesting point is that the increase in quality factor as a function of temperature is largest at low powers. This can also be explained when the limiting factor is due to TLS. With low powers, the TLS are almost exclusively excited thermally, while at higher powers, the excitation of TLS is not only due to thermal excitations, but also from photon absorption.

If one looks at the highest temperatures, it seems that the increase in quality factor as a function of temperature seems to slowly approach a saturation point. One reason is that the TLS are approaching their saturation, and so increasing the temperature further will have little effect on the percentage of time that the TLS are in the excited state. As will be shown in the next section, at 400 mK the quality factor of the resonator is close to its maximum value, and will decrease as temperature is further increased.

4.4 TEMPERATURE TRACKING

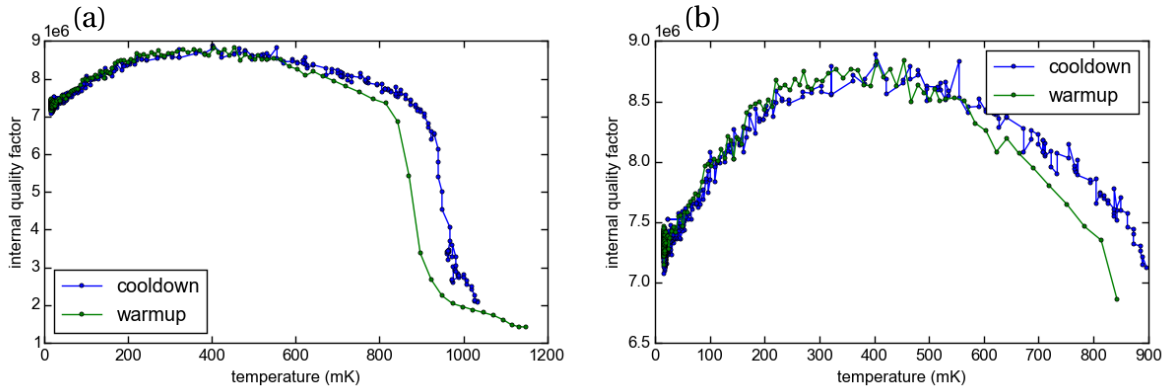


Figure 4.4: Internal quality factor versus temperature for the resonator with resonance frequency f_0 =2.75 GHz. Quality factor was continuously measured as the sample was cooled down and warmed up four days later. Panel (a) shows the full temperature range up to the helium condensation cycle. Panel (b) shows a close-up of the region until 900 mK.

To further investigate the temperature dependence of the resonator, a continous measurement was performed on the resonator with resonance frequency 2.75 GHz during a cool-down and a subsequent warm-up of the fridge four days later. Measurements were performed for temperatures ranging from base temperature (15 mK) to roughly 1 K. Above this temperature, the fridge entered a cyclic helium condensation/evaporation process. All temperatures were measured at an input power of -113 dBm, corresponding to roughly 5×10^5 photons. In Figure 4.4 the internal quality factor versus temperature is shown during a cooldown and subsequent warm-up of the fridge. As can be seen, the quality factor reaches a maximum quality factor at a temperature of ~400 mK. Below this temperature, the quality factor is likely limited by the presence of TLS (see sections 4.2 and 4.3). Above this temperature however, the quality factor decreases, indicating that TLS are not the limiting factor anymore for Q_i. One likely explanation is that the main source of dissipation is now due to the presence of quasiparticles in the resonator. At even higher powers other effects, such as vortices and enhanced radiation, contribute more and more significantly to the decay of the quality factor.

From Figure 4.4 it seems that there is some hysteresis at high temperatures. However, this is likely due to the fact that the thermometer is at a different position in the fridge as the sample, and does not thermalize equally fast. There may therefore be a delay between the temperature of the thermometer, and the actual temperature of the sample.

Aside from the internal quality factor, another quantity of interest is the resonance frequency f_0 of the resonator, which also depends on the temperature. The result from of tracking the resonance frequency of the resonator during cooldown and subsequent warm-up is shown in Figure 4.5. As can be seen in both cases, the resonance frequency reaches a maximum around 500 mK.

Between the cooldown and warm-up the resonance frequency seems to have shifted by roughly 500 Hz. As the sample was kept at 15 mK, it is unlikely that this decrease in resonance

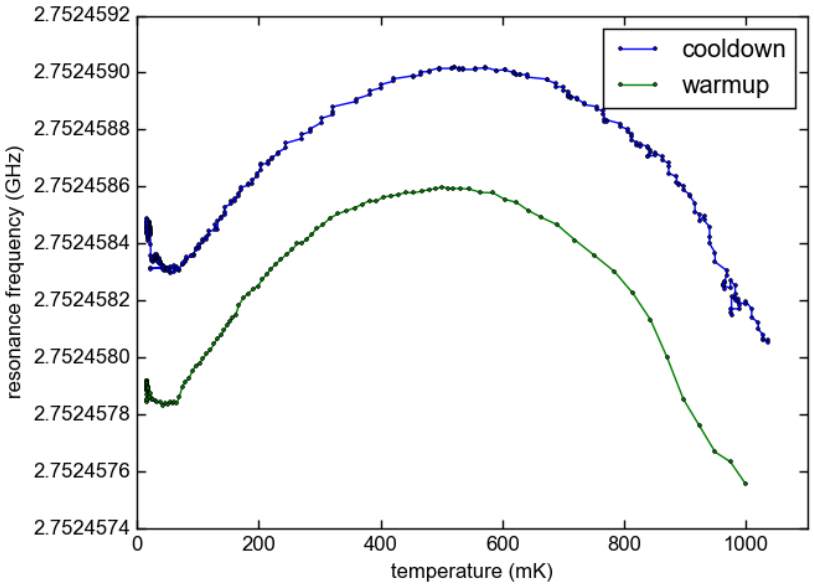


Figure 4.5: Resonance frequency versus temperature during a cooldown and subsequent warm-up four days later. In the period between cooldown and warm-up the resonance frequency has shifted by ~500 Hz, possibly due to phase noise.

frequency is due to degradation of the sample. One possible explanation is that this change in resonance frequency is due to phase noise, which is known to shift the resonance frequency of the resonator. Further measurements are, however, required to determine if this is the case.

The decrease in resonance frequency at higher temperatures can be explained by the presence of quasiparticles, which increase the kinetic inductance [11, p91]. The resonance frequency is inversely proportional to the square root of the total conductance [1], and so an increase in kinetic inductance leads to a decrease in resonance frequency. For measurements done by Barends et al. [1], the change in resonance frequency due to changes in the kinetic inductance seem to roughly correspond with the decrease in center frequency measured in Figure 4.5.

The decrease in resonance frequency at lower temperatures can be explained due to TLS still being present. A model is presented by Gao et al. [10], in which they describe the decrease in resonance frequency due to the presence of TLS. As can be seen in Figure 4.6 the model corresponds well with the data at low temperatures. At higher temperatures the model deviates from data, which may be explained by quasiparticles dominating as source of dissipation. An interesting thing to note is that an increase in resonance frequency was predicted at the lowest temperatures, but as they did not reach temperatures sufficiently low they could not confirm this effect. In Figure 4.6 however, this increase in resonance frequency is observed. This supports the claim that at low temperatures the resonator is still limited by TLS, even after treatment of HMDS and DRIE.

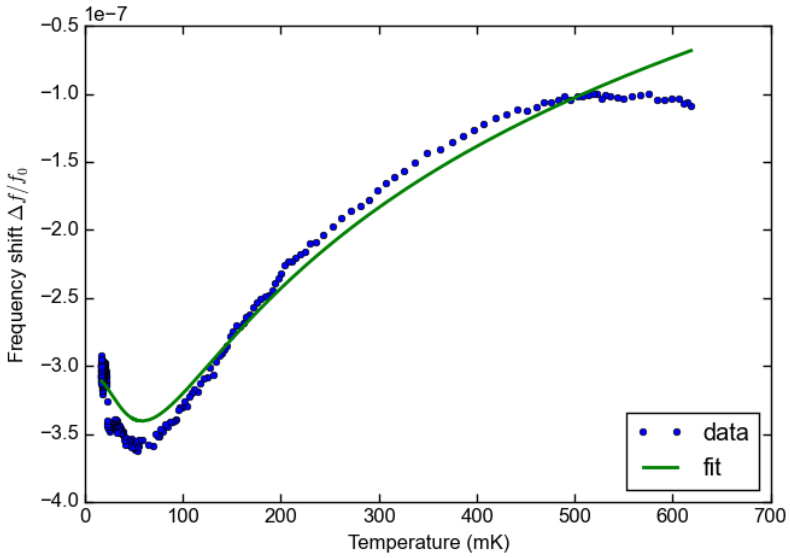


Figure 4.6: Frequency shift versus temperature for low temperatures, along with fit (green). The fit was performed using the model by Gao et al. [10]. The fit corresponds well with data, even describing the frequency peak at the lowest temperatures.

Chapter 5

Conclusion and future work

Since the quality factor of all resonators is found to decrease with decreasing input power, this indicates that at low temperatures and powers the limiting factor is still due to TLS. The fact that the quality factor initially increases with higher temperatures, supports this claim. By also measuring the resonance frequency as a function of temperature, the curve obtained is in good agreement with a model describing the resonance frequency shift due to TLS [10]. The curve even shows an increase in resonance frequency at the lowest temperatures, which was also predicted by the model. These results suggest that even after HMDS surface treatment and deep-reactive ion etching was applied, the internal quality factor of the resonator at low temperatures and power is still limited by the presence of TLS.

Nevertheless, As shown by Bruno et al. [5], the application of HMDS surface treatment and DRIE resulted in an improvement of the internal quality factor of the resonators by almost an order of magnitude. More research must be done to determine at what interface the dissipation due to TLS is greatest after these two treatments.

The next step is to perform the same treatments (HMDS and DRIE) on transmon qubits, to study what the influence will be on coherence times.

Part II

Muxmon experiment

Chapter 6

Challenges in scaling up

6.1 Frequency re-use

Even in large-scale quantum circuits individual qubit control is required. When multiple qubits are connected to a single drive line, being able to control the qubits individually requires their frequencies to be sufficiently separated. The amount by which the frequencies must be separated depends on the length of the pulses applied to the qubit. The shorter the pulse, the larger its corresponding frequency bandwidth, as can be seen in Figure 6.1. Longer pulses, on the other hand, result in less operations being possible within the decoherence times of the qubit. It is therefore desirable to have pulses with a short duration. This means that the separation between qubit frequencies must be large. Since there is only a finite frequency spectrum, having many qubits, each with a different frequency, results in frequency crowding.

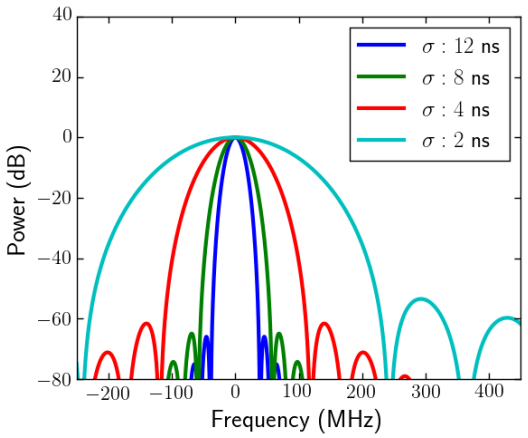


Figure 6.1: The frequency bandwidth corresponding to Gaussian pulses with different widths σ.

If the qubits are not connected to the same drive line, they can be individually controlled, even when they share the same frequency. This concept is known as frequency re-use, and is a possible way to overcome frequency crowding. Frequency re-use can be implemented in the surface code architecture. One possible implementation of frequency re-use in the surface code is shown in Figure 6.2, where every resonator is connected to four qubits. In this set-up four distinct frequencies suffice to enable individual driving of every qubit on the lattice.

Frequency re-use does pose potential issues. Two unwanted effects in particular arise when multiple qubits that are directly or indirectly coupled to each other share the same frequency. The first effect is cross-coupling, where an excitation can be transferred from one qubit to the other. This will be discussed in Section 7.4.1. The second effect is cross-

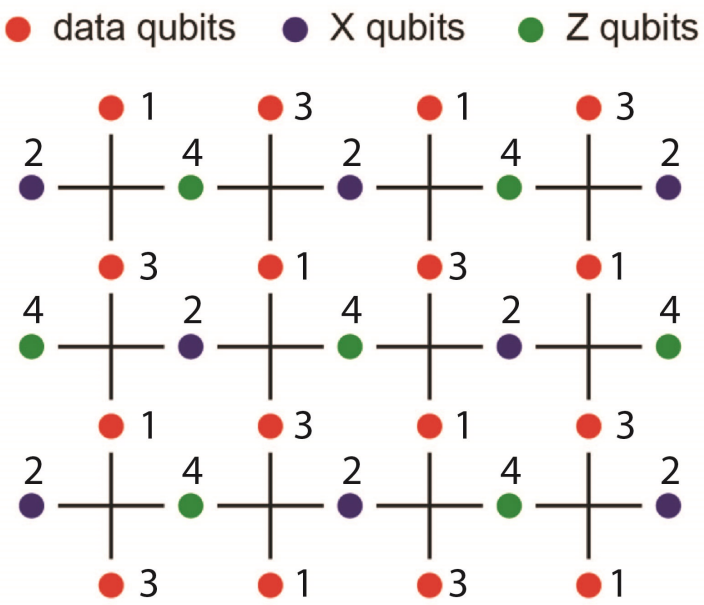


Figure 6.2: A surface code architecture with the implementation of frequency re-use. The different digits correspond to distinct frequencies.

driving, where a pulse that partially leaks through the components will drive other qubits. The components separating qubits do not act as perfect filters, and so the pulse will always partially leak through. This does not pose a problem as long as the frequency of the pulse is sufficiently detuned from the frequency of the other qubits. In the case of frequency re-use, however, the frequencies of the qubits are the same, and so a pulse leaking through will result in cross-driving. This effect is discussed in Section 7.4.2.

6.2 SELECTIVE BROADCASTING

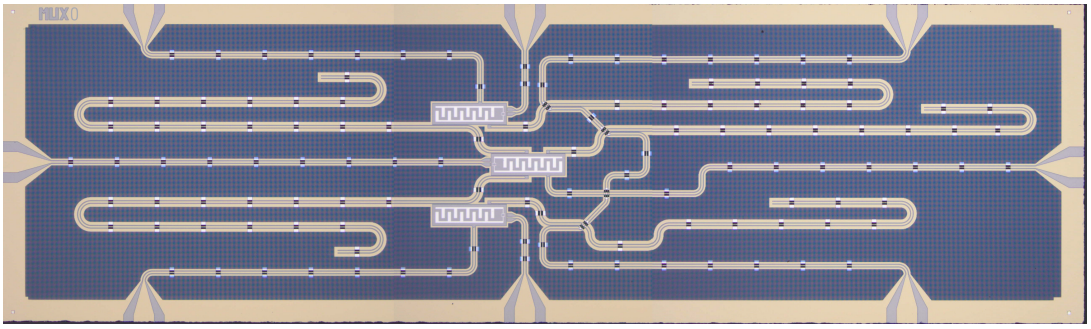
A second challenge that exists when scaling up is that as the number of qubits grow, so do the instruments needed to control them. When thinking of a large-scale quantum computer, it is unrealistic to think that each qubit will have its own RF generator, AWG, and other necessary instruments. Therefore, alternatives have to be devised to combat this scaling of instruments. One such device is the Duplexer, shown in Figure 6.3. The Duplexer is a patented vector switch matrix designed by Duije Deurloo, who is working at TNO. It has four input ports and two ouput ports. Signals going through each of the input-output port combinations pass through the following successive components:

Figure 6.3: The Duplexer

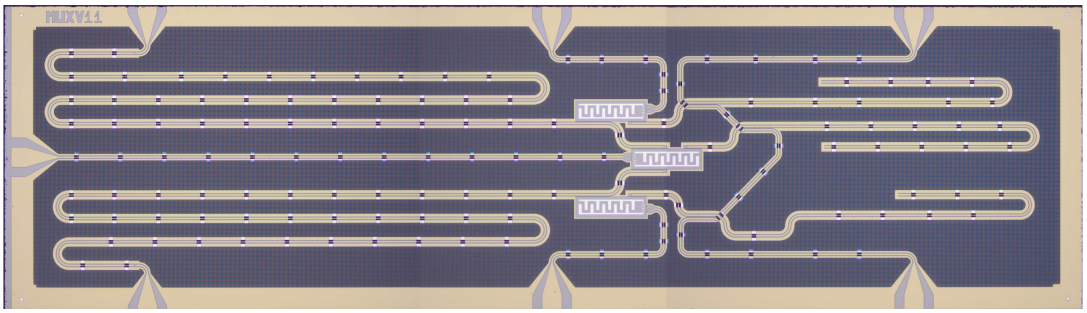
1. Digital switch
2. Variable phase-shifter
3. Variable attenuator
4. Amplifier

The Duplexer's digital switches have an RF-switching time of 4 ns. Using these switches, pulses may be sent to either of the two output ports, or to both output ports simultaneously. This allows for selective broadcasting, where pulses are routed to either or both of the two output ports at the nanosecond scale. When the output ports are connected to drive lines that are connected to qubits sharing the same frequency, the result is that two qubits can be controlled using a single generator and AWG. This property will be exploited in Randomized Benchmarking in Chapter 9.

6.3 THE MUXMON EXPERIMENT



(a) The Muxmon0 chip, in which the qubit drive lines are directly coupled



(b) The Muxmon1 chip, in which the qubit drive lines are capacitively coupled

Figure 6.4: The Muxmon0 and Muxmon1 chips. The qubits of Muxmon0 have direct drive lines, while the qubits of Muxmon1 have drive lines capacitively coupled to the resonator buses.

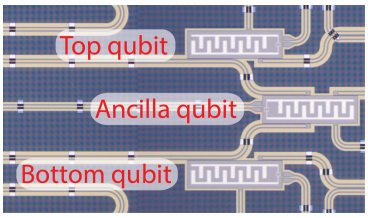


Figure 6.5: Naming of the three qubits.

The Muxmon experiment was performed specifically to study frequency re-use and selective broadcasting using the Duplexer. Two chips were designed, Muxmon0 and Muxmon1, shown in Figure 6.4. Both chips have three transmon qubits connected to them, all three of which are flux-tunable. For convenience, these qubits are named top, ancilla, and bottom qubit, as shown in Figure **??**. Air-bridges are used, not only to connect the ground-planes in the coplanar waveguides, but also such that the feedline can cross over other lines. The key difference between the two chips is the approach used to drive the qubits. In Muxmon0 all three qubits have their individual directly coupled drive lines. The top and bottom qubit are connected to the ancilla qubit through a bus. In the Muxmon1 the bus and drive lines are combined into two drive lines, each of which is capacitively coupled to the ancilla qubit and one of the other two qubits. The disadvantage of the direct drive lines in Muxmon0 is that they are an added decoherence channel for the qubits. However, the three drive lines of Muxmon0 ensure that individual qubit control of all three qubits is possible, even when they share the same frequency. This is in contrast to Muxmon1, where the ancilla qubit frequency must differ from that of the other two qubits. In the frequency re-use structure of the surface code, as explained in section 6.1, this should not be a problem, as frequency re-use in not applied to neighbouring qubits. Having capacitively coupled drive lines would result in less lines being required, both on-chip and outside the chip. Furthermore there would be one decoherence channel less. Therefore the capacitive coupled drive lines of Muxmon1 seemed the more attractive of the two options.

During initial characterization of the qubits on both chips it was found that the decoherence times of the qubits in Muxmon1 were considerably worse than those in Muxmon0 (see Appendix C.2. It is, however, unlikely that this is due to the capacitively coupled drive lines, since one would expect the absence of direct drive lines to result in one less decoherence channel. Instead it is likely that the lower decoherence times are simply due to some error in the chip fabrication process. Due to the lower decoherence times of the qubits in Muxmon1, the focus of the experiment was on Muxmon0. Therefore, unless states otherwise, the Muxmon chip used in the experiment refers to the Muxmon0 chip.

The Muxmon experiment set-up is shown in Figure 6.6. The main AWG is used for generating the Gaussian and DRAG pulses, and to trigger other devices. A second AWG is used for pulses at the excited-state to second excited-state transition frequency. Two of the four input ports of the Duplexer are used. One input port receives the main pulse, the other receives the DRAG pulse (see Section 7.3.1 for more info). The main pulse and DRAG pulse are combined in the Duplexer. The output ports are connected to the drive lines of the top and bottom qubit.

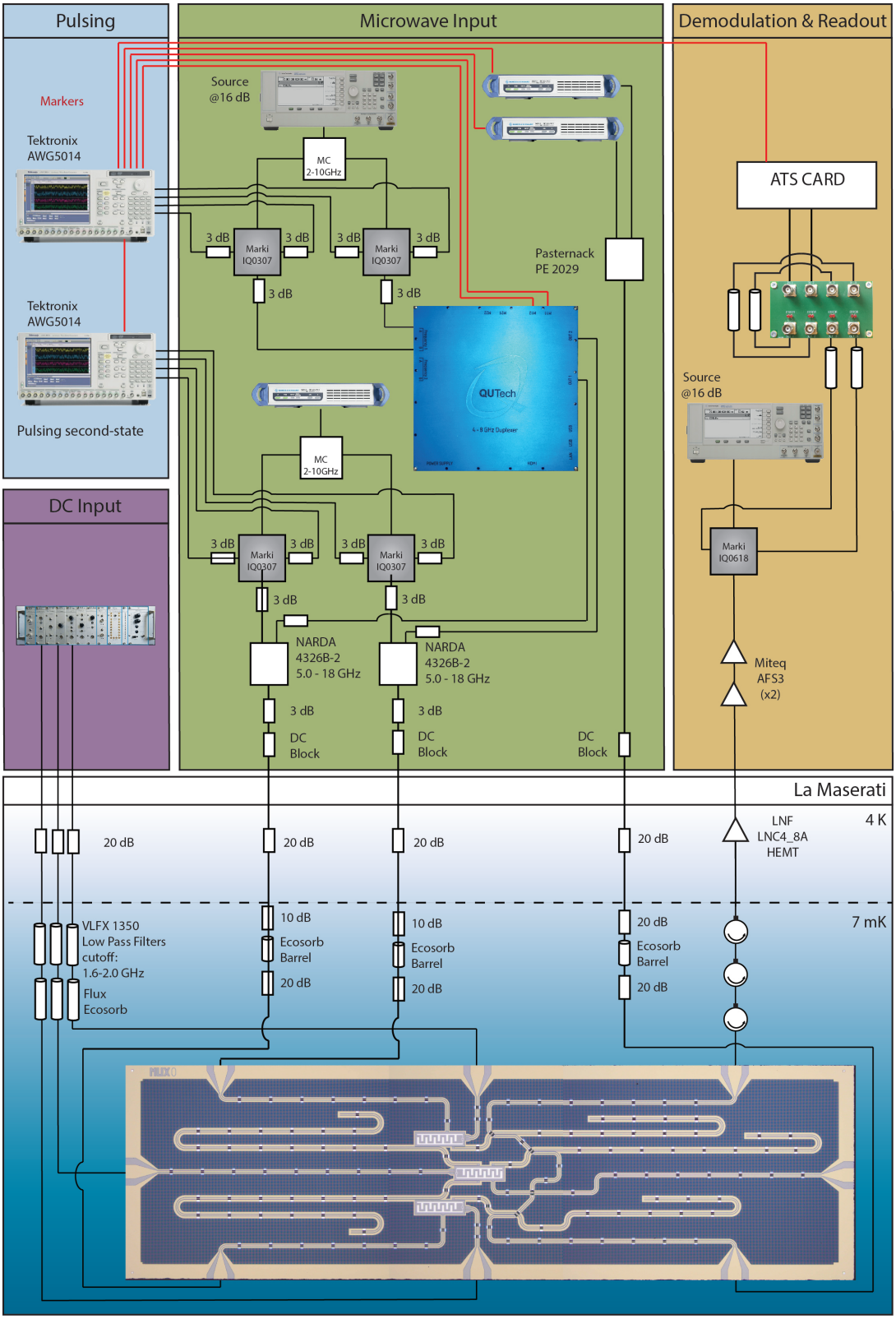


Figure 6.6: The Muxmon set-up

Chapter 7

Muxmon device characterization

This chapter describes the measurements performed to characterize the Muxmon chip. Section 7.1 gives a short introduction of the transmon qubit. The initial characterization of the chip is done using continuous-wave measurements, described in Section 7.2. These are measurements used to determine properties such as the energy levels of the resonators and qubits. Once the energy levels of the resonators and qubits are found, the specific properties of the qubits, such as their decoherence times, can be measured. These measurements are known as time-domain measurements, and are described in Section7.3. With the qubits characterized, in Section 7.4 frequency re-use is studied by tuning the top and bottom qubit to the same frequency.

7.1 THE TRANSMON

The qubits on the Muxmon chip are all transmon qubits, which are modified versions of the Cooper-pair box qubit [4, 21]. The Cooper-pair box qubit consists of a superconducting island separated from a superconducting bulk by a Josephson junction. The Josephson junction allows Cooper-pairs to travel between the island and the bulk. The Hamiltonian describing the Cooper-pair box is given by:

$$H = 4E_C(n - n_g)^2 - E_J \cos\phi \tag{7.1}$$

Here E_C is charging energy, E_J is the Josephson energy, n is the operator corresponding to the number of Cooper pairs on the island, n_g is a charge offset, and ϕ is the Josephson phase operator. The Cooper-pair box operates in the regime where $E_C \gg E_J$, resulting in well-defined number of Cooper pairs on the island, which determine its energy levels. The Cooper-pair box is therefore known as a charge qubit. The Josephson junction is a nonlinear inductor, resulting in an anharmonicity between the energy levels of the Cooper-pair box. Due to tihs anhamonicity the different energy levels can be individually addressed, which is crucial for a qubit. The qubit states in a Cooper-pair box qubit are the two states having the lowest energy $|n\rangle$ and $|n+1\rangle$. The number of Cooper pairs n having the lowest energy is determined by an applied gate voltage, corresponding to n_g.

The Cooper-pair box qubit suffers from being sensitive to fluctuations in the charge offset n_g, which result in fluctuations in its energy levels. The transmon qubit is a modification to the Cooper-pair box, which is insensitive to charge fluctuations by operating the regime where $E_J \gg E_C$ [14, 29]. This is achieved by adding a large shunting capacitor to the Cooper-pair box, which increases its capacitance to ground, thereby reducing E_C. The charge sensitivity decreases exponentially with increasing E_J/E_C ratio. The cost of increasing E_J/E_C is that the anharmonicity between the energy levels decreases, albeit with a polynomial depence. The transmon usually operates in the regime where E_J/E_C is somewhere between 20 and 100, where the charge sensitivity is largely suppressed, while still having sufficient anharmonicity. In the transmon regime the transition frequency from the ground-state to the excited-state of the qubit is given to good approximation by [26, p.52]:

$$\hbar\omega_q = \sqrt{8E_J E_C} - E_C \tag{7.2}$$

The performance of a qubit is determined by its ability to retain information, which are quantified by its decoherence times. The more isolated a qubit is from its environment, the less decay channels it will have, and the longer its decoherence times will be. On the other hand it is also necessary to interact with the qubit. In a cQED chip a signal is sent through a feedline. Directly connecting a qubit to this feedline would result in a strong decay channel. The connection between a transmon and the feedline is therefore mediated by a capacitive coupling to a resonator, in this case a coplanar waveguide (see Part I). The resonator acts as a filter, thereby limiting the decay channel. Nevertheless due to the Purcell effect the resonator is still a decay channel for the qubit, the rate of which depends on the quality factor of the resonator.

7.2 CONTINUOUS-WAVE MEASUREMENTS

The first step in the characterization of a chip is to look for signs of life, which are the energy levels of the resonators and qubits. These manifest themselves as resonance frequencies of the resonators and of the qubits that are coupled to them. To determine the energy levels we send continuous tones through the feedline, and measure response in the transmission S_{21}. These measurements are known as continuous-wave measurements.

7.2.1 Scanning for resonators

Since communication with the qubits is mediated through their capacitive coupling to resonators, the first step is to find the resonator frequencies. This is done using a transmission measurement of the feedline, in combination with heterodyne detection, and has been explained in section **TODO:** Create section in Resonator chapter.

There is one difference in measuring a resonator when there is a qubit coupled to it. When considering the qubit as a two-level system, the behaviour of the coupled resonator-qubit system is governed by the Jaynes-Cummings Hamiltonian [14]:

$$\hat{H} = \hbar\omega_{\mathrm{r}}\left(\hat{a}^{\dagger}\hat{a} + \frac{1}{2}\right) + \frac{\hbar\omega_{\mathrm{q}}}{2}\hat{\sigma}_z + \hbar g\left(\hat{a}^{\dagger}\sigma_{-} + \hat{a}\sigma_{+}\right) \tag{7.3}$$

where ω_{r} is the bare resonance frequency of the resonator, ω_{q} is the resonance frequency of the qubit's ground-to-excited-state transition, and the qubit's two states are in the spin-representation. This Hamiltonian consists of three terms. The first term corresponds to the energy level of the resonator, the second to the energy level of the transmon, and the third is an interaction term between the two with coupling strength g. The coupling strength g determines the rate at which the qubit and resonator exchange excitation. If g is larger than the decay rates of both the resonator and qubit, the system is in the strong-coupling regime, and so an excitation can travel multiple times between the resonator and qubit before decaying.

The difference between the resonator frequency ω_{r} and the qubit frequency ω_{q} is given by the detuning $\Delta = \omega_{\mathrm{q}} - \omega_{\mathrm{r}}$. If the magnitude of the detuning is large compared to the coupling strength g, the system is in the dispersive regime. In this case the Hamiltonian can be approximated by the dispersive Jaynes-Cummings Hamiltonian:

$$\hat{H} = \frac{\hbar\omega_{\mathrm{q}}^{'}}{2}\hat{\sigma}_z + \left(\hbar\omega_{\mathrm{r}}^{'} + \hbar\chi\hat{\sigma}_z\right)\hat{a}^{\dagger}\hat{a} \tag{7.4}$$

The coupling between the qubit and resonator causes both the qubit frequency and the resonator frequency to shift: $\omega_{\mathrm{q}}^{'} = \omega_{\mathrm{q}} + \chi_{01}$, $\omega_{\mathrm{r}}^{'} = \omega_{\mathrm{r}} - \chi_{12}/2$, where $\chi_{ij} = \frac{g_{ij}^2}{\omega_{ij} - \omega_r}$ are the partial dispersive shifts. We see that in this approximation we must not only take into account the lowest two states of the transmon, but also the second excited-state of the transmon. The difference between the first excited-state to second excited-state transition frequency ω_{12} and the ground-state to first excited-state transition frequency ω_{01} is given by the anharmonicity $\alpha = \omega_{12} - \omega_{01} \simeq E_C$ [14]. The anharmonicity α determines the degree in which the transmon behaves as a qubit, without experiences excitations to higher excited-states.

Aside from experiencing a frequency shift dependent on the amount of detuning, Equation 7.4 shows that the resonator also experiences a shift depending on the state of the qubit. The resonator's frequency is decreased by an amount 2χ when the qubit is in the excited-state. The parameter χ is the dispersive shift, and is given by:

$$\chi = \chi_{01} - \chi_{12}/2 \approx \frac{g^2}{\Delta}\frac{E_{\mathrm{c}}}{\hbar\Delta - E_{\mathrm{c}}} \tag{7.5}$$

Due to this coupling between resonator and qubit, it is important to choose the right RF power. When the amount of photons in the resonator reaches a critical photon number, this coupling will result in the resonator experiencing nonlinear effects. The resonator will thereby lose its Lorentzian lineshape. Therefore the RF power should be kept sufficiently low to avoid these nonlinear effects, while still maintaining a good signal-to-noise ratio.

In Figure 7.1 the three resonators belonging to the top, ancilla, and bottom qubit are shown, with resonator frequencies 6.700 GHz, 6.733 GHz, and 6.800 GHz respectively.

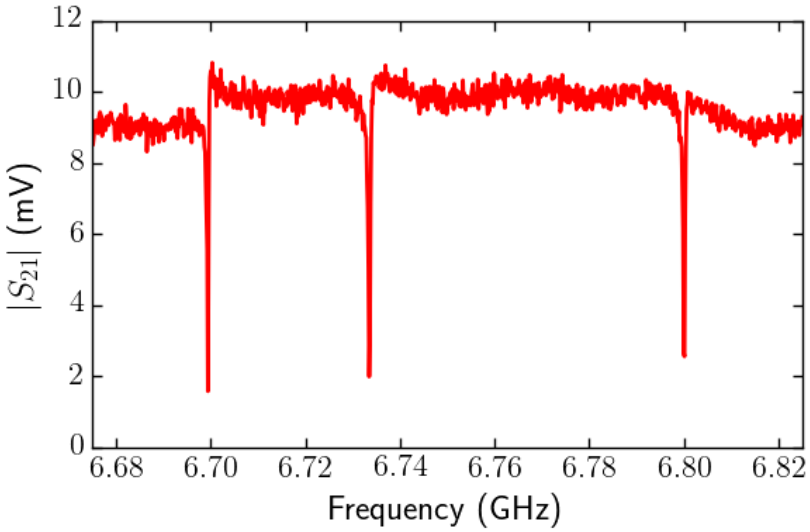


Figure 7.1: The three resonators coupled to the top, ancilla, and bottom qubits in ascending frequency.

7.2.2 Powersweeping the resonators

Once the resonators have been located, the next stage is to find the qubit that is capacitively coupled to each of the resonators. First some initial measurements are performed aimed at gaining information about our resonator and qubit, which will allow us to search for the qubit frequencies with more focus.

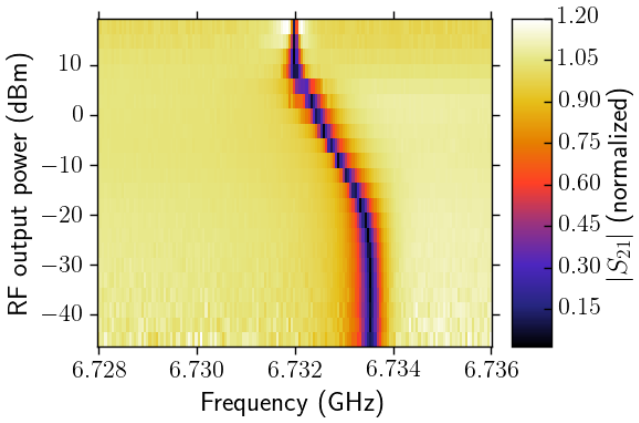


Figure 7.2: A powersweep of the resonator coupled to the ancilla qubit. The dispersive shift found is equal to $\chi \approx 30$ MHz.

As explained previously, the capacitive coupling between the resonator and qubit shifts the resonator frequency ω_r from its bare frequency. When the amount of photons in the resonator reaches a certain point, the resonator experiences nonlinearity, thereby losing its Lorentzian lineshape. When increasing the RF power even further, at a certain point the resonator regains its Lorentzian lineshape. In doing so its resonance frequency has shifted to its bare frequency ω_{bare}. For more information see Reed's thesis [26]. Observing this frequency shift reveals that a qubit is coupled to the resonator. This frequency shift is commonly measured in a powersweep, where a resonator scan is performed for a range of powers. In Figure 7.2 a powersweep is shown of the ancilla qubit. As can be seen the resonator frequency shifts during the transition from the low-power regime to the high-power regime. Since the resonator frequency shifts downwards when entering the high-power regime it can be concluded that the ancilla qubit frequency lies below the resonator frequency.

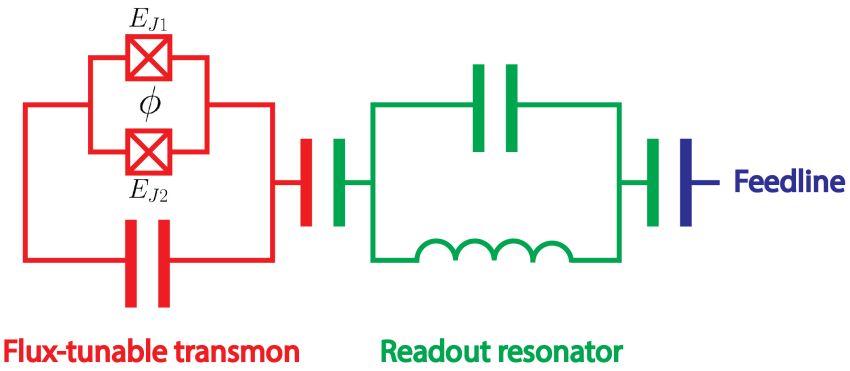


Figure 7.3: Schematic of a flux-tunable transmon coupled to a readout resonator. The two Josephson junctions with Josephson energies E_{J1} and E_{J2} form a SQUID loop, resulting in an effective Josephson energy E_J^{eff} that depends on the flux ϕ through the SQUID loop.

A powersweep additionally provides information about at what power the resonator enters the nonlinear regime. For measurements involving the qubit the readout power must be below this threshold power. Furthermore, from the frequency shift between the dressed cavity frequency and the bare cavity frequency, the amount of detuning between the qubit and the resonator can be estimated using Equation 7.5.

If no shift is observed, it could mean that the qubit is dead, due to an open or shorted Josephson junction. However, this is not necessarily the case. An alternative possibility is that the detuning between qubit and resonator is very large, and as a result the frequency shift cannot be discerned.

7.2.3 Scan for qubit sweet-spots

All qubits in the Muxmon device have a tunable resonance frequency. This is achieved by replacing the Josephson junction with a superconducting quantum interference device (SQUID), as shown in Figure 7.3. Here the two islands that compose the transmon qubit are connected by two Josephson junctions instead of one, effectively forming a loop. If the two Josephson junctions in a SQUID loop have the same Josephson energy $E_J = E_{J1} = E_{J2}$, this results in an effective Josephson energy [26, pp.54-56]:

$$E_J^{\text{eff}} = 2E_J * \cos\left(\pi \frac{\Phi}{\Phi_0}\right) \tag{7.6}$$

where $\Phi_0 = h/2e$ is the magnetic flux quantum and Φ is the flux passing through the SQUID loop.

As can be seen in Equation 7.6, the effective Josephson energy E_J^{eff} of a SQUID loop can be controlled by the flux through the SQUID loop. This is commonly done by having a flux bias line in close proximity to the SQUID loop. Current flowing through the flux-bias line alters the magnetic field in the vicinity of the SQUID loop, thereby changing the amount of flux through the SQUID loop. A fixed current from a digital-to-analog (DAC) converter is used to set the flux through the SQUID loop, thereby changing the effective Josephson energy

E_J^{eff}. According to Equation 7.2 the qubit frequency depends on the Josephson energy, and so changing the magnetic flux through the SQUID loop changes the qubit frequency. Qubits having a SQUID loop are therefore called flux-tunable. Since the effective Josephson energy E_J^{eff} has a cosine-dependence on the flux, the qubit frequency is periodic with respect to the flux. It- has a maximum when Φ is a multiple of Φ_0, in which case the qubit is said to be at its sweet-spot.

For flux-tunable qubits, flux corresponding to the sweet-spot of the qubit can be found without knowledge of the qubit frequency. This can be done by sweeping the DAC current and measuring the shift in the resonator frequency. Because the qubit frequency varies as the current through the flux-bias line changes, the detuning between the qubit and the resonator changes. As a result the dispersive shift χ, and therefore the resonator frequency, also varies. At the sweet-spot of the qubit, the resonator's frequency ω_r is at a maximum. This is irrespective of whether the qubit's frequency ω_q is above or below the resonator's frequency.

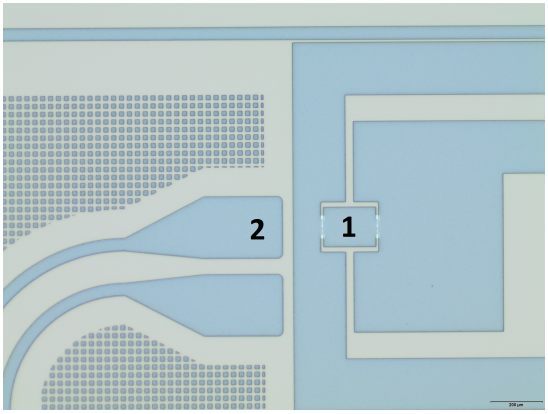


Figure 7.4: A SQUID loop (1) in close proximity to a flux-bias line (2).

Finding the sweet-spot can be done by performing a series of resonator scans as the DAC voltage is varied. The result of such a 2D scan shown in Figure 7.7. There is an alternative, faster measurement which can be performed, and is explained in Appendix D.1.1. It is not necessarily the case that the qubit sweet-spot is at zero DAC current, as trapped magnetic flux may result in a flux offset.

In the case where the powersweep showed no measurable frequency shift, these measurements are also useful in discerning whether or not the qubit actually has broken junctions, or whether it was simply far detuned from the resonator.

7.2.4 Qubit spectroscopy

The measurement to perform in order to find the qubit depends on the amount of detuning between the resonator and qubit, which can be estimated from powersweep measurements. If the amount of detuning is large compared to the coupling strength ($\Delta \gg g$), the system is in the dispersive regime. In this case one commonly performs a two-tone spectroscopy to find the qubit's frequency. If the detuning is comparable to the coupling strength ($\Delta \sim g$), the qubit and resonator are hybridized, and experience an avoided crossing. In such cases a normal transmission measurement suffices. In the Muxmon device the sweet-spot frequencies of all qubits were considerably lower than their corresponding resonator frequencies, and so the qubits were always in the dispersive regime. Therefore two-tone spectroscopy was performed to find the qubit frequencies.

When finding the qubits the DAC currents were chosen such that the qubits were close to their sweet-spots. It is important that the qubit is not close to its anti-sweet-spot, which

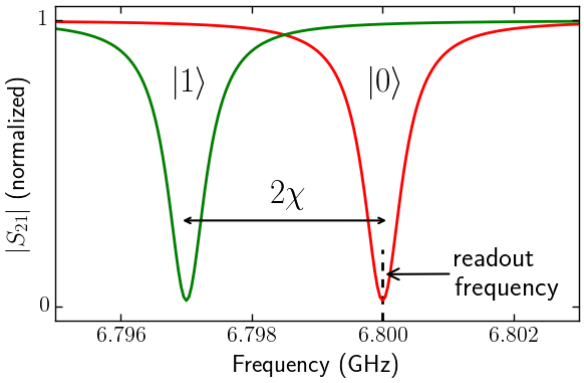


Figure 7.5: The resonator frequency shifts by an amount 2χ dependent on the state of the qubit, where χ is the dispersive shift. In a two-tone spectroscopy measurement the readout frequency is fixed at the resonant frequency of the top qubit, while a second tone is swept with varying frequency. The transmission is low when the second tone is off-resonant with the qubit frequency and high if the second tone is resonant with the qubit frequency.

would make it difficult, if not impossible to find.

As was explained in section 7.2.1, in the dispersive regime the resonator experiences a 2χ frequency shift dependent on the state of the qubit. This is shown in Figure 7.5. For a resonator capacitively coupled to the feedline, the transmission experiences a dip at the resonator frequency, which shifts when the qubit's state is switched. The transmission at the resonator's dip when the qubit is in the ground state will therefore be dependent on the state of the qubit (low when the qubit is in the ground state, high when the qubit is in the excited-state). This is the property exploited in a two-tone spectroscopy measurement.

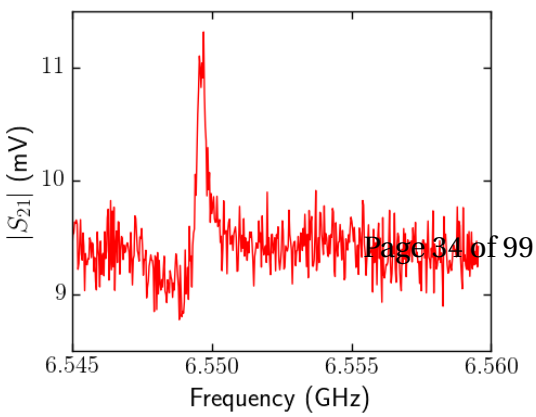
In a two-tone spectroscopy measurement two tones are sent through the feedline.

1. A drive tone with varying frequency ω_d.

2. A readout tone at the resonator frequency when the qubit is in the ground state.

In figure 7.6 the result of a two-tone spectroscopy is shown. When the drive frequency ω_d is detuned from the qubit's frequency ω_q, the drive is off-resonant with respect to the qubit, and so we measure a low transmission due to the readout tone being at the resonator frequency. However, when the drive frequency ω_d approaches the qubit's frequency ω_q the qubit will start to oscillate between its ground- and excited-state, at a rate dependent on the detuning between the drive frequency ω_d and the qubit's frequency ω_q. The qubit will therefore have a partial population in the excited-state, resulting in a shift of the resonator frequency, dependent on the population in the excited-state. The result is an increase in transmission

The minimum linewidth of the qubit using spectroscopy is set by its dephasing time T_2, which can be seen as the uncertainty in its frequency. However, the power of the drive tone causes an additional increase in the linewidth,



due to stimulated emission of the qubit. This effect is known as power broadening, and can be quite useful for finding qubits, especially when designing high-quality qubits with a very narrow intrinsic linewidth. The optimal power for the drive strength is further dependent on the amount of detuning Δ between the qubit and resonator. The resonator effectively acts as a bandpass filter, centered around the resonator frequency. Therefore, the stronger the detuning, the more the drive tone is suppressed.

The dispersive shift 2χ is also dependent on the amount of detuning Δ between the qubit and resonator. If the detuning Δ is large, the dispersive shift will be very small, and so the difference in transmission will also decrease. Since the resonator has a Lorentzian lineshape, at the resonance frequency this Lorentzian is flat, and so is insensitive to small deviations. To increase the contrast between the resonant and off-resonant transmission, it is usually advantageous to measure at a slight detuning δ away from the resonance frequency, where the transmission slope is high. Since the resonator frequency shifts down when the qubit is excited, it is better to have a positive detuning δ to ensure that the transmission increases as the drive frequence ω_d approaches the qubit frequency ω_q and decreases as it leaves the qubit frequency.

A better approach to a two-tone spectroscopy measurement is to separate the drive tone and the readout tone in time. This is known as pulsed spectroscopy, and has the advantage that during readout the photon population in the resonator will be low, resulting in a more accurate measurement. This does, however, require a more complicated set-up, where the pulses need to be accurately timed.

7.2.5 Tracking the qubits

For flux-tunable qubits one is usually not interested in finding the frequency at one specific flux value, but how this frequency changes with varying flux. A simple approach is to perform a two-dimensional scan of a fixed frequency range versus flux. This however has several disadvantages. First of all it requires a new measurement to be set-up every time the qubit frequency moves out of the frequency range. Furthermore the largest part of the measurement will be spent measuring off-resonant signal, containing no useful information.

As an alternative approach I have created a

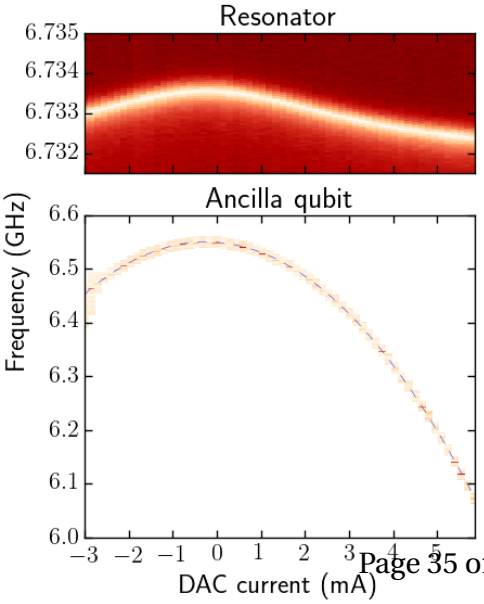


Figure 7.7: Tracked spectroscopy of the ancilla qubit. The blue dotted line is the corresponding fit to the qubit frequency, showing excellent agreement with measurements

modified version of the spectroscopy versus flux scan, known as tracked spectroscopy. In this approach after every spectroscopy scan the qubit frequency is extracted through fitting. From the qubit frequencies measured in previous scans the expected frequency at the next flux value is extrapolated. The frequency window of the next spectroscopy scan are then centered around the expected qubit frequency. The qubit is therefore tracked as its frequency changes, without requiring any human intervention. The same method is applied to update the resonator frequency. For more information on the tracked spectroscopy algorithm see appendix **??**.

In Figure 7.7 the results of tracked spectroscopy are shown for the ancilla frequency. As can be seen the qubit is tracked over a wide range of frequencies, while the frequency window in each scan is relatively small. As the qubit frequency approaches the resonator frequency, the dispersive shift χ increases, and so the resonator frequency shifts upwards. InIn this scan one can clearly see the sweet-spot of the qubit, and the curve has the shape of the square root of a cosine, consistent with Equations 7.2 and 7.6.

7.2.6 Second transition of the qubit

Once the qubit's grequency is known, it is possible to find the qubit's excited-state to second-excited-state transition ω_q^{12}, which shall be referred to as the 12-transition. For consistent notation we shall temporarily denote the qubit's ground-state to excited-state transition frequency as ω_q^{01}. The anharmonicity α is then given by the difference in transition frequencies: $\alpha = \omega_q^{12} - \omega_q^{01}$. Knowledge of the anharmonicity α allows determination of the coupling energy E_c through **TODO:**.

The 12-transition can be found using three-tone spectroscopy, in a manner similar to two-tone spectropscopy measurement explained in section 7.2.4. The following three tones are used in three-tone-spectroscopy:

1. A drive tone with fixed frequency $\omega_{\text{drive}}^{01}$ at the qubit frequency ω_q^{01}.
2. A drive tone with varying frequency $\omega_{\text{drive}}^{12}$ to scan for the 12-transition.
3. A third tone at the resonator frequency when the qubit is in the excited-state.

The first drive tone with frequency $\omega_{\text{drive}}^{01}$ is used the drive the qubit to the excited-state. Since the first drive tone results in the excited-state being (partially) populated, the second tone with frequency $\omega_{\text{drive}}^{12}$ is then able to drive the qubit from the excited-state to the second-excited-state. The transmission should therefore change when $\omega_{\text{drive}}^{12}$ is on resonance with ω_q^{12}.

In principle the first drive tone with frequency $\omega_{\text{drive}}^{01}$ can be kept fixed at the 01-transition frequency ω_q^{01}. However, by performing a two-dimensional scan, where $\omega_{\text{drive}}^{01}$ is also varied in a small region around ω_q^{01}, the 12-transition becomes much clearer. This can be seen in

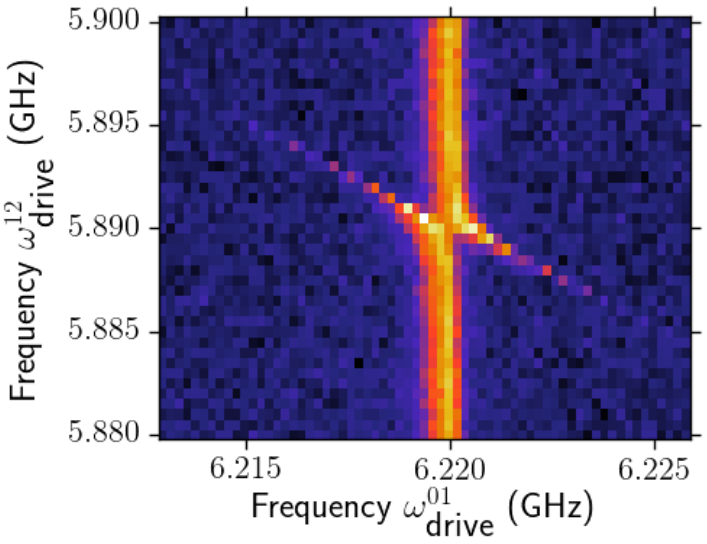


Figure 7.8: Two-dimensional three-tone pulsed spectroscopy results for the bottom qubit while at its sweet-spot. The second-excited state has a frequency equal to $\omega_{12} = 5.89\,\text{GHz}$, corresponding to an anharmonicity $\alpha = \omega_{12} - \omega_{01} = -330\,\text{MHz}$.

Figure 7.8, where a shift in transmission is observed when $\omega^{01}_{\text{drive}} = \omega^{01}_q$. This corresponds to the 12-transmission frequency ω^{12}_q. Additionally, a transmission shift is observed at a line crossing ω^{12}_q. At this line $\omega^{01}_q + \omega^{12}_q = \omega^{02}_q$, resulting in some population in the second-excited-state.

It is worthwhile to note that three-tone spectroscopy produces much more accurate results when using pulsed spectroscopy, as the difference in transmission is generally small compared to two-tone spectroscopy.

Finding the resonator frequency when the qubit is in the excited-state can be tricky, as pulsed spectroscopy is usually required. Nevertheless a rough estimation can be found using a modified version of spectroscopy, where the drive tone ω_{drive} is kept fixed, while the resonator frequency is swept. The lower peak will be a rough estimation of the resonator frequency when the qubit is in the excited-state. Alternatively it is also possible to use the resonator frequency when the qubit is in the ground-state, although this will reduce the contrast between driving on-resonance and driving off-resonance.

- Once the 12-transition transition is found, and hence the qubit's anharmonicity α, it is the possible to calculate the coupling energy E_c. **TODO:** cite Reed.

7.2.7 Flux matrix

When a current is passed through a flux-bias line, it is not only the flux through the SQUID of the qubit directly connected to it that is affected. Instead, the fluxes through SQUIDs of neighbouring qubits are also affected, albeit to a lesser extent. Therefore changing the frequency of one qubit by changing its corresponding flux-bias line current also affects the

frequencies of the other flux-tunable qubits on the chip. For the Muxmon experiment the frequencies of the three qubits have to be individually tuned to very specific frequencies, and so the frequency responses of the qubits have to be decoupled. This is done by implementing a flux matrix, which corrects for the flux cross-coupling effects.

A flux matrix $\boldsymbol{F}$ is an $n \times n$ matrix, where n is the number of flux-tunable qubits. Each row $F_i = [F_{i1}, \ldots, F_{in}]$ corresponds to the ratio's by which each of the DAC currents should be changed such that only the frequency of qubit i changes while the other frequencies remain fixed. This results in n decoupled virtual flux parameters.

For a flux matrix F knowledge of the frequency responses of each of the qubits due to each of the flux-bias lines is required. It is important that the qubits are tuned away from their sweet-spots, to a point where the frequency is sensitive to changes in flux. In Figure 7.9 the frequency responses of the three Muxmon qubits to currents flowing through each of the flux-bias lines are shown. As can be seen the degree to which a flux-bias line affects other qubits can be quite severe. From the frequency responses a frequency response matrix $\boldsymbol{M}$ can be constructed:

$$\boldsymbol{M} = \begin{bmatrix} \frac{\partial f_1}{\partial I_1} & \cdots & \frac{\partial f_1}{\partial I_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial I_1} & \cdots & \frac{\partial f_n}{\partial I_n} \end{bmatrix} \tag{7.7}$$

where $\frac{\partial f_i}{\partial I_j}$ is the frequency response of qubit i to a change in the DAC current corresponding to flux-bias line j. Each rows M_i should be divided by the frequency response of the main qubit $\frac{\partial f_i}{\partial I_i}$, such that the elements are relative frequency responses. The diagonal elements of matrix $\boldsymbol{M}$ should be equal to one.

For each qubit i the DAC current I_i^0 corresponding to its sweet-spot is known from tracked spectroscopy. However, since the other DAC currents must be set to zero, the corresponding qubits will likely not be at their sweet-spots. There is a certain combination of DAC currents $\vec{I}^{ss}$ where all qubits are at their simultaneous sweet-spot. This can be found using matrix $\boldsymbol{M}$. Since $\boldsymbol{M}$ contains the frequency responses of each qubit to each of the flux-bias lines, the qubit i remains at its sweet-spot as long as $\boldsymbol{M}_i \vec{I} = I_i^0$ holds, where $\vec{I}$ is the vector containing the DAC currents. Therefore the simultaneous sweet-spot $\vec{I}^{ss}$ is found by solving $\boldsymbol{M}\vec{I}^{ss} = \vec{I}^0$, where $\vec{I}^0$ contains the DAC currents of the individual sweet-spots of the qubits.

The flux matrix $\boldsymbol{F}$ can be found by inverting matrix $\boldsymbol{M}$. As mentioned earlier, each row F_i of matrix $\boldsymbol{F}$ corresponds to the ratio by which the DAC currents of the flux-bias lines need to be varied, such that only the frequency of qubit i is changed. Each row may therefore be multiplied by a different factor, as long as the ratio between the elements in each row remains constant. By dividing each row F_i by F_{ii} the DAC current of the main flux-bias line is equal to its corresponding virtual flux. The main qubit therefore behaves identically when changing the virtual flux, except that all the other qubit frequencies remain fixed.

A virtual flux vector $\vec{\Phi} = [\phi_1, \ldots \phi_n]$ can be converted to the corresponding DAC current vector $\vec{I}$ through:

$$\vec{I} = \boldsymbol{F}\vec{\Phi} + \vec{I}^{ss} \tag{7.8}$$

By adding the simultaneous sweet-spot DAC voltages $\vec{V}^{ss}$, we obtain the additional property that the simultaneous sweet-spot of the virtual fluxes is set to zero.

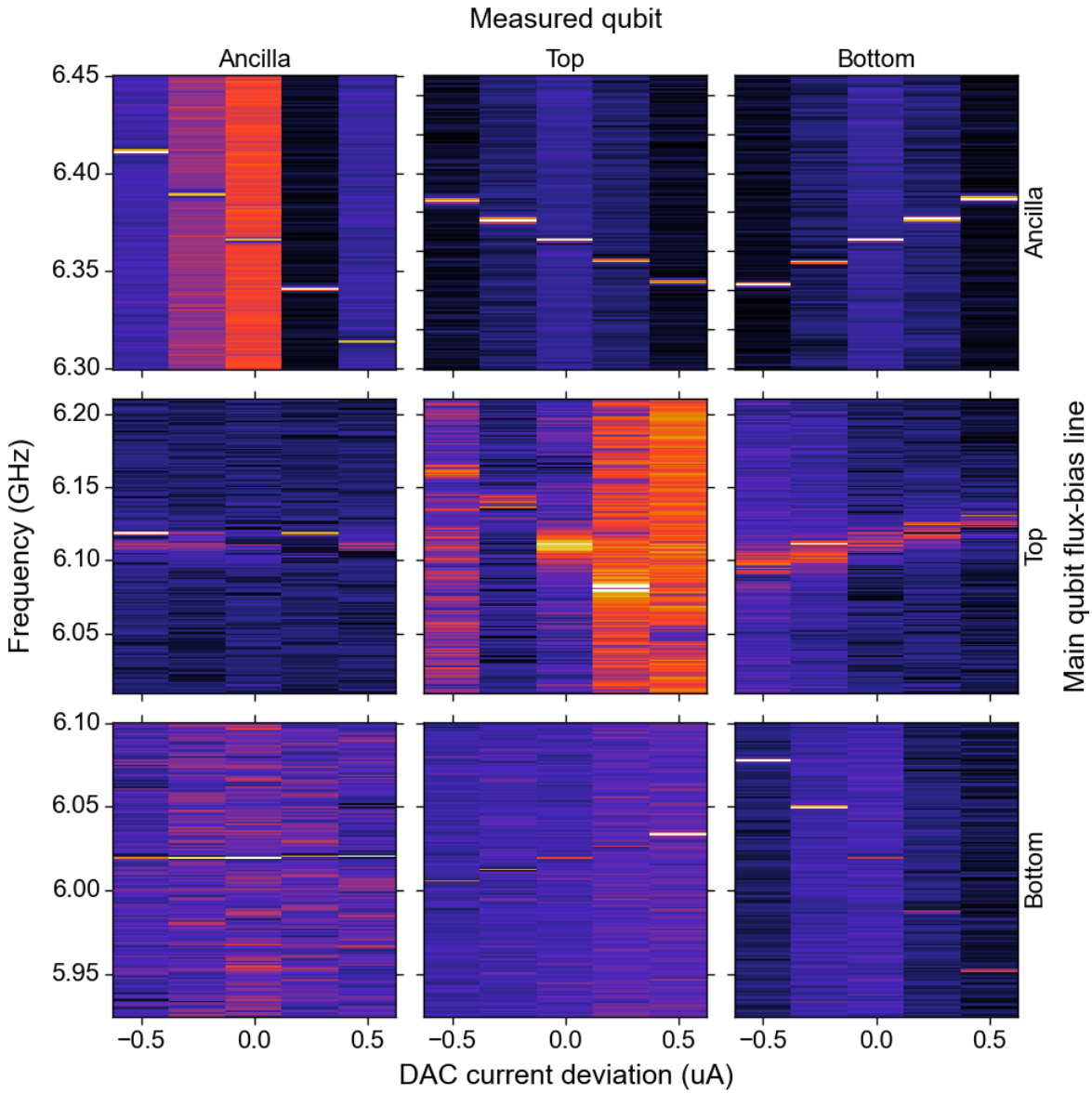


Figure 7.9: Frequency responses of each qubit to each flux-bias line. Each qubit has been tuned away from its sweet-spot such that the frequency-response is higher. The frequency response slopes determine the coefficients of matrix $\boldsymbol{M}$, used for creating the flux matrix $\boldsymbol{F}$.

7.3 TIME-DOMAIN MEASUREMENTS

7.3.1 Qubit control

The measurements described so far are able to determine the energy levels of the qubits and resonators using continuous tones. However, for qubit properties such as their decoherence times, time-domain measurements are required. Here well-calibrated pulses accurately control the state of the qubit, and correspond to gates being applied to the qubit.

A simple time-domain measurement usually consists of two parts: qubit control, and qubit readout. The qubit is initially in the ground-state due to relaxation. During qubit control precisely timed pulses modify the state of the qubit. During qubit readout a readout tone is applied in a manner similar to spectroscopy. The state of the qubit can be inferred from the response in transmission. More complicated measurements involve feedback, where additional qubit control can be applied depending on the measurement outcome. Feedback measurements are not performed in this experiment, and are therefore beyond the scope of this thesis.

Measuring the qubit will project its state onto the Z-axis, and so it will be either measured in state $|0\rangle$ or in state $|1\rangle$, with a probability determined by its population in the two states. One single measurement, however, provides little information about the actual populations in the two states before the measurement. To determine the actual population in the two states it is required to repeat the measurement many times. The number of times the qubit is measured in either states is then an estimate for the actual state populations of the qubit. It is important that the experimental repetition rate is low enough such that the qubit has sufficient time to relax to its ground-state, which is determined by its relaxation time T_1 (see Section 7.3.3.1).

Qubit pulses usually have a Gaussian shape, because the corresponding frequency bandwidth is narrow and because it has a smooth shape in the frequency domain. These pulses can be generated by modulating a carrier signal from an RF generator, and is commonly done using an Arbitrary Waveform Generator (AWG). For the Muxmon experiment the Tektronix AWG5014 is used, which has four voltage channels, each of which can control its voltage output at the nanosecond scale. It further has eight marker channels, which are used to trigger devices, such as RF generators and the Duplexer.

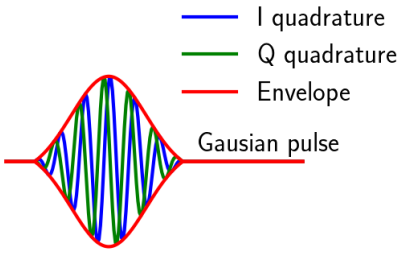


Figure 7.10: Schematic representation of the Gaussian pulse. The signal is convoluted with a sine and cosine with sideband modulation frequency ω_{sb}, resulting in single-sideband modulation.

The carrier signal is sent through the LO port of the mixer, where it is modulated by the AWG. In the Muxmon experiment IQ-modulation is used, and so two AWG channels are used to independently modulate the in-phase and quadrature components of the signal in an IQ-mixer. This allows for single sideband modulation, which shifts the frequency of the signal. Single sideband modulation has the advantage that the carrier frequency ω_c is shifted away from mixer leakage. For more information on mixer leakage, see section **??**. A schematic representation

of a Gaussian pulse with single sideband modulation is shown in Figure 7.10.

An important consideration is the duration of a pulse, also known as the pulse length. Shorter pulses allow for more qubit operations within its decoherence time. However, the pulse length is inversely proportional to its frequency bandwidth. If the frequency bandwidth is too broad, there will be a nonnegligible signal at the qubit's excited-state to second excited-state transtion frequency. This will cause leakage to the second excited-state, thereby leaving the two-state Hilbert space. It is therefore desirable to have a bandwidth, which is the inverse of the width σ of the Gaussian pulse, that is small compared to the anharmonicity of the qubit.

Performing gates on qubits requires knowledge of the parameters that define the pulse, such as the amplitude, phase, and pulse duration. The phase determines the axis along which the qubit is rotated. The duration and amplitude of the pulse determine the rotation angle. Rotating the qubit by a specific angle can be achieved by either varying the pulse duration or the pulse amplitude. Both methods have advantages and disadvantages. Varying the pulse duration ensures that the maximum amplitude is roughly fixed, regardless of the rotation angle. Varying the amplitude, on the other hand, ensures that pulses applied to multiple qubits end simultaneously, and so it is more natural to speak of a pulse clock cycle. In the measurements for this thesis the amplitude is varied, while the pulse duration is kept fixed.

7.3.2 Drive amplitude calibration

Determining the correct amplitude for qubit gates is commonly performed using a Rabi measurement. In this measurement the amplitude of a pulse with rotation along the X-axis is varied monotonically. The pulse will cause the qubit to rotate at a Rabi rate Ω_R, which is proportional to the pulse amplitude.

After application of a pulse with Rabi rate Ω_R and duration τ, the wavefunction of the qubit will be in state $|\psi\rangle = \cos\left(\frac{\Omega_R}{2}\tau\right)|0\rangle + \sin\left(\frac{\Omega_R}{2}\tau\right)|1\rangle$. During a subsequent measurement the qubit will be in the excited-state with probability $\sin^2\left(\frac{\Omega_R}{2}\tau\right)$ [26].

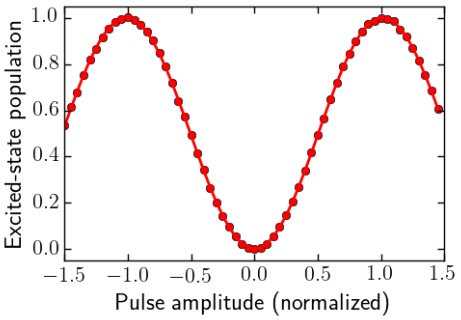


Figure 7.11: A Rabi measurement. In the center peak the pulse has zero amplitude, corresponding to the qubit being in the ground-state. The other two peaks therefore correspond to $X_{-\pi}$ and X_{π} pulses, as the qubit is in the excited-state.

In a Rabi measurement the amplitude is swept monotonically from a negative value to a positive value. The result should look like a cosine, as shown in Figure 7.11. The center of this cosine is where the amplitude is zero, and therefore corresponds to the ground-state of the qubit. At the other peaks, where the deviation from the ground-state is maximal, the qubit is in the excited-state. The amplitude of the left peak corresponds to an $X_{-\pi}$ pulse, and the amplitude of the right peak to an X_{π} pulse.

TODO:

- Mention calibration points

7.3.3 Qubit decoherence

Once the pulse amplitude has been calibrated the qubit state can be controlled. Ideally the state of a qubit remains fixed in the absence of pulses. However due to the qubit's interaction with its environment the qubit will experience decoherence. Decoherence can be viewed as the qubit experiencing random interactions with its environment, and as a result we lose information about the state of the qubit. The decoherence of the qubit can be characterized by its relaxation time T_1, and its dephasing time T_2.

7.3.3.1 Qubit relaxation: T_1

One type of decoherence is relaxation, which results in the excitation of the qubit slowly leaking away to its environment through different relaxation channels. Some of these relaxation channels have been discussed in 2.3. The relaxation can be measured in a T_1-measurement, a schematic of which is shown in Figure 7.12. In a T_1-measurement an initial pi-pulse is applied, after which the qubit is in its excited state. After waiting for a monotonically increasing wait time τ, the state of the qubit is measured. During the wait time τ the qubit experiences an exponential decay due to relaxation, with a corresponding decay time T_1.

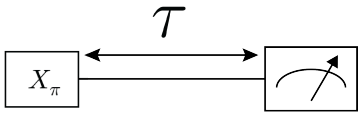


Figure 7.12: Schematic of a T1 sequence. An X_π pulse is applied to the qubit, and after a waiting time τ the qubit is measured.

- Purcell limit

7.3.3.2 Qubit dephasing: Ramsey

Aside from relaxation, the qubit also experiences dephasing, resulting from phase noise. Phase noise can be seen as fluctuations in the qubit frequency, and results in the random accumulation of phase. It can be visualized on the Bloch sphere as the transversal component of the qubit's state decreasing in magnitude, as the qubit's phase has increased uncertainty. In the limiting case this will result in all phase information being lost, with the qubit's state thereby lying on the Z-axis.

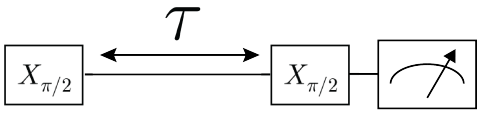


Figure 7.13: Schematic of a T_2^* sequence.

The amount of dephasing is characterized by the dephasing time T_2^*, which has an upper bound equal to 2 T_1 due to dissipation [3, pp56-58]. However it can decrease significantly due to other sources of phase noise, such as charge noise, fluctuating cavity photon number, and flux noise for flux-tunable qubits [30, p126].

Measuring the dephasing time T_2^* is done in a Ramsey measurement, a schematic of which is shown in Figure 7.13. A Ramsey sequence consists of an initial $X_{\pi/2}$ pulse, after which the qubit lies on the equator of the Bloch sphere. After a certain wait time τ, a second $X_{\pi/2}$ pulse is applied to the qubit. The combination of the two pulses should result in the qubit ending up at the excited-state. However, during the wait time τ the qubit experiences dephasing, causing it to deviate from its original position on the equator. Therefore the final

state of the qubit will deviate from the qubit's excited-state. The probability of the final state to end up in the excited-state has an exponential decay, asymptotically approaching 0.5 (all phase information lost).

During the wait time τ, when the qubit lies on the equator, it does not only experience dephasing. When the frequency of the driving tone ω_d is different from the qubit frequency ω_q, the qubit acquires a phase with respect to the rotating frame of the drive. As a result the qubit will precess along the equator with a frequency equal to the frequency difference $\omega_d - \omega_q$. Due to this precession the Ramsey measurement also exhibits an oscillation. The frequency of the oscillation can in fact be used to accurately determine the qubit frequency (see Section **??**).

7.3.3.3 Fast frequency qubit dephasing: Echo

The dephasing time T_2^* is a combination of several different phase noise sources. Some of these sources produces high frequency (fast) noise, while others produce low frequency (slow) noise. It is possible to distinguish these two effects by performing a second dephasing measurements that filters out slow noise, called an Echo measurement, and its schematic is shown in Figure 7.14.

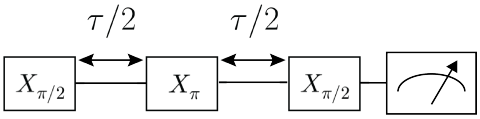


Figure 7.14: Schematic of a T_2^E sequence.

An Echo measurement is quite similar to a Ramsey measurements, where two $X_{\pi/2}$ pulses are applied, separated by a wait time τ. The difference is that in the middle of this wait time, at $\tau/2$, an additional refocusing X_π pulse is sent. The refocusing pulse flips the qubit state on the Bloch sphere around the X-axis. Any slow phase noise, which we can view to be quasi-static, is thereby cancelled. Fast phase noise, however, will vary considerably during the wait time τ, and so will still cause dephasing. In the absence of decoherence, the final state of the qubit is the ground-state, which is in contrast to a Ramsey measurement, where the final state is the excited-state.

An Echo measurement is performed by monotonically increasing the wait-time τ, while keeping the three pulses relative to the wait time τ fixed. The result is similar to a Ramsey measurement, showing an exponential decay, with corresponding Echo dephasing time T_2^E. This value should be higher or equal to the dephasing time T_2^*. The refocusing pulse has the additional effect that any precession due to detuning is also cancelled, thereby inhibiting the oscillatory behaviour that is present in Ramsey measurements. To be able to better estimate the Echo dephasing time T_2^E, the phase of the final $X_{\pi/2}$ pulse is monotonically shifted, resulting in an artificial oscillation.

7.3.4 Second excited-state

When viewing the transmon as a qubit the higher energy levels are neglected. However, qubit pulses may result in some population leakage to the second excited-state. This is especially the case when the

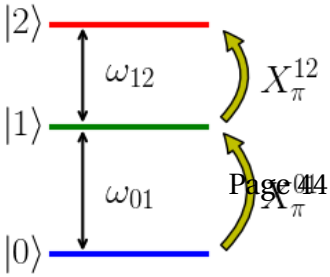


Figure 7.15: Schematic of the pulse

pulse length is short and when the DRAG parameter is not well-calibrated (see Section 8.2.4 for information on the DRAG pulse). Measuring the population in the second-excited state due to leakage requires being able to excite the transmon to its second excited-state. This can be done by first exciting the transmon to its first excited-state, and then applying a second pulse at the second transition frequency of the transmon, as shown in Figure 7.15. Measuring this frequency has been explained in Section 7.2.6.

Determining the correct amplitude for the pulse at the second transition frequency can be done using a second transition Rabi measurement, where an initial X_π-pulse is applied to excite the transmon to its first excited-state. The amplitude of the second pulse is varied from negative to positive, similar to a standard Rabi measurement. Optionally a third X_π-pulse at the first transition frequency of the transmon. The third pulse transfers any remaining excitation in the first excited-state back to the ground-state.

The pulse-amplitude required to pulse to the second excited-state has been calibrated for both the top and bottom qubit. In Figure 7.16 a second transition Rabi measurement is shown for the top qubit. The measurement is performed both with and without a third X_π pulse. As can be seen the signal reaches above the signal of the first excited-state, and so it can be concluded that the second-excited state is indeed populated. Furthermore, since the signal at the second excited-state peaks are identical for the measurements with and without the final pulse, it can be concluded that at these amplitudes there is no residual excitation in the first excited-state, and that indeed the transmon is fully excited to the second excited-state.

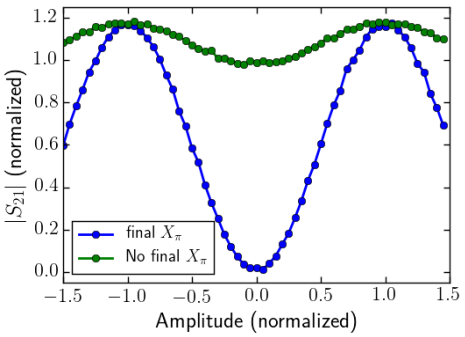


Figure 7.16: A second transition Rabi measurement of the second excited-state, both with and without final X_π pulse. The signal is rotated and normalized such that $|S_{21}| = 0$ corresponds to the transmon in the ground-state, and $|S_{21}| = 1$ corresponds to the transmon in the first excited-state.

7.4 Exploring Frequency re-use

The frequencies of the qubits and their corresponding resonators are shown in Table 7.1. The resonator buses have a frequency of 4.88 GHz and 4.97 GHz for the bus connecting the ancilla qubit to the top and bottom qubit respectively (see Appendix C.4).

To study frequency re-use, the top qubit frequency has been tuned to that of the bottom qubit, while the ancilla and bottom qubits were kept at their respective sweet-spots. Under these conditions the decoherence times of the three qubits are shown in Figure 7.17. The dephasing time T_2^* of the top qubit is considerably lower than of the ancilla and bottom qubit.

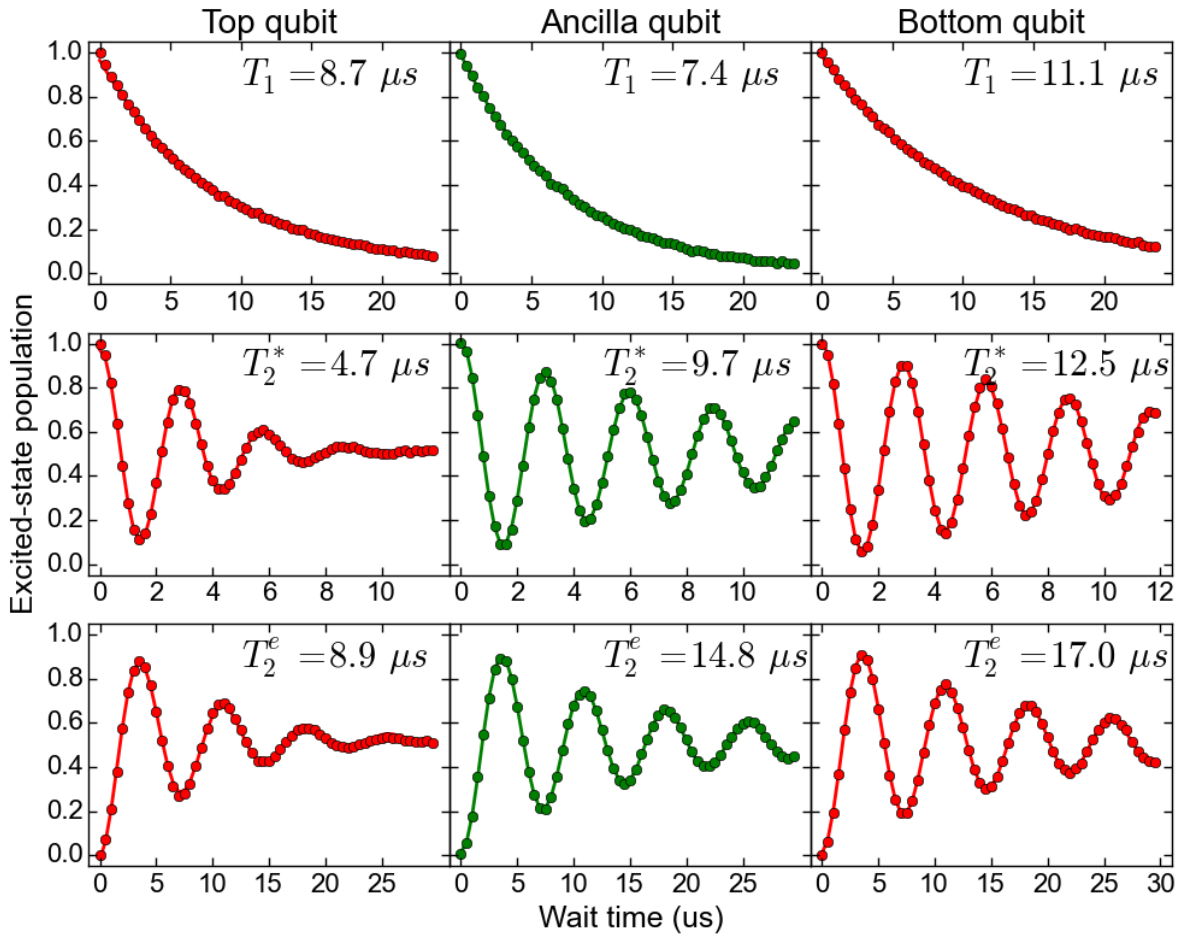


Figure 7.17: Decoherence times of the three qubits in Muxmon0. The top qubit frequency is tuned to match the bottom qubit frequency (6.22 GHz). When measuring the coherence times of either the top or bottom qubit the other qubit was detuned by 50 MHz to suppress cross-coupling effects. The fit for the dephasing time T_2^* of the top qubit was performed using a Gaussian noise model.

The reason is that the top qubit is tuned away from its sweet-spot to match the frequency of the bottom qubit. As a result its frequency response to flux is higher, and so it is more susceptible to flux noise.

Having the top and bottom qubit at the same frequency gives rise to cross-coupling and cross-driving effects. It is important to characterize these effects, as they can potentially result in gate errors.

7.4.1 Cross-coupling

The top qubit is coupled to the botom qubit via the following three successive components:

1. The resonator bus coupling the top qubit to the ancilla qubit.
2. The ancilla qubit.

Qubit	f_{max} (GHz)	f_{res} (GHz)
Top	6.277	6.700
Ancilla	6.551	6.733
Bottom	6.220	6.800

Table 7.1: Sweet-spot frequencies f_{max}, and resonator frequencies f_{res} of the three qubits in the Muxmon0 chip

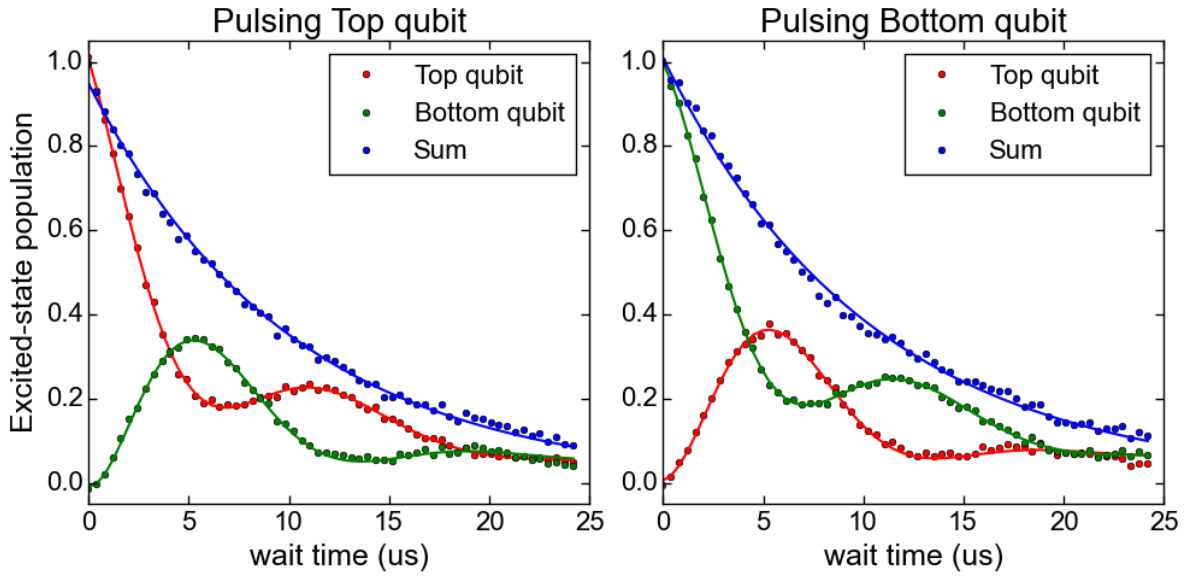


Figure 7.18: After initially exciting one qubit, and measuring the excited-state population of both qubits versus time, an excitation swap is observed. The extracted coupling strength is equal to $J2\pi = 72.0 \pm 1.8$ kHz

3. The resonator bus coupling the two ancilla qubit to the bottom qubit.

When the top and bottom qubit are tuned into resonance, the two qubits experience an exchange interaction. This interaction leads to an excitation in one qubit being able to transfer to the other qubit. This can result in the swapping of excitation between the top and bottom qubit, at a rate given by the interaction strength J. The excitation swapping of the top and bottom qubit when tuned into resonance can be seen in Figure 7.18. From these results the coupling J is found to be equal to $J/2\pi = 72. \pm 1.8$ kHz. For more information on the exchange interaction see section 4.3.2 of the thesis by Chow [6].

7.4.2 Cross-driving

When driving one of the qubits through its dedicated drive line, the signal can partially leak through to the other qubits. The cause of this leakage may be on-chip, where the components separating the qubits do not fully filter the signal, or may be off-chip, due to for instance

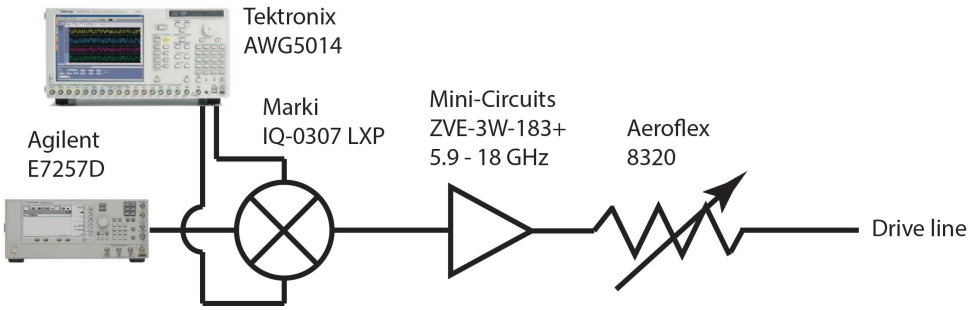


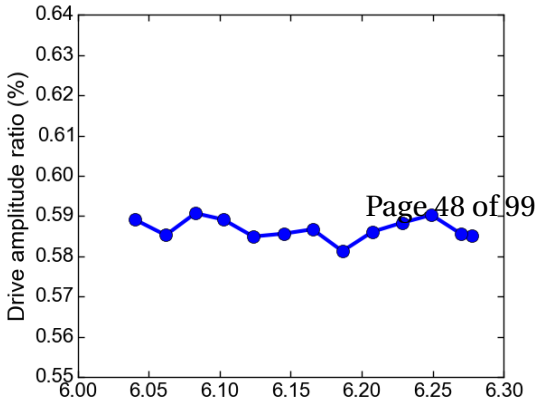
Figure 7.19: Schematic for measuring the amount of cross-driving using the direct drive lines of the top and bottom qubit. The combination of amplifier and variable attenuator allow a range range of drive strengths.

qubit	cross-driving (%)	
	top drive line	bottom drive line
Top	–	0.23
Ancilla	0.79	0.66
Bottom	0.57	–

imperfect isolation between the cables or other components. This signal leakage results in cross-driving effect, where driving one qubit will also partially drive the other qubits. Cross-driving effects affect the performance of the qubits when the frequency of the driven qubit and the cross-driven qubit are in resonance, as is the case with the Muxmon experiment.

The amount of cross-driving can be determined by measuring the pulse amplitude required to drive each of the three qubits through a drive line. The cross-driving due to the finite isolation of the Duplexer has been separately measured (see Appendix B). At the frequencies used in the experiment, the isolation of the Duplexer was found to be typically around 50 dBm. The amount of cross-driving due to other sources was determined by measuring the pulse amplitude required to perform a Rabi on each of the qubits using a fixed drive line. The measurements were performed using the set-up shown in Figure 7.19. The results of a single cross-driving measurement is shown in Figure 7.20, where the top and bottom qubit are driven through the drive line of the top qubit. The pulse amplitude is normalized to the amplitude required for applying pi pulse to the qubit directly connected to the drive line. The amount of cross-driving is equal to the ratio of the pulse amplitude required for a pi rotation for the main qubit, and for the cross-driven qubit. The cross-driving ratio's are shown in Tables **??** and **??**. The cross-driving ratio's are found to be less than one percent, and are higher for the ancilla qubit than for the other cross-driven qubit. This indicates that the main source of cross-driving is likely on-chip. Furthermore it can be seen that the cross-driving is stronger from the drive line of the top qubit than from the drive line of the bottom qubit.

As a check that the effects observed are indeed due to cross-driving, and not due to cross-coupling, the cross-driving when driving the bottom qubit through the drive line of the top qubit



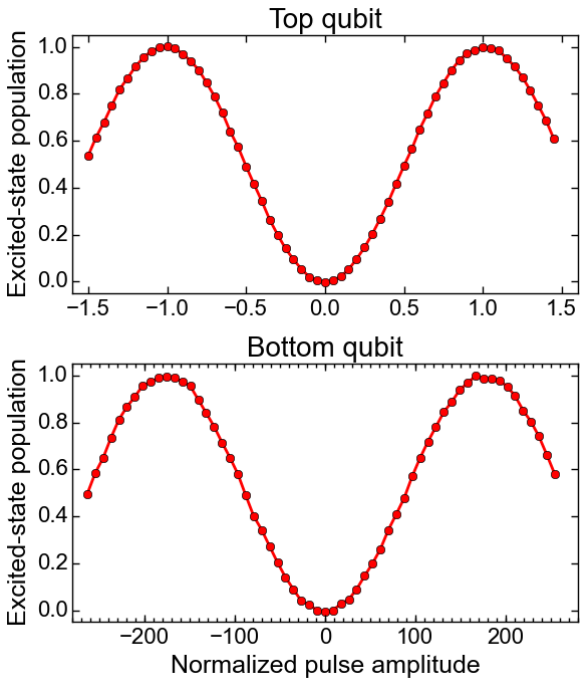


Figure 7.20: Required drive amplitude required to drive the top and bottom qubit through the top qubit drive line. The amplitude has been normalize to the amplitude required for a pi pulse on the top qubit.

has been measured as the frequency of the top qubit is varied. The results are shown in Figure 7.21. If the effect is due to cross-coupling the amount of cross-driving should depend strongly on the detuning between the top qubit and bottom qubit. Instead we see that the cross-driving is approximately constant, indicating that this is indeed a cross-driving effect instead of a cross-coupling effect.

Chapter 8

Calibration routines

In the Muxmon experiment we want to find out what the best possible qubit performance is in a frequency re-use set-up. The qubit pulses must therefore be optimally tuned. Section 8.1 deals with the calibration routines used to calibrate the mixer and Duplexer, which would otherwise add to the gate error. Section 8.2 describes the calibration routines used to accurately tune the qubit and its pulses. These routines are have been automatized, and so can also be applied periodically during long measurements.

8.1 Instrument calibrations

8.1.1 IQ mixer calibration

An important calibration routine which must not be forgotten is correcting for mixer imperfections. An IQ mixer must be calibrated for two types of mixer imperfections: the mixer carrier leakage, and the mixer skewness. These imperfections result in two spurious signals, which are shown in Figure 8.1.

For a perfectly balanced mixer, a signal of given frequency ω_0 at the LO port should produce no output at the RF port if no modulation signal is applied in the inphase and quadrature ports. However, any imperfections, due to for instance diode mismatches in the mixer, may lead to some signal at frequency ω_0 leaking through. This leakage can be compensated to a large extent by adding a DC offset to the inphase and quadrature IF signals leaving the AWG, for which the leakage is minimized. Determining the optimal DC offset can be performed by sending a carrier signal into the LO port of the mixer, and a continuous DC signal from the AWG to both IF ports of the mixer. The leakage can be measured as signal exiting the RF port at the carrier frequency ω_0, and can be minimized by varying the offsets of the individual AWG channels.

Another type of IQ mixer imperfection is mixer skewness. The carrier signal entering the LO port is split into its inphase and quadrature component, where it is mixed with an inphase and quadrature IF signal. Ideally the inphase and quadrature components of the LO signal are perfectly orthogonal. However, in reality this is not the case, and so a small amount of skewness is present. Ideally the signal should be shifted in frequency to the desired

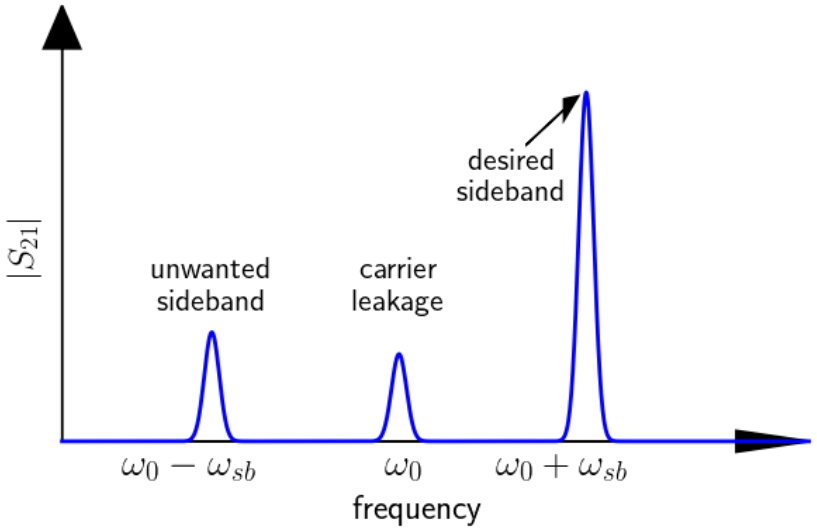


Figure 8.1: Schematic showing spurious signals resulting from mixer leakage and mixer skewness,when using single sideband modulation with frequency ω_{sb}.

sideband frequency $\omega_0 + \omega_{sb}$. However, any mixer skewness will lead to the signal being partially shifted in the opposite direction, resulting in some signal at the unwanted sideband frequency $\omega_0 - \omega_{sb}$. Additionally any amplitude skewness between the two IF ports will also result in signal at the unwanted sideband frequency. The signal at the unwanted sideband can be measured for a given carrier frequency ω_0 by adding a sine and cosine with sideband frequency ω_{sb} to the inphase and quadrature ports respectively. This can be corrected during the generation of the pulses through a transformation $(I,Q) \rightarrow \left(I',Q'\right) = \left(I - Q\tan\phi, Q\sec\phi\right)$, where ϕ is the phase skewness. The skewness can be corrected by varying the phase and amplitude of one of the AWG channels, and minimizing the signal at the unwanted sideband. The inphase and quadrature signals have to be transformed for every phase. Note that the mixer skewness is dependent on both the carrier frequency ω_0 and the sideband modulation frequency ω_{sb}

8.1.2 Duplexer phase calibration

The Duplexer has phase-shifters for each of input-output combinations. That means that the phases of each of the eight channels can be tuned individually.. In the case of the Muxmon experiment, where we have independent control over the main pulse and the DRAG pulse, we require the two channels to share the same phase. The phase of two Duplexer channels can be tuned to one another by sending two signals with opposite phase into the Duplexer. If the relative phase shift of these two signals induced by the Duplexer, these two signals should (partially) cancel each other, resulting in a dip in transmission. The amount of transmission at this dip depends on the amplitude difference between the two signals.

Calibrating the Duplexer phase can be done by splitting a signal from an RF generator, and then using single sideband modulation on both signals individually, where the phase of

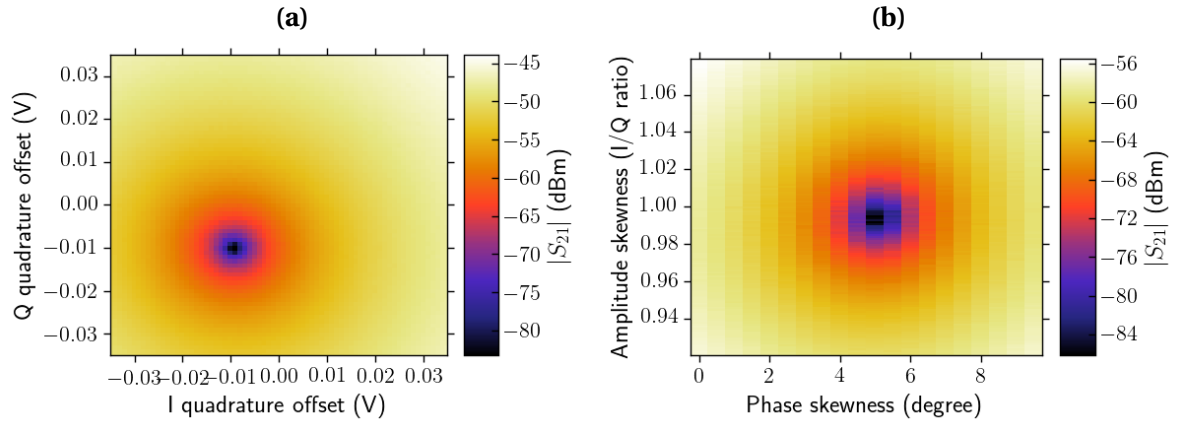


Figure 8.2: Mixer carrier leakage (a) and mixer skewness (b). Without any corrections both the mixer carrier leakage and the signal due to mixer skewness are equal to −58 dBm. Compared to the main signal with output power equal to −32 dBm, the mixer imperfections would result in two additional signals, each with strength −26 dB.

one of the sidebands is shifted by 180°, which can be realized using an AWG. As the phase of one of the channels is varied, a dip in transmission corresponds to both Duplexer channels sharing the same phase shift.

8.2 QUBIT CALIBRATIONS

8.2.1 Accurate frequency estimation

Spectroscopy provides an estimate for the frequency of the qubit. However, even with pulsed spectroscopy, the accuracy is limited to about 1 MHz. As explained in Section 7.3.3.2, any difference between a drive frequency ω_d and qubit frequency ω_q will result in the qubit precessing around the Z-axis.

In a Ramsey measurement this precession can be measured, from which the qubit frequency can be inferred. This provides a very accurate estimate for the qubit frequency ω_q. The longer the wait time τ, the more the qubit precesses around the equator. Small differences between the drive frequency ω_{drive} and the qubit frequency ω_q can be detected. The upper bound on the accuracy of being able to determine the qubit frequency is set by the qubit's dephasing time T_2^*, because it corresponds to the inherent fluctuation in the qubit frequency.

The main goal of the Muxmon experiment is to determine the performance of selective broadcasting combined with frequency re-use, which requires the top and bottom qubit to be at the same frequency. Any effect due to weak coupling between the qubits is only present if the two qubits are accurately tuned to the same frequency. Furthermore, any frequency detuning from the drive frequency ω_d will lead to a decrease in qubit performance. The top qubit frequency was initially tuned using spectroscopy. Accurate frequency tuning was then performed using Ramsey measurements. The bottom qubit frequency ω_q^B was first

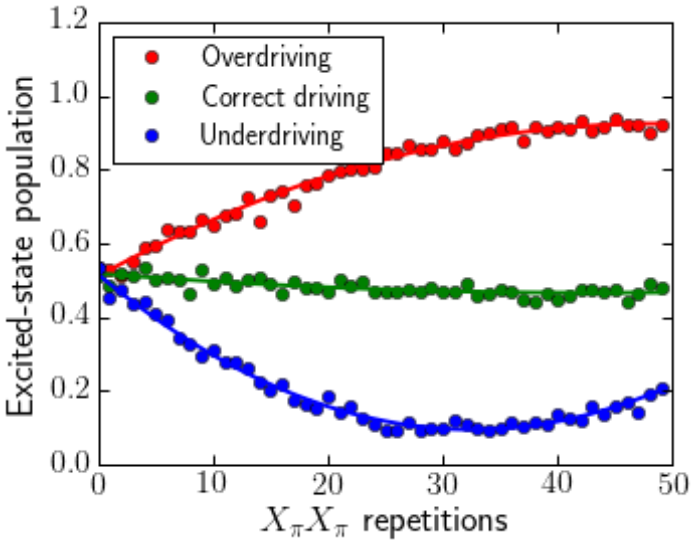


Figure 8.3: Accurate drive calibration using repeated pi-pulses. The initial slope determines if the system is overdriving (positive slope), or underdriving (negative slope)

determined, after which the top qubit frequency ω_q^T was tuned to match ω_q^B. Using Ramsey measurements the inidividual frequencies could be determined to within 10 kHz, and the two frequencies could be tuned to within 50 kHz of one another. The limit to tuning the frequencies was not determined by the Ramsey measurement accuracy, but due to the IVVI having a finite DAC voltage stepsize.

8.2.2 Multiplexed readout calibration

In standard heterodyne detection the output signal is downconverted in a mixer using an LO frequency ω_{LO}that is slightly different from the output frequency ω_{out}. The resulting signal is a combination of the sum $\omega_{\text{LO}} + \omega_{\text{out}}$ of the two frequencies, and the difference $\omega_{\text{LO}} - \omega_{\text{out}}$ of the two frequencies. The high-frequency $\omega_{\text{LO}} - \omega_{\text{out}}$ is filtered out using a low-pass filter. The difference frequency $\omega_{\text{LO}} - \omega_{\text{out}}$ is known as the intermediate-frequency (IF), of which the signal is measured.

If the output signal is the combination of multiple signals with different frequencies, these can be... measured. This is known as multiplexed readout.

8.2.3 Accurate drive amplitude calibration

To test the limits of performance using the Duplexer all gates need to be calibrated to a very high accuracy. The Rabi measurement explained in section 7.3.2 is able to tune the drive amplitude up to a certain degree. However, the degree to which one can tune the drive amplitude using Rabi is limited, and for fine-tuning different methods are required.

In the Muxmon experiment, the method used for accurate drive amplitude calibration is based on applying repeated pi-pulses on the qubit. The entire sequence can be summarized

as $(X_\pi)^{2N} X_{\pi/2}|0\rangle$, where N is the segment number. In the absence of gate errors and decoherence, the qubit should return to the equator, regardless of the segment number N. However, any amount of overdriving or underdriving results in small rotations that are added coherently, resulting in a positive or negative slope respectively. These slopes serve as very accurate measures for the optimal drive amplitude. Three such examples are shown in Figure 8.3. If the difference in driving strength is large, oscillations will be present, corresponding to rotations around the Bloch sphere.

8.2.4 DRAG parameter calibration

Ideally a transmon has two energy levels, and so can be treated as a qubit. However, in Section 7.2.1 it was shown that the degree to which we can ignore the second excited-state depends on the anharmonicity $\alpha = \omega_{12} - \omega_{01}$. The presence of the second excited-state (and to lesser extent higher excited-states) can have a considerable influence on the gate performance in two different ways. First of all gates can cause a direct excitation into the second excited-state, thereby resulting in leakage outside the qubit two-state Hilbert space. Ensuring that the frequency bandwidth corresonding to the gate is small compared to the anharmonicity α can greatly reduce this leakage. However, an additional form of leakage is manifested as a virtual excitation to the second excited-state during the pulse, resulting in an added phase error.

A modification to the Gaussian, known as the Derivative Removal by Adiabatic Gate (DRAG) pulse [20], can be used to suppress leakage errors. The DRAG pulse relies on quadrature control, where one quadrature consists ofthe main Gaussian pulse, while the quadrature consists of the derivative of the main Gaussian pulse. The DRAG pulse can reduce leakage and virtual excitation-induced phase errors by roughly an order of magnitude. The DRAG pulse is shown in Figure 8.4.

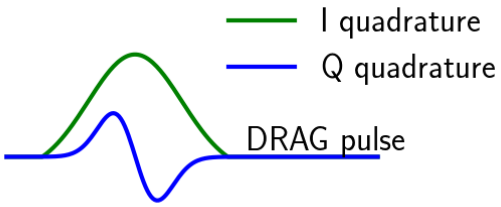


Figure 8.4: Schematic representation of the DRAG pulse. The I quadrature pulse is the main Gaussian pulse, while the Q quadrature pulse is the derivative of the Gaussian.

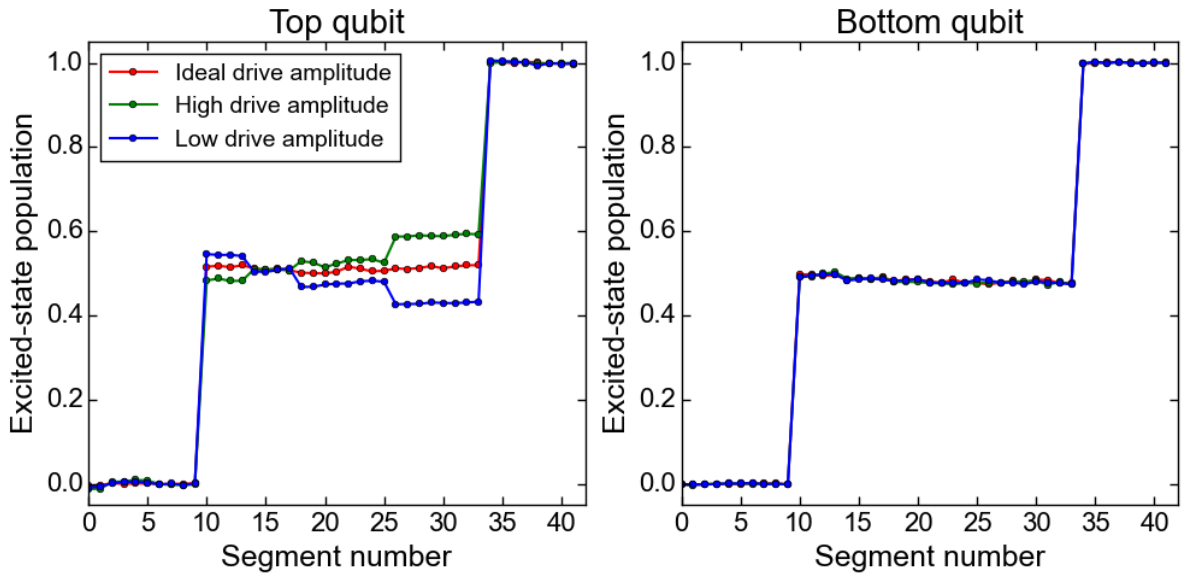
The DRAG parameter determines the amplitude of the derivative pulse, and must be accurately calibrated. It has been shown that the virtual excitation-induced phase error is present up to first order in a pi/2-pulse [15]. In the Muxmon experiment the DRAG parameter has been calibrated using a method suggested by Reed [26], by measuring the difference in signal between an $X_\pi Y_{\pi/2}$ pulse and a $Y_\pi X_{\pi/2}$ pulse. Ideally for both pulses the final state should lie on the equator. However, any phase error would result in these two measurement shifting away from the equator in opposite direction. The DRAG parameter is found by minimizing this difference.

8.2.5 AllXY

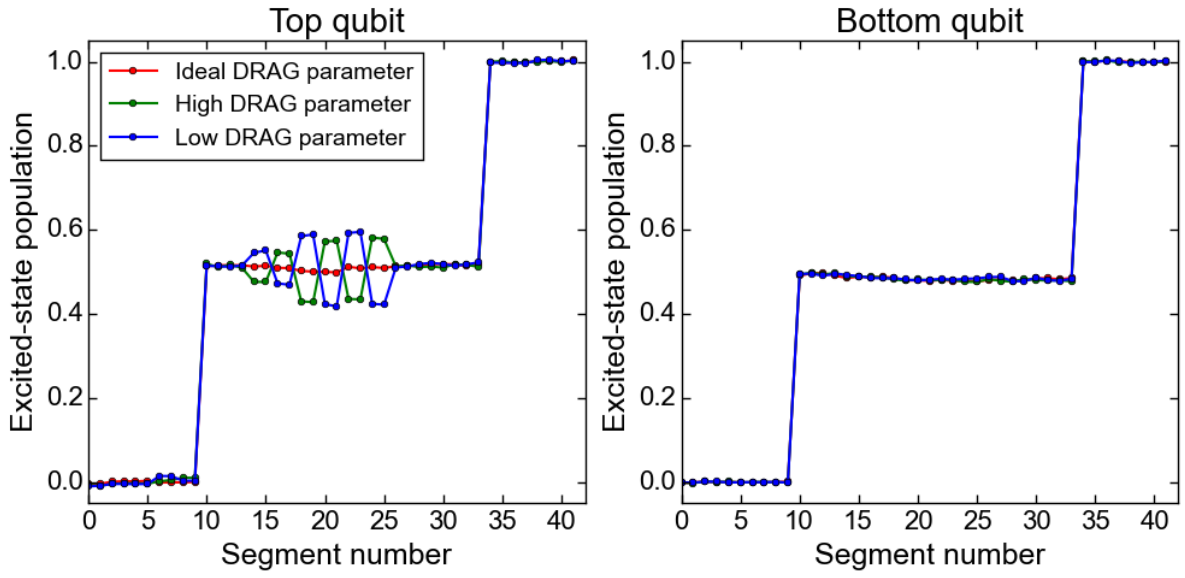
There is a good measurement to test how well the qubit has been tuned. It is known as the AllXY measurement, and consists of all 21 possible two-gate combinations of $\{Id, X_\pi, Y_\pi, X_{\pi/2}, Y_{\pi/2}\}$. Each combination is susceptible to different types of gate errors to a different degree. The AllXY combinations have therefore been arranged in such a way that the final state of the first five combinations is the ground-state, the final state of the second twelve combinatios is the equator, and the final state of the last four combinations is the excited-state. Furthermore the combinations are arranged in such a way that the most common sources of gate errors can be distinguished. The full list of combinations in correct order can be found in Appendix E.

The errors that can be distinguished are extensive: drive amplitude, DRAG parameter, frequency detuning, signal reflections, and several more. This is the strength of AllXY, but simultaneously its weakness. If there are multiple errors present, their errors may interfere, resulting in symptoms which are difficult to diagnose. nevertheless, it is a powerful tool, especially if one source of gate error is dominant. For a detailed analysis of the AllXY symptoms produced by different types of gate errors, see Reef [26].

The Duplexer can modify the signal of each of its input-output port combinations individually. In the set-up used for the Muxmon experiment, both the drive amplitude and DRAG parameter can be separately tuned for each of the two qubits. In Figure 8.5 the AllXY results are shown for the top and bottom qubits, as either the drive amplitude or DRAG parameter of the top qubit is varied using the Duplexer. As can be seen there is no noticeable change in the AllXY of the bottom qubit, even though the top qubit's performance changes considerably. This shows that the Duplexer allows for individual drive amplitude and DRAG parameter control for each qubit, without affecting the other.



(a)



(b)

Figure 8.5: AllXY measurement of the top and bottom qubit as one parameter of the top qubit is varied. The parameter varied is (a) drive amplitude, (b) DRAG amplitude. The AllXY sequences are placed on top of each other. As can be seen the bottom qubit has no noticeable change in the AllXY measurements when the parameters of the top qubit are varied.

Chapter 9

Randomized benchmarking

9.1 Introduction

When preparing the qubits for a quantum algorithm it is important to know what the performance is of the qubits, which can be characterized by its gate errors. One method has already been discussed in Section 8.2.5, namely the AllXY method. Although the AllXY is good at detecting specific errors, it is a crude method, which is unable to characterize gate errors with high accuracy. Another method used for gate characterization is quantum process tomography. Quantum process tomography measures the output density matrix resulting from the application of a specific to each of the system's basis states. Due to linear superposition of quantum states, the full effect of a gate is thereby characterized, including the different types of gate errors.

Quantum process tomography suffers from some important drawbacks. The first is that it is sensitive to state preparation and measurement errors, which make it difficult to distinguish whether the gate errors originate from the actual gate being characterized, or from the gates used for preparation and measurement. Another related drawback is that it is unable to measure small gate errors, which are required for fault-tolerant quantum computing. Additionally, quantum process tomography suffers from an exponential scaling in measurement with the number of qubits. These drawbacks result in quantum process tomography being an inadequate gate characterization method for gates with small errors.

An alternative method to quantum process tomography is randomized benchmarking, described by Knill et al. [13]. It is based on repeated application of operations drawn randomly from a set of unitary operations, after which the qubit fidelity to the theoretical final state (in the absense of errors) is measured. This fidelity can then be translated to an average unitary operation error. The set of unitaries used in randomized benchmarking may be single-qubit or multi-qubit operations. Randomized benchmarking has the advantage that it is insensitive to state preparation and measurement errors, and is a relatively fast measurement, which is able to accurately determine the average unitary operation error. Compared to quantum process tomography the drawback of randomized benchmarking is that it does not provide additional information about the type of gate error. However, randomized benchmarking has been used to distinguish unitary gate errors, such as overrotations, from non-unitary gate

errors, such as decoherence effects [31].

Modifications of the randomized benchmarking protocol have resulted in a versatile range of applications. Interleaved randomized benchmarking [17], a variation to the randomized benchmarking method, is able to characterize the performance of a single gate or unitary operation. This is done by interleaving the specific operation with the randomized benchmarking unitary operations, resulting in a uniform sampling of the targeted operation. Another variation of the randomized benchmarking scheme has furthermore been used as a metrological tool to study phase noise at very low timescales [24]. In a more recent work, Johnson et al. [12] demonstrated a quantum gate tomography scheme based on randomized benchmarking. This method has the advantage of being insensitive to state preparation and measurement errors.

In the Muxmon experiment randomized benchmarking is used to characterize the performance of the qubits when using the Duplexer, and using selective broadcasting. The set of unitary operations is the single-qubit Clifford group, which will be discussed in Section 9.2.

9.2 CLIFFORD GROUP

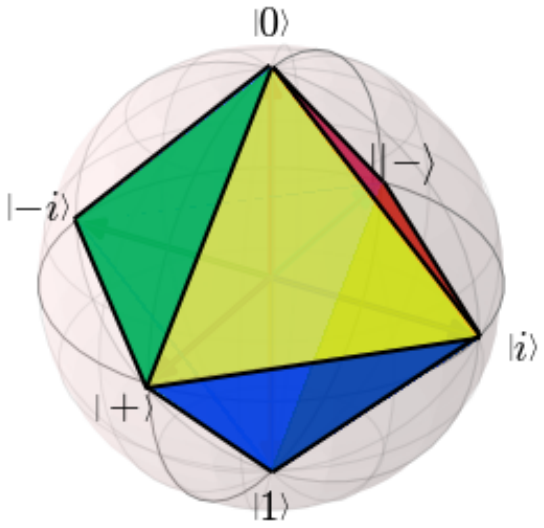


Figure 9.1: The single-qubit Clifford group can be visualized using an octahedron whose vertices correspond to the six cardinal states. The 24 Cliffords composing the single-qubit Clifford group correspond to the 24 distinct rotations of the octahedron such that the vertices are mapped onto themselves.

The single qubit Pauli group consists of the two-dimensional identity operator and the three Pauli matrices. The n qubit Pauli group is generated by the tensor product of n single qubit Pauli groups. The n qubit Clifford group $\mathcal{C}_n$ is equal to the set of unitary transformations

on the n qubit Pauli group up to a global phase. The n qubit Clifford group $\mathcal{C}_n$ is generated by $\{H_i, P_i, CNOT_{ij}\} \forall i, j \in (1, \ldots, n)$, where:

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \qquad P = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}, \qquad CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{9.1}$$

The single qubit Clifford group $\mathcal{C}_1$ can be visualized using an octahedron in the Bloch sphere, as shown in Figure 9.1. Each Clifford in $\mathcal{C}_1$ corresponds to a distinct rotation of the octahedron, where the six cardinal states are mapped onto themselves. If one sets the constraint that the state $|0\rangle$ is mapped to itself, four distinct rotations are possible. Since the state $|0\rangle$ can be mapped to each of the six cardinal states, there are $6 \times 4 = 24$ distinct rotations, and hence 24 Cliffords in the single qubit Clifford group $\mathcal{C}_1$.

Each of the Cliffords in the single qubit Clifford group $\mathcal{C}_1$ can be decomposed into rotations along the X and Y axis, as is shown in Appendix F.1. Each Clifford can be decomposed into a minimum of zero (identity) gates and a maximum of three gates. If one includes the identity as a gate, a Clifford requires on average 1.875 gates.

The Clifford group does not form a universal set of gates. This is because the rotations only map cardinal states onto themselves. Since single qubit gates and CNOT is universal [22], the Clifford group combined with the T-gate ($\pi/4$ rotation) constitute a set of universal quantum gates.

9.3 THE RANDOMIZED BENCHMARKING PROTOCOL

Randomized benchmarking is a method to characterize the performance on qubits. It is based on repeated application of operations on the system, and measuring the fidelity to the ideal final state. The operations are randomly chosen from a set of unitary operations.

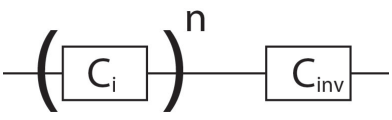


Figure 9.2: Schematic of the randomized benchmarking protocol

In the case of the Muxmon experiment randomized benchmarking is performed on individual qubits, and the set of unitary operations is the single qubit Clifford group. The randomized benchmarking protocol used in this experiment, depicted in Figure 9.2, is as follows:

1. Initialize the qubit in the ground state
2. Apply n consecutive Cliffords $(C_1, \ldots, C_n)$, where $C_i \in \mathcal{C}_1 \ \forall \ i \in [1, \ldots, n]$
3. apply final inverting Clifford $C_{\text{inv}} = (C_n \ldots C_1)^{-1}$
4. Measure the state of the qubit

After the final inverting Clifford C_{inv} the qubit should return to the ground-state. However, gate errors and decoherence result in the final state having a nonzero population in the

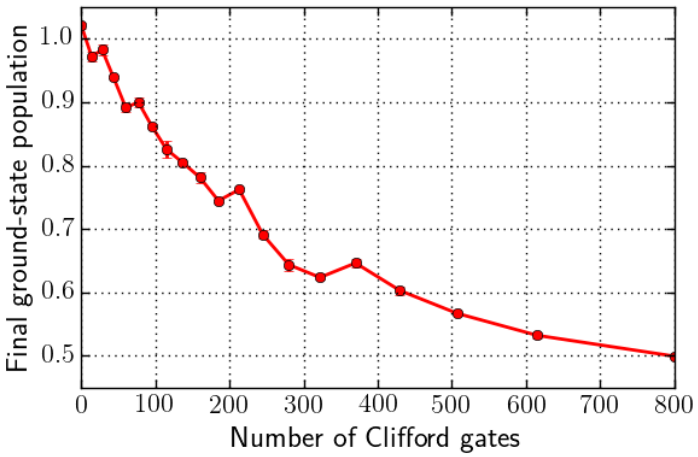


Figure 9.3: Randomized benchmarking results using a single seed. The final ground-state population decreases exponentially as the number of Cliffords n is increased. The points include errorbars, determined by the spread of three separate measurements.

excited-state. The final population in the ground-state P_0 decreases exponentially as the number of Cliffords n is increased. In the limiting case where all information about the qubit is lost the qubit has equal population in the ground- and in the excited-state. The final ground-state population follows an exponential curve given by [13]:

$$P_0 = \frac{1}{2}p^n + \frac{1}{2} \tag{9.2}$$

where p is the decay parameter, related to the average fidelity per operation F_0 by [16]:

$$F_o = \frac{1+p}{2} \tag{9.3}$$

With knowledge of the average number of gates n_g per operation, the average fidelity per operation can be translated to an average fidelity per gate F_g via:

$$F_g = F_o^{1/n_g} \tag{9.4}$$

Qubit relaxation places an upper limit on the number of gates that can be performed, and therefore on the fidelity per gate. This upper limit $F_{\max}$ is given by: [13]:

$$F_{\max} = \frac{1}{6}\left(3 + 2e^{-t_g/2T_1} + e^{-t_g/T_1}\right) \tag{9.5}$$

where t_g is the duration per gate, including any buffer between successive gates. As is expected this upper limit increases with shorter gates.

To characterize the shape of the exponential decay in randomized benchmarking measurements, the successive values of n in a measurement are separated by an exponentially increasing amount. Each set of random Cliffords is known as a seed. The randomized benchmarking results using a single seed are shown in Figure 9.3. Twenty points are chosen between

$n = 0$ and $n = 800$ Clifford operations. As can be seen there is a clear exponential decay, in agreement with Formula 9.2. However, upon closer inspection the curve does not have a perfect exponential shape. This deviation is not due to insufficient averaging, but due to the fact that the Cliffords do not all have the same number of gates. A Clifford is on average composed of 1.875 gates, but may be composed of anywhere between 1 and 3 gates. The result is that in a single seed some segments contain on average more than 1.875 gates per Clifford, while other contain less. For a fixed error per gate, this results in fidelities deviating from the fidelity corresponding to exactly 1.875 gates per Clifford. To obtain an accurate estimate for the exponential decay rate, multiple seeds must be used to average out the random fluctuations from the average gate per Clifford.

9.4 SINGLE QUBIT RANDOMIZED BENCHMARKING

9.4.1 Driving a single qubit versus driving both qubits

After tuning up the top and bottom qubit using the calibration methods explained in Chapter 8, their performance was determined using randomized benchmarking. To test whether the performance of the qubits change when driving only a single qubit or when simultaneously driving both qubits using the Duplexer, randomized benchmarking was performed in both cases. A total of 50 seeds were used for randomized benchmarking. For each seed 20 different number of Cliffords n are used, varying between $n = 0$ and $n = 800$, with exponentially separated points. Re-calibration of the drive amplitude and top qubit frequency was performed every 5 seeds to correct for small fluctuations in time. Pulses of 16 ns were used, with 4 ns buffer between pulses. For 800 Cliffords this corresponds to a total duration of 30 µs.

The randomized benchmarking results when driving only a single qubit or when driving both qubits simultaneously are shown in Figure 9.4. As the number of Cliffords is increased, the ground-state population exponentially approaches the limiting value of 0.5. The performance of both qubits reaches 99.9% gate fidelity, and the performance of the bottom qubit is considerably better than that of the top qubit. This can be expected, as the relaxation time and the dephasing time of the top qubit ($T_1 = 8.7\,\mu s$, $T_2^* = 4.7\,\mu s$) is lower than that of the bottom qubit ($T_1 = 11.1\,\mu s$, $T_2^* = 12.5\,\mu s$). Nevertheless, the gate fidelities of both qubits are still close to the limit imposed by T_1 (see Figure 9.4). This shows that the qubit performance is mainly limited by its decoherence times, and not by the instruments including the AWG and Duplexer.

In Figure 9.4 it can furthermore be seen that there is no discernible difference in qubit performance between driving a single qubit and driving both qubits simultaneously. This indicates that cross-driving and cross-coupling effects do not affect the qubit performance during randomized benchmarking. One would expect that the measured cross-driving would place a lower bound on the qubit gate error (0.57% when driving the bottom qubit via the top qubit drive line, and 0.23% when driving the top qubit via the bottom qubit drive line). However this does not seem to be the case. One possible explanation is that the transmon is a nonlinear system, and since the cross-driving measurements were performed at a high power, the cross-driving at low power might be significantly less.

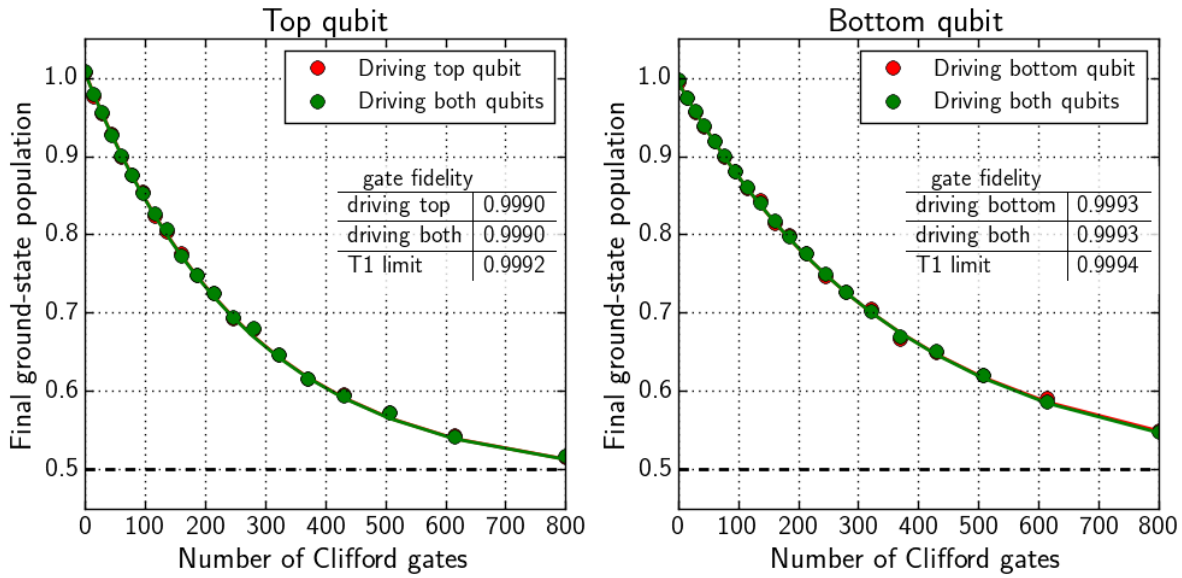


Figure 9.4: Randomized benchmarking results when driving a single qubit versus driving both qubits simultaneously. As can be seen there is no noticeable difference between driving only one qubit versus driving both qubits simultaneously. A total of 50 different seeds were used. T1 limits are calculated using Equation 9.5.

Figure 9.5 shows the ground-state population during randomized benchmarking when only the other qubit is driven. This should ideally be unity irrespective of the number of Cliffords applied, but one can see that application of pulses on one qubit also affects the state of the qubit not being driven. The excitation of the qubit not being driven may be due to cross-coupling, or due to cross-driving, or a combination of the two effects. This effect is much stronger when driving the top qubit and measuring the bottom qubit than the other way around. This is in agreement with the cross-driving measurements in Section **??**, inidicating that the effect is mainly due to cross-driving, and not cross-coupling. Furthermore, the period of the excitation swap was previously found to be equal to $J^{-1} \approx 14\,\mu s$, whereas the duration of the longest pulse sequence, corresponding to $n = 800$ Cliffords, is equl to $30\,\mu s$. Cross-coupling effects should not depend on which qubit is being driven, and so the fact that the top qubit is not excited much, even after $n = 800$ Cliffords, places an upper bound on the cross-coupling effects. This further supports the claim that the main source of the undriven qubit's excitation is due to cross-driving.

9.4.2 Second-state leakage

The exponential fits to the single qubit randomized benchmarking measurements show that the asymptote of the exponential curve actually lies slightly below a ground-state population of 0.5. This is in contrast to expectation, as in the limiting case in randomized benchmarking, where all information about the qubit is lost, a measurement of the qubit should result in the qubit being foud with equal probability in the ground-state and in the excited-state. One

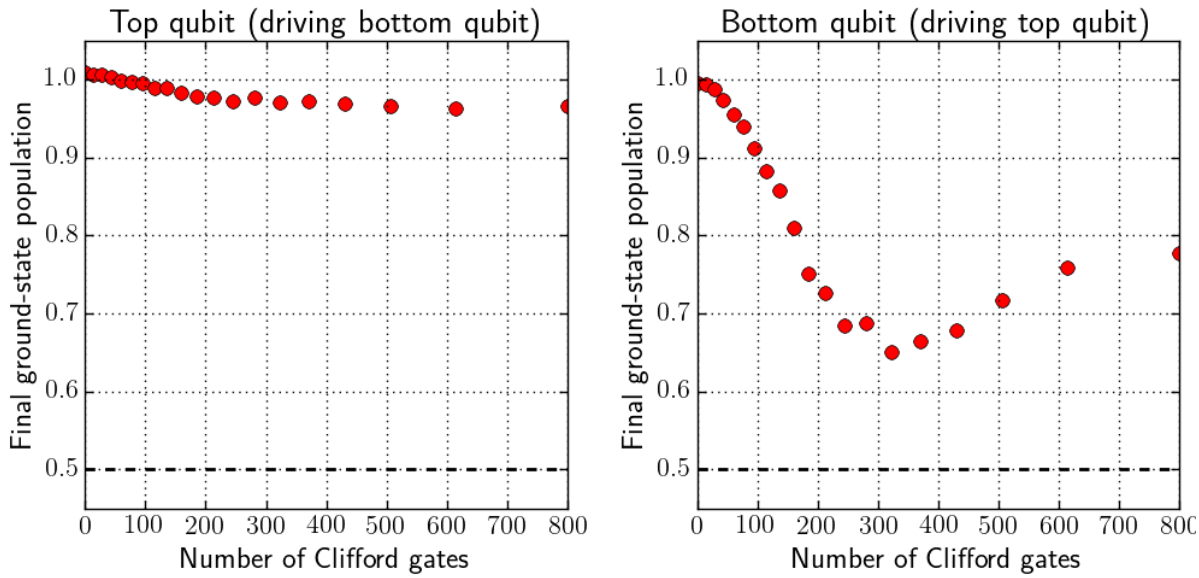


Figure 9.5: Excitation in qubit when only the other qubit is driven during randomized benchmarking. This effect is mostly due to cross-driving and is considerably stronger for the top qubit than for the bottom qubit.

possible explanation for this deviation from 0.5 is that during the randomized benchmarking sequence the qubit experiences leakage to the second excited-state, which would shift the measured signal.

To test whether the exponential curve not saturating at 0.5 is indeed due to leakage, an identical randomized benchmarking measurement set has been performed using the same 50 seeds. At the end of each Clifford sequence a final pi pulse is applied. The result is that the final ground-state and excited-state populations are swapped. In this case the final state should therefore ideally be the excited-state.

If we assume there is no further leakage into states higher than the second-state, these two measurements result in the following set of equations:

$$\begin{aligned} p_0 V_0 + p_1 V_1 + p_2 V_2 &= S_0 \\ p_1 V_1 + p_0 V_1 + p_2 V_2 &= S_1 \\ p_0 + p_1 + p_2 &= 1 \end{aligned} \tag{9.6}$$

where p_i is the final population of state $|i\rangle$, V_i is the measured signal corresponding to state $|i\rangle$, and S_i is the signal measured for a fixed number of Cliffords n where the qubit's final state should ideally by $|i\rangle$. If the signals corresponding to all three states are known, the three populations can be extracted (see Appendix F.3 for details).

Using the methods explained in Section **??**, the signal V_2 of the second-state has been measured and added as a calibration point to the randomized benchmarking sequence. With knowledge of the signals corresponding to all three states, the populations of the first three

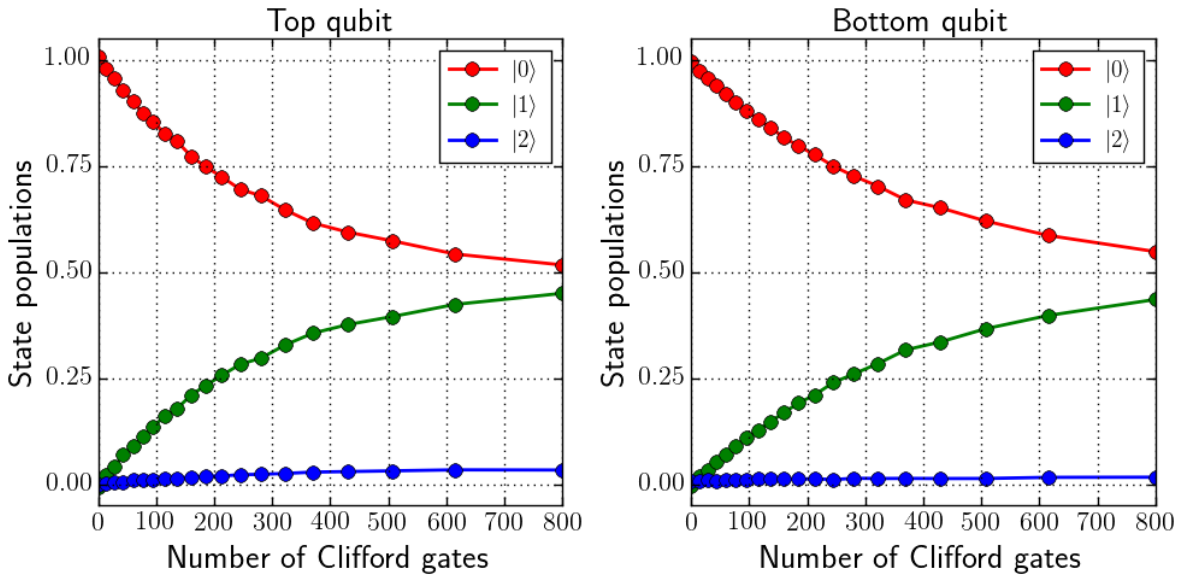


Figure 9.6: The populations of the first three states of the qubit during randomized benchmarking. Populations are extracted by comparing randomized benchmakring results with and without a final pi pulse (see Appendix F.3). As can be seen, both qubits suffer from a small amount of leakage to the second-state. The leakage rate for the top qubit is higher than for the bottom qubit.

states of the qubits during randomized benchmarking have been determined. The results are shown in Figure 9.6. As can be seen there is indeed a small amount of leakage present. The leakage rate is higher for the top qubit than for the bottom qubit. Increasing the pulse length would result in a smaller frequency bandwidth, and is therefore expected to lower the leakage rate. This has, however, not yet been tested.

- Mention that drag can be optimized for leakage, although this would result in a higher gate error. Must find paper

9.5 TWO QUBIT RANDOMIZED BENCHMARKING

So far randomized benchmarking on two qubits has only been performed in the case where both qubits receive the same pulses. In a more realistic scenario one would like to be able to control both qubits individually. The Duplexer, in combination with frequency re-use, offers this possibility. The switches of the Duplexer are able to route pulses at nanosecond-scale. Using a single sequence of pulses, these switches allow each individual pulse to be sent to either of the two qubits, or both qubits simultaneously. This is known as selective broadcasting, and enables individual control of both qubits simultaneously.

The performance of controlling both qubits simultaneously using selective broadcasting has been studied using randomized benchmarking. In this case each of the two qubits has an individual seed composed of n randomly chosen Cliffords. There are many different ways in which randomized benchmarking on two qubits can be performed. Three methods have been

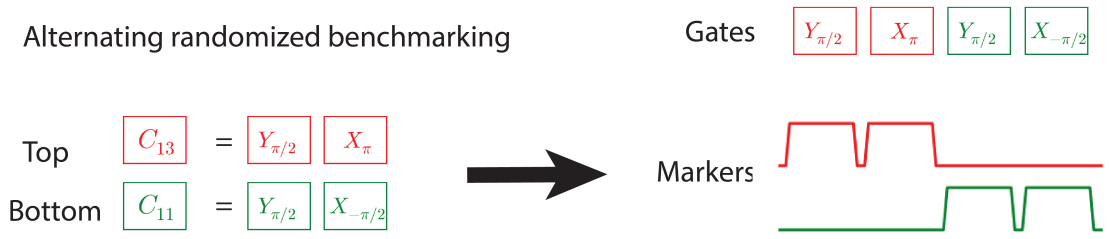


Figure 9.7: Alternating randomized benchmarking

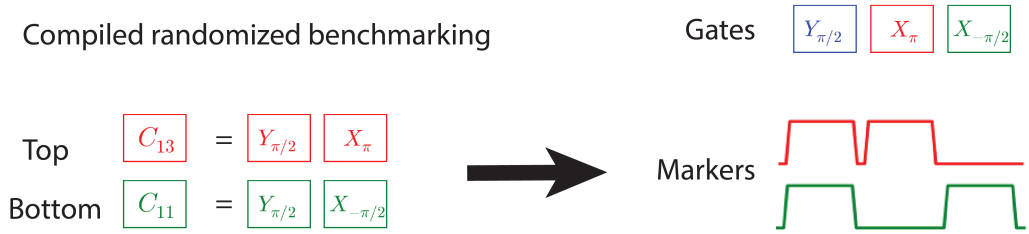


Figure 9.8: Compiled randomized benchmarking

explored: alternating randomized benchmarking, compiled randomized benchmarking, and 5 primitives randomized benchmarking. These will be explained in the following sections.

- Additionally since the qubits are driven by the same pulses, their states should be identical.

9.5.1 Alternating randomized benchmarking

The first method explored is alternating randomized benchmarking. In this method the Cliffords of the two qubits are applied alternatingly, as shown in Figure 9.7. When a Clifford of the top qubit is applied, the switches of the top qubit are on, while the switches of the bottom qubit are off, and vice versa. If each qubit has a seed composed of n Cliffords, the total duration of the sequence will be twice as long as in single qubit randomized benchmarking. Therefore the average duration of a pair of Clifford is equal to $2 \times 1.875 = 3.75$ gates.

9.5.2 Compiled randomized benchmarking

In alternating randomized benchmarking pulses are never simultaneously applied to both qubits. However, it is quite possible that a pair of Cliffords has one or more gates in common. Compiled randomized benchmarking is a more efficient alternative to alternating randomized benchmarking, where pairs of Cliffords are compiled into the least amount of physical gates required to perform both Cliffords. The schematic is shown in Figure 9.8,

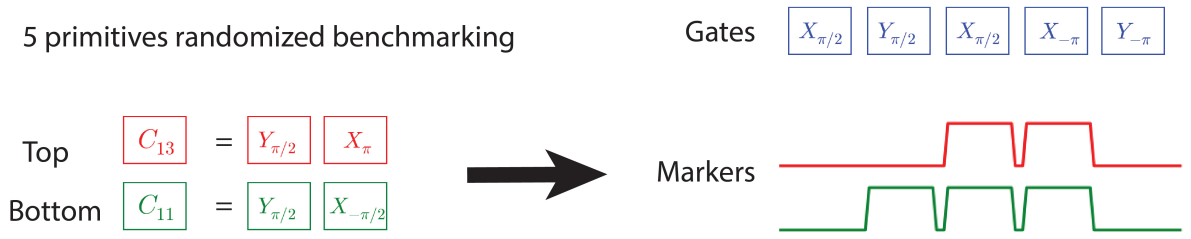


Figure 9.9: 5 primitives randomized benchmarking

where the same Clifford combination is shown as in the schematic of alternating randomized benchmarking (Figure 9.7). As can be seen both Clifford decompositions share the $Y_{\pi/2}$ gate, and so can be combined into a single pulse sent to both qubits simultaneously. The result is that the total number of gates is reduced from four to three.

Compiled randomized benchmarking even goes one step further. The Clifford decomposition shown in Appendix **??** is one way in which the 24 Cliffords can be decomposed. There are, however, many other possible decompositions for each of the Cliffords into rotations along the X and Y axis. It is not at all obvious which decompositions of the two Cliffords would result in the optimal gate compilation for a given Clifford pair, i.e. the least amount of physical gates required to perform the pair of Cliffords on the two qubits.

If we ignore decompositions where successive gates cancel, such as $X_{\pi/2}$, followed by $X_{-\pi/2}$, and only look at decompositions up to 4 gates, there are 903 possible combinations of gates, resulting in an average of approximately 38 possible decompositions per Clifford. This corresponds to $38^2 = 1444$ possible pairs of decompositions. In compiled randomized benchmarking the optimal compilation from all these possible decomposition pairs is found. (see Appendix G.1 for details on the algorithm used). The result of this compilation is that on average the duration of a pair of Cliffords is equal to 2.925 gates, which is 22% less than alternating randomized benchmarking (3.75 gates per Clifford pair).

9.5.3 5 primitives randomized benchmarking

When exploring different two qubit randomized benchmarking methods, one very interesting property of the Clifford group was discovered. It was found that each of the 24 Cliffords can be decomposed into a subset of 5 primitive gates, while retaining the gate order of the 5 primitives. The 5 primitive gates are:

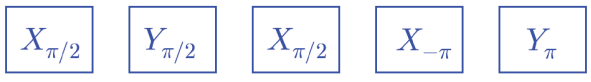


Figure 9.10: The 5 primitive gates, in order of application in time.

Using the fixed set of 5 primitive gates, and by selectively routing the pulses to each of the qubits, arbitrary Cliffords can simultaneously be applied both qubits. The Cliffords applied

to the two qubits need not be equal. In the 5 primitives method the gate sequence stays fixed, and could even be continuously repeated. The markers corresponding to the state of the switches determine the Clifford operations of the qubits.

There is another advantage of the 5 primitive gates, which is that it decreases the amount of cross-driving when one of the two qubits is not being driven. The first three gates composing the 5 primitives are positive $\pi/2$ gates along the X and Y axes. The final two gates are negative π gates along the X and Y axes. Cross-driving corresponds to rotations where the angle is a fraction of the full rotation angle. After one full 5 primitive gates cycle the state of the idle qubit should therefore return close to its initial position. Even if not all 5 gates are applied, the cross-driving effects should still decrease. This is because the cross-driving rotations are small, and in randomized benchmarking a large number of random Cliffords are applied, and so over the course of an entire randomized benchmarking run these cross-driving effects will still partially cancel out.

An even better approach to eliminating cross-driving is to alternate between a cycle of the 5 primitive gates, and a cycle where the 5 primitive gates are inverted, i.e. negative $\pi/2$ rotations and positive π rotations, and switching of the gate order. Since the Clifford set is a group, each Clifford has a unique inverse which is also a Clifford. Inverting a Clifford is equal to inverting the gates of its 5 primitives decomposition (including reversal of the gate order). Since all the inverted gates are elements of the inverted 5 primitive gates, we see that we can indeed also decompose all 24 Cliffords using the inverted 5 primitive gates. Alternating between the 5 primitive gates and the inverted 5 primitive gates is expected to reduce the effects of cross-driving even further, as any deviation from returning to the original position in the first 5 primitives cycle is compensated for by the inverted cycle.

The 5 primitives randomized benchmarking method alternates between the 5 primitives cycle and the inverted 5 primitives cycle. The 5 primitives randomized benchmarking requires 5 gates per Clifford pair, which is considerably more than compiled randomized benchmarking (2.925 gates per Clifford pair) and alternating randomized benchmarking (3.75 gates per Clifford pair).

RB method	Clifford fidelity		gate fidelity	
	top qubit	bottom qubit	top qubit	bottom qubit
Alternating	0.9962	0.9972	0.9990	0.9992
Compiled	0.9970	0.9978	0.9990	0.9992
5 primitives	0.9947	0.9964	0.9990	0.9993

Table 9.1: Corresponding gate fidelities of top and bottom qubit using three randomized benchmarking methods. The gate fidelities are obtained using the corresponding average gates per Clifford.

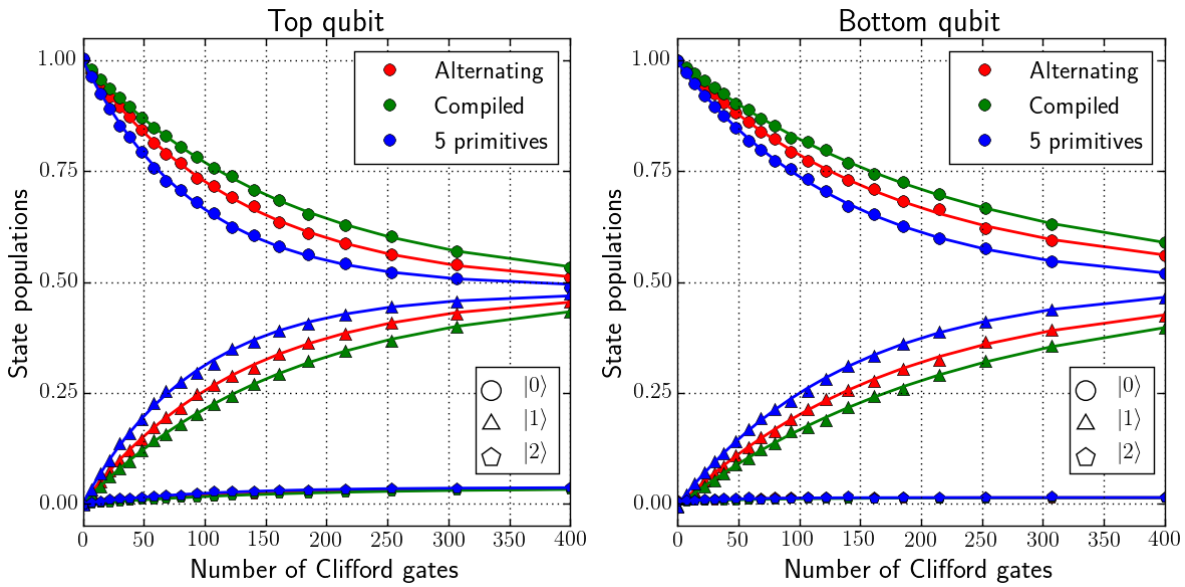


Figure 9.11: State populations using three different two qubit randomized benchmarking sequences. Number of Cliffords correspond to each qubit. A total of 50 seeds was used. The measurements were performed with and without a final pi pulse, from which the populations of the first thee states of the qubit were determined.

9.6 TWO QUBIT RANDOMIZED BENCHMARKING RESULTS

The three different two qubit randomized benchmarking methods have been performed for a total of 50 seeds. Up to $n = 400$ Cliffords per qubit were used, and the pulse length was kept at 16 ns with a 4 ns buffer between pulses. By performing the measurements both with and without a final pi pulse, the populations of the first three states of the qubit have been determined (see Appendix F.3). The results are shown in Figure 9.11. For all three methods the bottom qubit performs considerably better than the top qubit, which is expected considering the decoherence times of the qubits. For both qubits compiled randomized benchmarking has the lowest error per Clifford, after which alternating randomized benchmarking, while 5 primitives randomized benchmarking has the highest error per Clifford. This is in agreement with the average gates per Clifford, which is smallest for compiled Randomized benchmarking, and largest for 5 primitives randomized benchmarking. Nevertheless we see that for both qubits all three methods result in fidelities per Clifford exceeding 99%. The fidelity per Clifford can be converted to a corresponding fidelity per gate using the average gates per Cliffords of the three methods. The corresponding gate fidelities are shown in Table 9.1. As can be seen the gate fidelities are nearly identical, and furthermore are equal to the single qubit randomized benchmarking gate fidelities. This comparison is not entirely fair, as idle gates (when only the other qubit is being pulsed) are also counted as gates. Nevertheless, it can be concluded that there is no significant decrease in gate fidelity when using selective broadcasting. These results show that simultaneous control of two qubits can be achieved with an error rate below the single qubit suface code fault tolerant threshold.

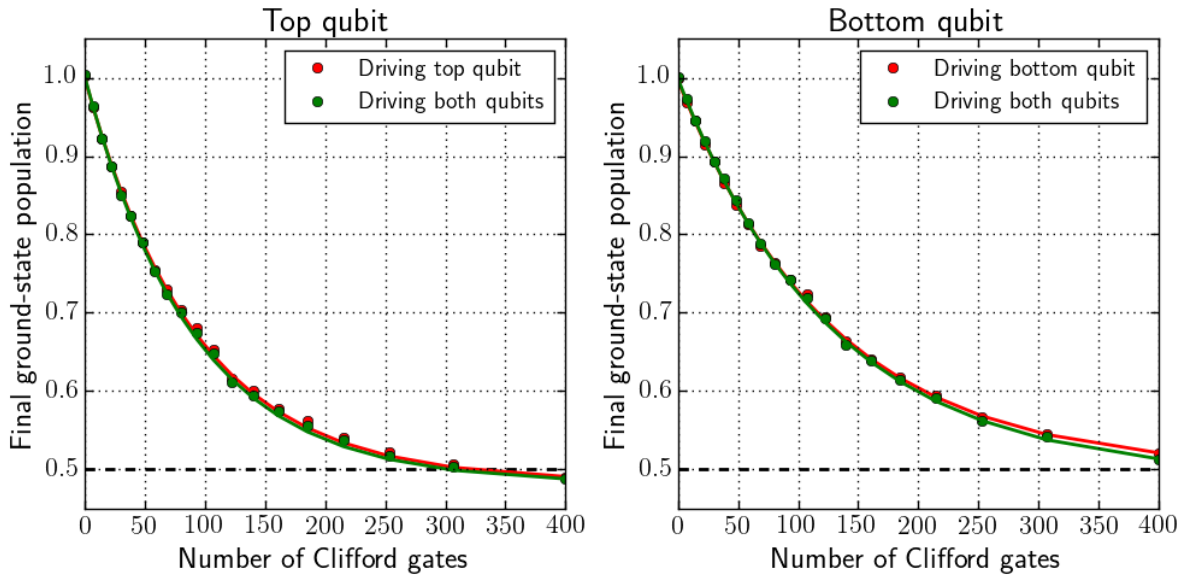


Figure 9.12: 5 primitives randomized benchmarking results when driving a single qubit versus driving both qubits simultaneously. A total of 50 different seeds were used.

In Figure 9.12 the effect of driving a single qubit versus driving both qubits simultaneously is shown using 5 primitives randomized benchmarking. As can be seen there is no noticeable change in qubit performance.

In Figure 9.13 we can see the effect of cross-driving during two qubit randomized benchmarking. In this case the switches corresponding to one qubit are continuously closed, while the other qubit is being driven. As can be seen the bottom qubit experiences much more cross-driving than the top qubit. We see, however, that this cross-driving is greatly reduced when using the 5 primitives method. This is due to the fact that the pulses in the 5 primitives method are chosen such that the cross-driving effects for idle qubits are reduced (see Section 9.5.3).

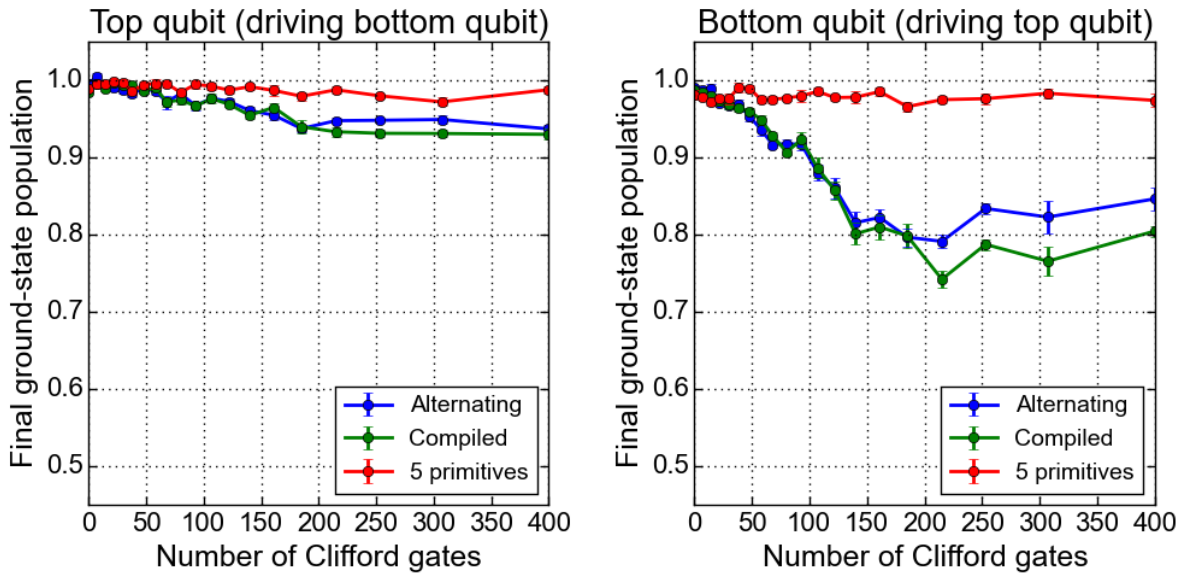


Figure 9.13: Excitation in qubit when only the other qubit is driven during randomized benchmarking. This effect is mostly due to cross-driving and is considerably stronger for the top qubit than for the bottom qubit. A total of 5 seeds was used

9.7 SCALING OF MULTI QUBIT RANDOMIZED BENCHMARKING

So far it has been found that for two qubits compiled randomized benchmarking clearly performs best, while 5 primitives method performs worst. The performance is directly related to the average number of gates per Clifford set. Figure 9.14 shows how the average number of gates per Clifford scale a function of the number of qubits for all three randomized benchmarking methods. At two qubits the 5 primitives method performs worst of all three methods. However, at three qubits the 5 primitives method already outperforms alternating randomized benchmarking. This is because the 5 primitives method has 5 gates per Clifford, irrespective of the number of qubits, whereas alternating randomized benchmarking scales linearly with the number of qubits.

Compiled randomized benchmarking always performs the best, as it per definition finds the least amount of gates required to perform a set of Cliffords on the qubits. However, compiled randomized benchmarking has a clear disadvantage, which is that the required time for finding the optimal compilation scales exponentially with the number of qubits. Each Clifford on average has 38 different decompositions, and so for five qubits this would already result in $38^5 \approx 7.9 * 10^7$ different combinations for each individual set of Cliffords. Furthermore, finding the average gates per Clifford requires knowledge of the gates per Clifford for all different combinations of Cliffords, and so for five qubits this would result in $6.3 * 10^{14}$ different combinations of decompositions. Using several optimization methods, which are discussed in Appendix G.1.2, the average gates per Clifford has been determined exactly for up to five qubits, and approximated up to ten qubits using random sampling. Nevertheless it is still computationally intensive to find the optimal gate compilation, even

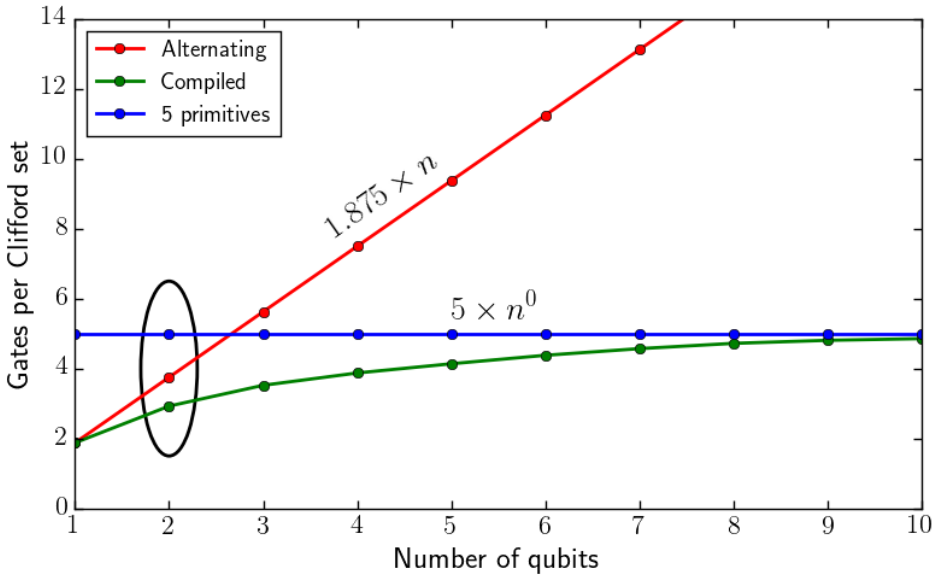


Figure 9.14: Comparison of the average gates per Clifford set versus the number of qubits using the three different multi qubit randomized benchmarking methods. The oval indicates the gates per Clifford at two qubits

for a modest amount of qubits.

In contrast to the compiled randomized benchmarking method, the 5 primitives randomized benchmarking method is computationally very easy to perform. Once a marker lookup table has been created (see Appendix F.1), determining the 5 primitives decomposition of an arbitrary number of Cliffords is simply a matter of looking up the corresponding marker sequences. Furthermore, at 5 qubits the difference in average gates per Clifford between the compiled randomized benchmarking and the 5 primitives randomized benchmarking is already less than 1, and so unless the number of qubits is small, the 5 primitives randomized benchmarking is the best and by far easiest of the three randomized benchmarking methods.

Chapter 10

Conclusions and outlook

When the top and bottom qubit are tuned to the same frequency, cross-coupling and cross-driving effects do become noticeable. This could potentially be a problem, as it results in correlated errors. The fact that the cross-driving effects in the Muxmon1 experiment are stronger than in the Muxmon0 device suggest that they dependent on the amount of components separating same-frequency qubits. The same holds for cross-coupling effects. Therefore, if it turns out that in the surface code architecture these effects must be reduced, a modified frequency re-use structure could be implemented. For instance, by using a total of eight frequencies instead of four, qubits sharing the same frequency can be separated by more components, thereby reducing these adverse effects.

When the top and bottom qubit share the same frequency, it has been shown that they can be simultaneously controlled by a single generator in combination with the Duplexer. Through randomized benchmarking it has been shown that for both qubits the gate fidelity reaches 99.9 per cent, thereby meeting the surface code fault-tolerant threshold. The gate fidelity is within a factor of two of limit imposed by its relaxation time T_1, indicating that the qubit performance is in fact decoherence limited.

It has furthermore been shown that there is no change in performance when driving a single qubit versus driving both qubits simultaneously. It is interesting to note that the gate fidelities exceed the maximum one would expect considering the cross-driving measured. One possible explanation is that the cross-driving measurements were performed at high power, and since the transmon is a nonlinear system, it is possible that the cross-driving is lower at low drive powers.

Furthermore, it has been shown individual control of both qubits simultaneously is possible through selective broadcasting. This has been demonstrated using multi-qubit randomized benchmarking, where each qubit has a different seed. Three different multi-qubit randomized benchmarking schemes have been implemented, namely alternating, compiled, and 5 primitives randomized benchmarking. It has been shown for all three methods the Clifford fidelity exceeds 99 per cent, thereby also meeting the surface code fault-tolerant threshold. Converting the Clifford fidelities to gate fidelities using their average gates per Clifford it has shown to be identical to the gate fidelities obtained by single qubit randomized benchmarking, showing that there is no significant decrease in performance when using selective broadcasting.

Of the three multi-qubit randomized benchmarking methods, alternating randomized benchmarking clearly scales the worst with number of qubits. Compiled randomized benchmarking per definition always scales the best. However, the complexity of finding the optimal gate compilation scales exponentially, making it a highly non-trivial challenge. The 5 primitives randomized benchmarking, remains fixed at 5 gates per Clifford, irrespective of the number of qubits. In contrast to the compiled randomized benchmarking however, finding the Clifford decomposition using the 5 primitives method can be done using a simple lookup table. furthermore, the average gates per Clifford using compiled randomized benchmarking converges to 5 quite rapidly, and so unless the number of qubits is small, the gain using compiled randomized benchmarking is small.

When randomized benchmarking is only performed on one of the qubits, while the other at the same frequency remains idle, cross-driving effects have been observed. However, it has been shown in the 5 primitives method that by cleverly choosing pulses, cross-driving effects when the other qubit is idle can largely be corrected.

During the randomized benchmarking sequences leakage to the second excited-state has been observed. The leakage rate is found to be worse for the top qubit than for the bottom qubit. The DRAG parameter has been calibrated to minimize the phase error. However, it is not necessarily the case that this also corresponds to the optimum for minimizing leakage. Changing the DRAG parameter could therefore potentially reduce leakage, although this would result in more phase error during gates. Additionally, leakage could be reduced by resorting to longer pulses.

Even though the Duplexer already proves to be very useful in a measurement set-up, it is only the first step in a path aimed at reducing the amount of intstruments. The next generation Duplexer is already being made, which will be a true multiplexer having a variable number of outputs. This will be a huge step in the road to scaling up.

There is one additional advantage of the 5 primitives method, which is that the 5 primitive gate sequence remains fixed. Therefore it is not needed to generate qubit pulses on demand, but only to control the qubit switch. This could potentially have an application in a quantum processor, where a central pulse generator continuously repeats the same 5 pulses. By selectively directing the pulses to each of the qubits using a switch matrix, each qubit could have access to any given Clifford at any time. The consequence is that for single-qubit Clifford gates the pulses do not need to be generated on demand, but can simply be continuously looped. The timing of the switches is the only thing that needs to be directed, which is a far easier challenge. If the 5 primitive gates are combined with a sixth gate, such as the T-gate, this would even result in any universal single-qubit gates being accessible to any qubit, simply through control of its corresponding switch. This technique could even possibly be extended to two-qubit gates, where the flux pulses could be similarly controlled. This would mean that a universal set of quantum gates could be generated for qubits by only controlling the corresponding switches. It might even be extended to qubit readout, as the readout tonescould be directed via switching as well.

Having continuously streaming pulses, which are directed to the qubits via selective broadcasting could have ... consequences, as this would shift the challenge from generating pulses when required to only controlling the switches. This would have important conse-

quences. First of all the feedback response time could improve significantly. Furthermore the amount of data required to be sent into the fridge could be reduced significantly. However, the most important consequence is probably that the amount of instruments required would be reduced significantly, potentially even reaching a constant as the number of qubits grow. This might prove to be an important step in the pathway to building a quantum computer.

Appendices

Appendix A

Noise characterization

When performing measurements one is faced with the reality that no component is ideal. When a signal passes through the different parts of a set-up, noise is constantly being added. Noise is the term given for all the random fluctuations that are added to the signal. These fluctuations are the cumulative result of several noise source contributions.

Thermal noise is one of the most common sources of noise. It is the result of the random thermal fluctuations of electrons. It is an example of frequency-independent noise, also known as white noise. Noise sources can also be frequency-dependent, such as TLS, as discussed in Chapter 1. This is an example of $1/f$-noise: the amount of noise added increases with decreasing frequency. In fact, truly frequency-independent noise does not exist, as even white noise has been observed to decrease at extremely high frequencies ($\sim 10^{15}$ Hz). At these frequencies a quantum correction needs to be added [32, p50].

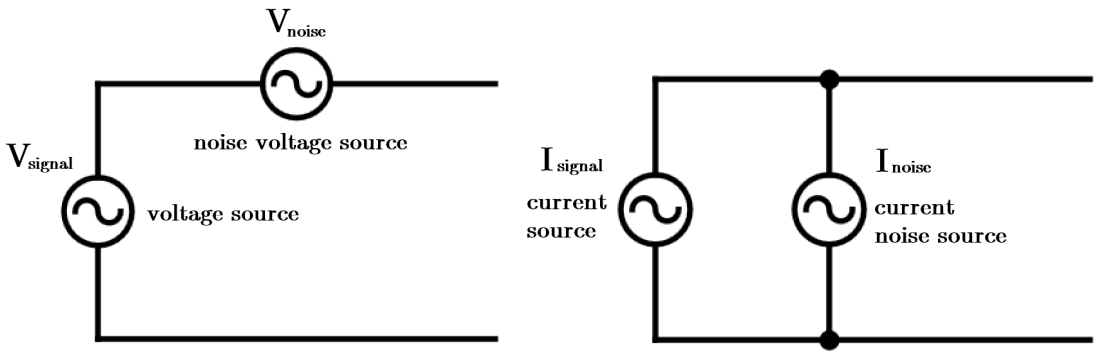
When performing measurements one important question to ask is how much noise is being contributed to the signal. In this chapter a model is presented for the general set-up used for measuring superconducting resonators and qubits. Using this model it is possible to characterize the amount of noise by determining its associated noise temperature. Finally, this model is applied to the set-up used to characterize the resonators presented in Chapter 1.

A.1 Characterizing noise

A.1.1 Circuit representations

There are two circuit representations in which we can depict a system with a noise contribution: The Thévenin representation, and the Norton representation. In the Thévenin representation we can model the system as a voltage source, and the noise added to the system is a noise voltage source connected in series. In the Norton representation the system is a current source and the noise added is a noise current source connected in parallel to the current source. These two representations are identical and can be converted to each other. In this section we will adopt the Thevénin representation, and so the signal will be a voltage source combined in series with a noise voltage source.

Assuming the signal to be at a fixed frequency ω and amplitude A, the combined voltage



(a) Thévenin representation **(b)** Norton representation

Figure A.1: Two equivalent representations of a system containing noise. Panel **(a)** shows the Thévenin representation, in which a noiseless voltage source is connected in series with a noise voltage source. Panel **(b)** shows the Norton representation, in which a noiseless current source is connected in parallel with a noise current source.

$v(t)$ is then given by:

$$\mathrm{v}(t) = \mathrm{v}_{\mathrm{signal}}(t) + \mathrm{v}_{\mathrm{noise}}(t) = A\cos\omega t + \mathrm{v}_{\mathrm{n}}(t) \tag{A.1}$$

Note that the mean value of the noise voltage $\overline{\mathrm{v_n}}$ is equal to zero. The amount of noise can be quantified by the root-mean square noise voltage $\mathrm{v}_{\mathrm{n}}^{\mathrm{rms}}$:

$$\mathrm{v}_{\mathrm{n}}^{\mathrm{rms}} = \sqrt{\overline{\mathrm{v}^2} - \overline{\mathrm{v}}^2} = \sqrt{\overline{\mathrm{v}_{\mathrm{n}}^2}} \tag{A.2}$$

A.1.2 Noise power spectral density

One way of quantifying the noise of the system is through the noise power spectral density $S(f)$, which is the distribution of noise power per unit bandwidth as a function of frequency. For the Thévenin representation, the noise spectral density is defined in terms of voltage. When the only noise in the system is white noise, the spectral density is independent of frequency. It is then given by:

$$S = \frac{\overline{\mathrm{v}_{\mathrm{n}}^2}}{\Delta f} \qquad [\mathrm{V}^2/\mathrm{Hz}] \tag{A.3}$$

In this equation Δf is the noise bandwidth. This is the bandwidth over which the noise is measured.

A.2 THE MODEL

As shown in the schematic in Figure A.2, we can model our set-up as a combination of four elements:

1. A noise source.
2. An amplifier to amplify the weak signal exiting the fridge.
3. A mixer to downconvert the signal.
4. A low-pass filter to remove unwanted high-frequency signal.

A.2.1 Noise source

Using the Thévenin model, we can approximate the components up to the first amplifier in the fridge as a voltage source, with a noise voltage source connected in series.

We can include the noise added by the amplifier in the noise voltage source, in which case we assume the amplifier to be ideal. Furthermore, we assume the signal to be amplified sufficiently, such that the mixer and low-pass filter add a negligible amount of noise. We also ignore effect such as mixer leakage. With these assumptions all of the noise is originated from the noise voltage source.

We can associate an effective noise temperature to the noise voltage source. The noise temperature is defined as the temperature at which a resistor would produce an equal amount of noise. Note that, since we are comparing the system to a resistor, the noise needs to have an (approximately) white spectrum.

According to Nyquist's theorem [32, p47], if the system experiences white noise, and is in thermal equilibrium, the root-mean square noise voltage v_n^{rms} is given by:

$$v_n^{rms} = \sqrt{\overline{v_n^2}} = 4k_B T R \Delta f \tag{A.4}$$

In this equation Δf is the bandwidth over which the noise is integrated, k_B is the Boltzmann constant, T is the noise temperature of the noise source, and R is the impedance of

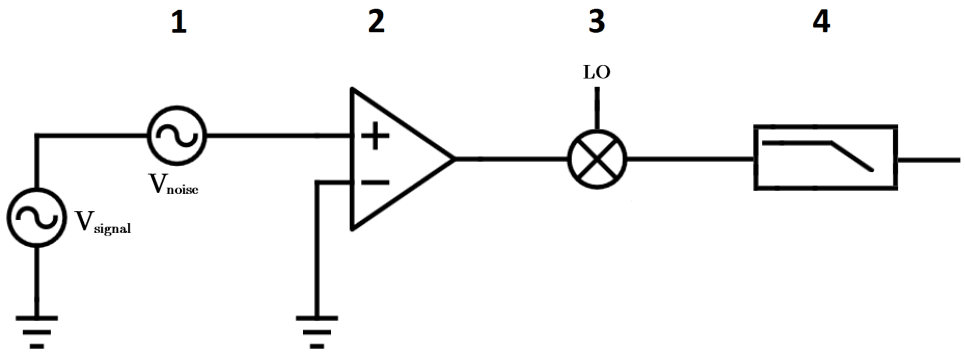


Figure A.2: Schematic representation of the measurement set-up including a noise source.

the system. We see that the noise added depends linearly on the bandwidth over which is integrated.

Combining Equations A.3 and A.4, the noise power spectral density $S_{\mathrm{v_n}}$ can be rewritten as:

$$S = \frac{\overline{\mathrm{v_n^2}}}{\Delta f} = 4k_B TR \tag{A.5}$$

A.2.2 Amplification

During the amplification stage both the signal and the noise is amplified by the same amount. This amount of amplification is determined by the gain G, which is defined as the ratio between the output voltage and the input voltage:

$$G = \frac{\mathrm{v_{out}}}{\mathrm{v_{in}}} \tag{A.6}$$

According to the maximum power transfer theorem, the maximum power transfer between a source and load occurs when the impedances of source and load are matched, in which case half of the power is transferred. From this it follows that in the amplification process half of the signal is dissipated. However, as G is defined as the ratio between $\mathrm{v_{out}}$ and $\mathrm{v_{in}}$ (Equation A.6), the factor $\frac{1}{2}$ is included in G.

During amplification not only the signal is amplified with gain G: the noise is amplified by the same amount. Defining the noise power spectral density before amplification as S^{in}, and after amplification as S^{out}, the following relation holds:

$$S^{\mathrm{out}} = G^2\, S^{\mathrm{in}} = G^2 4k_B TR \tag{A.7}$$

Note that in Equation A.7, the gain G is squared. This is due to the fact that the noise power spectral density depends quadratically on the root-mean square noise voltage (Equation A.3).

In our actual set-up the amplification occurs in multiple stages. Aside from amplifying the signal and its noise, at each stage additional noise, originating from the amplifier itself, is added as well. This added noise is then also amplified in the next amplification stage. Therefore it is always best to have the amplifier with highest gain and lowest noise temperature as the first amplifier in the chain. For more information see Friis formula [32, p103].

A.2.3 Downconversion

After amplification the frequency ω of the signal is still in the GHz range. In homodyne or heterodyne detection the signal is downconverted to DC (homodyne) or to a lower frequency (heterodyne), such that it can be measured more easily. To downconvert the signal, it is mixed in a mixer with a local oscillator (LO) signal having the same frequency ω (homodyne) or a slightly higher frequency $\omega + \Delta\omega$ (heterodyne). The mixer effectively multiplies the two signals. If the signal exiting the amplifier is given by $\mathrm{v}(t) = A\cos\omega t$, then, ignoring a possible phase difference, the signal at the output of the mixer is given by:

$$\begin{aligned} \mathrm{v}(t)\cdot\cos\omega t &= A\cos\omega t\cdot\cos\left(\omega+\Delta\omega\right)t \\ &= \frac{1}{2}A\left[\cos\left(2\omega+\Delta\omega\right)t+\cos\Delta\omega t\right] \end{aligned} \tag{A.8}$$

As can be seen from Equation A.8, the output signal contains both the sum and the difference of the two signals. However, as the sum of both frequencies is in the GHz range, it can be filtered out using a low-pass filter, leaving only the downconverted signal, which is the result of the difference between the two frequencies. Note that the amplitude of the signal is reduced by a factor two. The noise amplitude is also reduced by a factor 2. Furthermore, in the case of a homodyne set-up, the difference signal is simply a DC signal ($\Delta\omega = 0$), while in the case of heterodyne the signal still contains a slow frequency $\Delta\omega$. For simplification we assume our set-up to be a homodyne set-up, although the result is similar in the case of a heterodyne set-up.

A.2.4 Low-pass filtering

In the case of homodyne detection the signal at the frequency of interest is downconverted to DC. However, the signal at other frequencies have not disappeared; in the mixer these have also shifted in frequency. Since the signal of interest is at DC, a low-pass filter can be used to filter out signal above a certain frequency.

The frequency above which a low-pass filter will filter out the signal is defined by its cut-off frequency f_c. The cut-off frequency f_c is the frequency at which the signal is attenuated by 3 dB. For first-order low-pass filters the noise bandwidth Δf is related to the filter cut-off frequency f_c by [32, p81]:

$$\Delta f = \frac{\pi}{2} f_c \tag{A.9}$$

A.3 NOISE TEMPERATURE

In the previous sections the influence of each of the components on the signal and on the noise has been analyzed. Using this information it is possible to determine the signal-to-noise ratio (SNR), which is the ratio between the average power of the signal and the average power of the noise. The SNR is a measure for how well a signal can be separated from the noise, and is given by:

$$\begin{aligned} \mathrm{SNR} = \frac{\overline{\mathrm{v_{out}}}^2}{\overline{\mathrm{v_n^2}}} &= \frac{1/4\ G^2\ \overline{\mathrm{v_{in}}}^2}{\mathrm{S}_{\mathrm{v_n}}^{\mathrm{out}}\ \Delta f} \\ &= \frac{1/4\ G^2\ \overline{\mathrm{v_{in}}}^2}{G^2\ \mathrm{S}_{\mathrm{v_n}}^{\mathrm{in}}\ \frac{\pi}{2}\ f_c} \\ &= \frac{\overline{\mathrm{v_{in}}}^2}{2\,\pi\ k_B\ T\ R\ f_c} \end{aligned} \tag{A.10}$$

Note that the factor 1/4 is because the amplitude is lowered by a factor of 2 due to mixing. Equation A.10 can be rewritten such that we have an expression for the noise temperature of the system:

$$T = \frac{\overline{v_{in}}^2}{2\,\pi\,k_B\,R\,f_c\,\mathrm{SNR}} \tag{A.11}$$

A.4 RESULTS

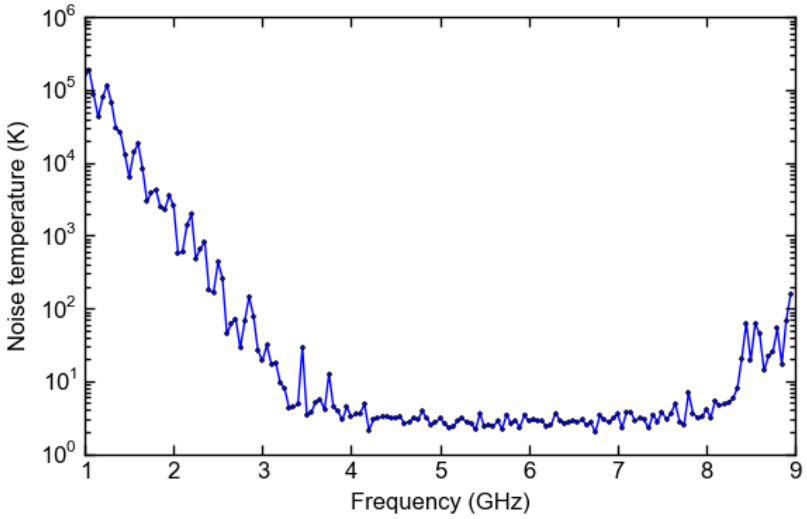


Figure A.3: Noise temperature versus frequency. The noise temperature has been calculated for 160 frequencies in the range 1–9 GHz. For each frequency 2001 points were measured, from which the signal, noise, SNR, and noise temperature was determined. Measurements were performed at an input power of −113 dBm and an IF bandwidth of Δf =300 Hz.

Using the Rhode & Schwarz ZVM vector network analyzer, The transmission has been measured for 160 equidistant frequencies in the range 1–9 GHz. For each frequency a total of 2001 points was measured with an IF bandwidth Δf = 300 Hz. From these measurements the signal-to-noise ratio has been determined for each frequency. With knowledge of the SNR, the noise temperature has then been determined as a function of frequency using Equation A.11 . The result is shown in Figure A.3.

From Figure A.3 it is clear that the noise temperature is highly temperature-dependent. In the frequency range 4–8 GHz the noise temperature is quite low, never reaching above 10 K. This is exactly the bandwidth of the cryogenic low-noise amplifier by Low Noise Factory, which is the first amplifier in the amplification chain. From the specifications of the amplifier, the noise temperature of the amplifier has been calculated at an ambient temperature of 8 K, and equals roughly 4 K for the entire bandwidth. Comparing the amplifier specifications with Figure A.3, it is likely that in the frequency range 4–8 GHz the first amplifier is the component contributing most to the total noise temperature.

Outside the 4–8 GHz frequency band, however, the noise temperature rapidly increases. This is partly due to the frequency lying outside of the bandwidth of the amplifier, in which case the amplification will be lower. However, this is not an adequate explanation for the fact that the noise temperature increases to several hundred thousand Kelvin. The reason for this unrealistic noise temperature is that in our model we did not take into account the noise added by components after the amplifier. While the gain of the amplifier decreases outside its bandwidth, the components after the amplification will still add the same amount of noise. When the gain decreases by a significant amount, the relative contribution of these post-amplification noise sources will increase. Furthermore the assumption that the noise spectrum is white is no longer correct at low frequencies, where $1/f$ noise starts to contribute.

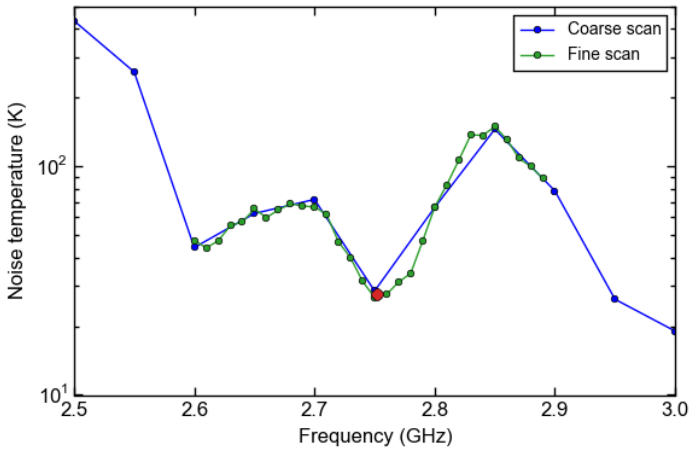


Figure A.4: Detailed scan of noise temperature versus frequency in the range 2.6–2.9 GHz. The resonator with f_0 =2.75 GHz (red dot) seems to reside at a local minimum of the noise temperature.

From Figure A.3 it can be seen that the noise temperature can vary by a large amount between consecutive points. To determine whether this variation is due to a large uncertainty, or due to the noise temperature actually fluctuating strongly with varying frequency, a more detailed scan has been performed in the frequency region 2.6–2.9 GHz, in which one resonator has a resonance frequency. The result is shown in Figure A.4. As the curve of the detailed scan follows the curve of the coarse scan pretty closely, it can be concluded that the noise temperature of the system in fact fluctuates quite strongly with varying frequency.

Another point of interest is that the resonance frequency of the resonator lies near the local minimum of the noise temperature in that region. This is quite a stroke of luck, as a slightly higher or lower frequency would have resulted in a much higher noise temperature.

A.5 Conclusion and future work

The noise temperature gives us an estimate of the noise added to the system. It has been shown that the noise temperature can fluctuate strongly with varying frequency. In the model used to estimate the noise temperature it has been shown that outside the bandwidth of the

amplifier the model breaks down. At this point the noise added by the components after the amplification, and indeed even the amplifiers themselves, needs to be taken into account to obtain accurate estimates of the noise temperature.

However, even outside of the bandwidth of the amplifier, there are frequency regions in which the noise temperature may still be acceptable. It is therefore a good idea to initially perform measurements of the noise temperature of the set-up. This will give an indication of the signal-to-noise ratio, from which accurate estimates can be made as to what the amount of measurement time is needed to obtain a desired SNR.

The noise temperature measurements were performed using the Rhode & Schwarz ZVM vector network analyzer. It would be interesting to see how other measurement set-ups would compare to the vector network analyzer. One interesting candidate would be a heterodyne detector. However, as the vector network analyzer can also measure phase, a fair comparison would also require the heterodyne detector to be able to measure the phase. This heterodyne detector is currently being set up, and will hopefully soon yield interesting results.

Aside from only comparing the noise temperature, other properties are also important when comparing two set-ups. One of these is the duty cycle, which is the percentage of time acually spent measuring. For the vector network analyser the duty cycle seems to be around 50%, provided that a single measurement sweep takes at least a few seconds. Other set-ups may therefore offer an improvement in the duty cycle. Furthermore, properties such as phase stability and uncertainty would also be interesting to compare.

Another interesting measurement would be to see if the noise temperature as a function of frequency remains the same in future cooldowns, and for different samples.

Appendix B

Duplexer isolation

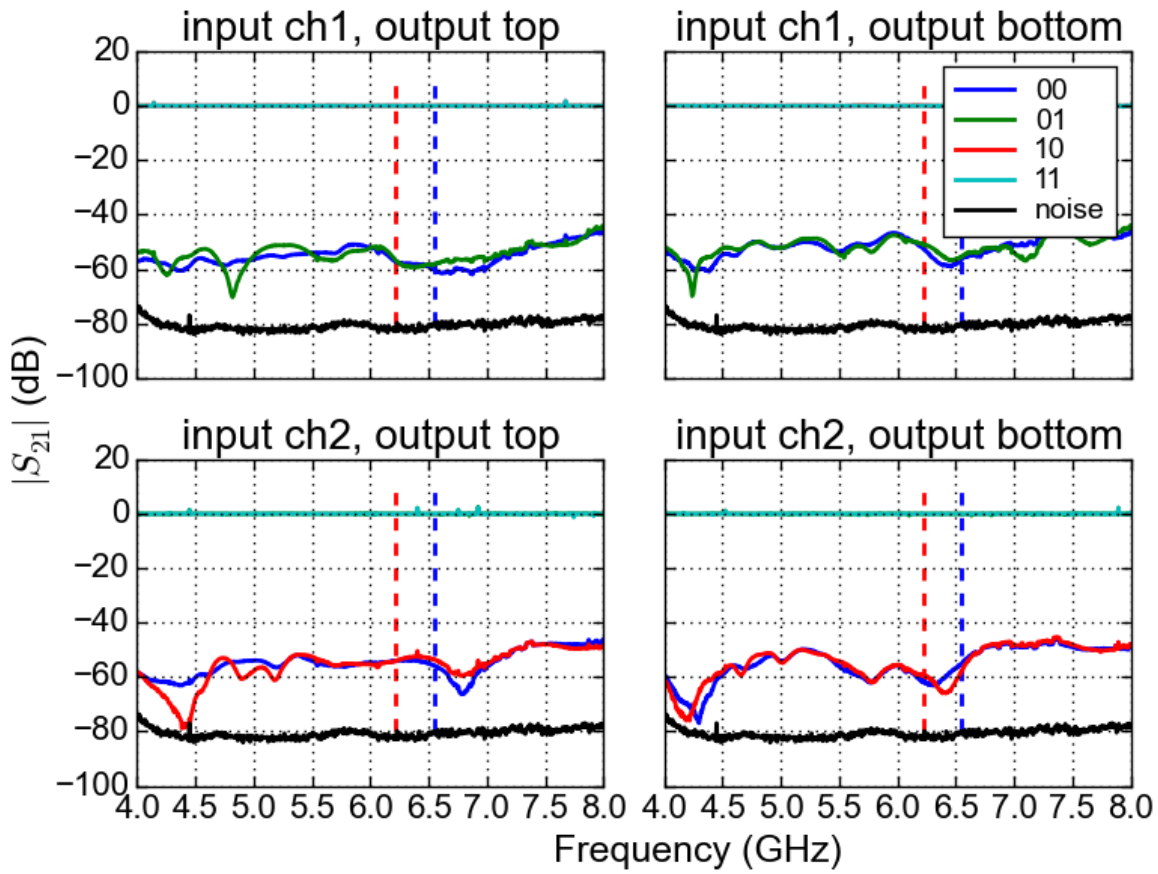


Figure B.1: Isolation of the Duplexer. Legend indices correspond to switch state(1 means open, 0 means closed, lsb corresponds to bottom drive line, msb to top drive line). Transmission is measured relative to transmission when all channels are open (11). The black line corresponds to the measured noise floor. The red dashed line indicates the frequency of the top and bottom qubit, and the blue dashed line to the ancilla qubit, during the Muxmon experiment.

The isolation of the Duplexer switches has been measured. During the measurements a signal was sent through input channel 1 and 2, corresponding to the signal of the main Gaussian pulse, and derivative pulse, respectively. For each input channel the transmission to each of the two output channels has been measured. These output channels are connected to the drive lines of the top and bottom qubit during the Muxmon experiment. For each input-output port combination the transmission has been measured for different switch configurations, which were controlled using the nanosecond-switches. Four different switch combinations were used, corresponding the the four distinct possibilities of the signal being directed to each of the two output channels.

The results of the isolation measurements are shown in Figure B.1. As can be seen in all four input-output combinations the isolation is around 50 dBm at the frequency of top and bottom qubit. Furthermore, the isolation does not seem to depend strongly on whether the switch to the other output port is open or closed.

Appendix C

Chip characterization

C.1 QUBIT COHERENCE TIMES VERSUS FREQUENCY

C.2 MUXMON1 DECOHERENCE TIMES

C.3 MUXMON1 CROSS-DRIVING

qubit	cross-driving (%)	
	top-ancilla drive line	bottom-ancilla drive line
Top	100	5.22
Ancilla	75	62.5
Bottom	1.05	100

C.4 MUXMON0 RESONATOR BUSES

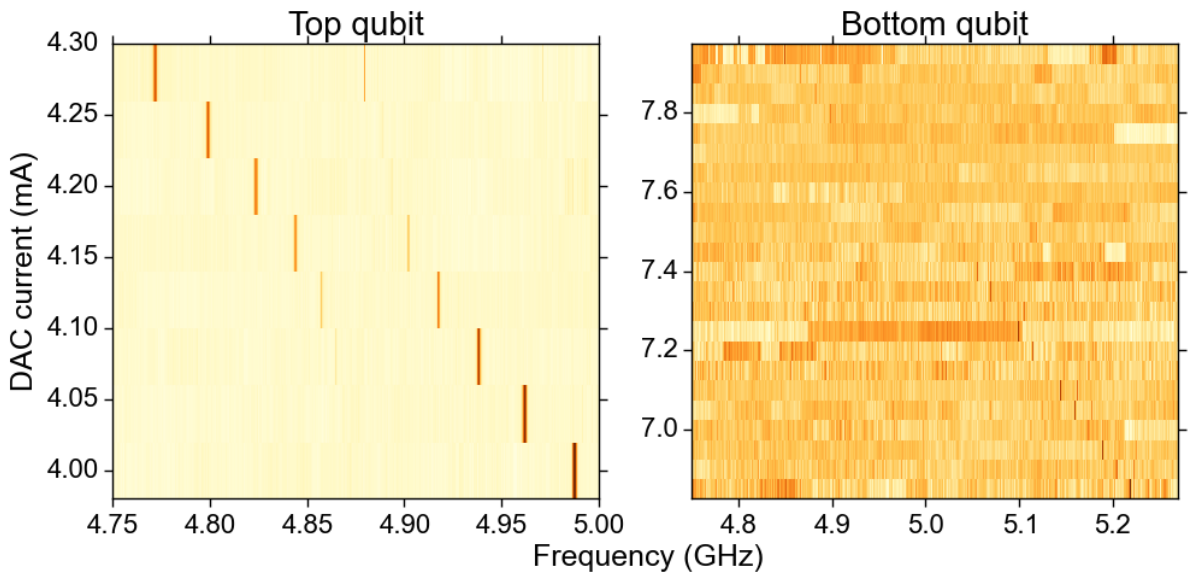


Figure C.1: Normalized transmission showing the resonator buses of the top and bottom qubit, at frequencies 4.88 GHz and 4.97 GHz respectively.

The top and bottom qubit are each coupled to a resonator bus. When the frequency of a qubit approaches that of the bus, they experience an avoided crossing. In Figure C.1 these avoided crossing are shown for the top and bottom qubit. For the top qubit the frequency of the resonator bus is found to be 4.88 GHz. For the bottom qubit the signal was considerably worse, but the frequency of the resonator bus is found to be roughly 4.97 GHz. For the top qubit the coupling strength g between the qubit and the bus can be extracted. It is equal to half the minimum distance, and approximately equal to 27 MHz.

Appendix D

Additional notes

D.1 QUBIT CHARACTERIZATION

D.1.1 Finding the qubit sweet-spot using a one-dimensional scan

As explained in Section 7.2.3, the sweet-spot of a qubit can be found by performing a series of resonator scans while varying the DAC current. There is however, a faster approach for finding the qubit sweet-spot, at the cost of providing less information. This is done by choosing a fixed frequency close to the resonator's frequency ω_r (preferrably slightly below, where the transmission slope is steepest). By measuring the amount of transmission as the DAC voltage is being varied, one obtains essentially a line-cut of the 2D scan. The idea this measurement is that if the qubit frequency f_q decreases, so does resonator frequency, resulting in a decrease in transmission (closer to ω_r). Likewise, if the qubit frequency ω_q increases, so does the resonator frequency, resulting in an increase in transmission (further away from ω_r).

At the qubit's sweet-spot, the resonator's frequency ω_r is at a maximum, and so the transmission should also be at a maximum. Furthermore, because the qubit frequency is symmetric with respect to its sweet-spot, so should the transmission. The sweet-spot can therefore be determined by finding the symmetric point in the transmission. In Figure D.1 one such linecut is shown. The transmission is clearly symmetric, and the symmetric point is equal to the qubit sweet-spot.

If the resonator's frequency ω_r shifts by a large amount in the course of this measurement, it becomes harder to determine where the sweet-spot is (although even then often it can still be discerned). Nevertheless, this method is considerably faster than performing a full two-dimensional scan of frequency versus DAC voltage, and in most cases does provide sufficient information to determine the sweet-spot.

D.1.1.1 Spectroscopy

- If the deviation in transmission becomes less due to more detuning, increasing the power can also increase the contrast.

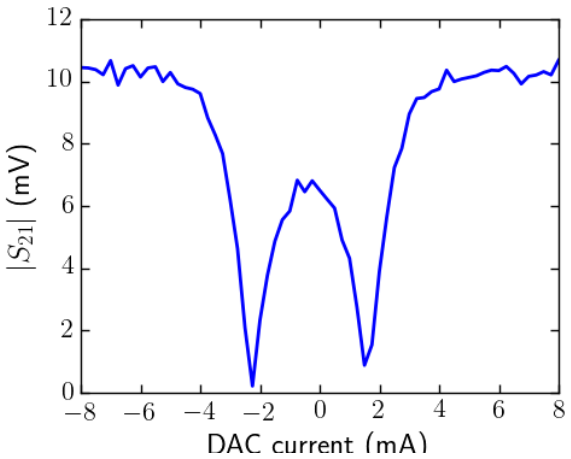


Figure D.1: Fixed frequency transmission versus DAC current. The frequency is chosen to be slightly below the resonator frequency found when the DAC current is set to zero. The symmetry point in the linecut corresponds to the sweet-spot of the ancilla qubit.

D.1.1.2 Flux matrix

After determining the flux matrix $\boldsymbol{F}$, there will still be some small remaining cross-coupling, which depends on the accuracy of the measurements. The process of creating a flux matrix can then be repeated, but instead of using DAC voltages as the varying parameters to construct matrix $\boldsymbol{M}$, the virtual fluxes should be used. Furthermore, as the cross-coupling is small compared to before, the flux range can be much greater, such that small slopes can be accurately measured. The resulting flux matrix $\boldsymbol{F}_2$ can then simply be multiplied with the first flux matrix $\boldsymbol{F}$ to obtain a more accurate final flux matrix.

Appendix E

AllXY pulse sequence

number	ideal $\langle z \rangle$	First pulse	Second pulse
1	1	I	I
2	1	X_π	X_π
3	1	Y_π	Y_π
4	1	X_π	Y_π
5	1	Y_π	X_π
6	0	$X_{\pi/2}$	I
7	0	$Y_{\pi/2}$	I
8	0	$X_{\pi/2}$	$Y_{\pi/2}$
9	0	$Y_{\pi/2}$	$X_{\pi/2}$
10	0	$X_{\pi/2}$	Y_π
11	0	$Y_{\pi/2}$	X_π
12	0	Y_π	$Y_{\pi/2}$
13	0	X_π	$X_{\pi/2}$
14	0	$X_{\pi/2}$	X_π
15	0	X_π	$X_{\pi/2}$
16	0	$Y_{\pi/2}$	Y_π
17	0	Y_π	$Y_{\pi/2}$
18	-1	X_π	I
19	-1	Y_π	I
20	-1	$X_{\pi/2}$	$X_{\pi/2}$
21	-1	$Y_{\pi/2}$	$Y_{\pi/2}$

Table E.1: The 21 pulses that comprises the AllXY pulse sequence.

In Table **??** the 21 pulse combinations are shown that comprise the AllXY pulse sequence. The AllXY measurement is able to diagnose specific sources contributing to gate errors (see 8.2.5). For a more detailed analysis on the AllXY see Reed's thesis [26].

Appendix F

Randomized benchmarking

F.1 CLIFFORD GATE DECOMPOSITION

Clifford nbr	gate decomposition	5 primitives decomposition				
		$X_{\pi/2}$	$Y_{\pi/2}$	$X_{\pi/2}$	$X_{-\pi}$	$Y_{-\pi}$
1	I	0	0	0	0	0
2	$Y_{\pi/2} - X_{\pi/2}$	0	1	1	0	0
3	$X_{-\pi/2} - Y_{-\pi/2}$	1	1	0	1	0
4	X_{π}	0	0	0	1	0
5	$Y_{-\pi/2} - X_{-\pi/2}$	0	1	1	0	1
6	$X_{\pi/2} - Y_{-\pi/2}$	1	1	0	0	1
7	Y_{π}	0	0	0	0	1
8	$Y_{-\pi/2} - X_{\pi/2}$	0	1	1	1	1
9	$X_{\pi/2} - Y_{\pi/2}$	1	1	0	0	0
10	$X_{\pi} - Y_{\pi}$	0	0	0	1	1
11	$Y_{\pi/2} - X_{-\pi/2}$	0	1	1	1	0
12	$X_{-\pi/2} - Y_{\pi/2}$	1	1	0	1	1
13	$Y_{\pi/2} - X_{\pi}$	0	1	0	1	0
14	$X_{-\pi/2}$	0	0	1	1	0
15	$X_{\pi/2} - Y_{-\pi/2} - X_{-\pi/2}$	1	1	1	0	1
16	$Y_{-\pi/2}$	0	1	0	0	1
17	$X_{\pi/2}$	0	0	1	0	0
18	$X_{\pi/2} - Y_{\pi/2} - X_{\pi/2}$	1	1	1	0	0
19	$Y_{-\pi/2} - X_{\pi}$	0	1	0	1	1
20	$X_{\pi/2} - Y_{\pi}$	1	0	0	0	1
21	$X_{\pi/2} - Y_{-\pi/2} - X_{\pi/2}$	1	1	1	1	1
22	$Y_{\pi/2}$	0	1	0	0	0
23	$X_{-\pi/2} - Y_{\pi}$	1	0	0	1	1
24	$X_{\pi/2} - Y_{\pi/2} - X_{-\pi/2}$	1	1	1	1	0

F.2 INDIVIDUAL RANDOMIZED BENCHMARKING MEASUREMENTS

F.3 DETERMINING POPULATION IN THREE STATES

If there were no leakage present during randomized benchmarking, the full information about the state populations can be extracted from the randomized benchmarking results. However, if leakage to the second-excited state is present, the full information about the populations of the three states can be extracted using two versions of randomized benchmarking, one without a final pi pulse, and one with a final pi pulse. The final pi pulse swaps the populations in the ground and excited state. We assume that the population of the second excited-state remains unaffected by this single final pulse.

Using these two randomized benchmarking sequences, two different signals S_0 and S_1 are measured, corresponding to a measurement without a final pi pulse, and a measurement with a final pi pulse, respectively. This leads to the following three equations:

$$\begin{aligned} p_0 V_0 + p_1 V_1 + p_2 V_2 &= S_0 \\ p_1 V_1 + p_0 V_1 + p_2 V_2 &= S_1 \\ p_0 + p_1 + p_2 &= 1 \end{aligned} \tag{F.1}$$

where p_i corresponds to the final population in state $|i\rangle$, and V_i is the signal of state $|i\rangle$. Filling in $p_2 = 1 - p_0 - p_1$ into the first two equations of F.1, we are left with the following set of equations:

$$\begin{bmatrix} V_0 - V_2 & V_1 - V_2 \\ V_1 - V_2 & V_0 - V_2 \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \end{bmatrix} = \begin{bmatrix} S_0 - V_2 \\ S_1 - V_2 \end{bmatrix} \tag{F.2}$$

This set of equations can be easily solved by matrix inversion, resulting in the following three populations:

$$\begin{aligned} \begin{bmatrix} p_0 \\ p_1 \end{bmatrix} &= \left((V_0 - V_2)^2 - (V_1 - V_2)^2\right)^{-1} \begin{bmatrix} V_0 - V_2 & -V_1 + V_2 \\ -V_1 + V_2 & V_0 + V_2 \end{bmatrix} \begin{bmatrix} S_0 - V_2 \\ S_1 - V_2 \end{bmatrix} \\ p_2 &= 1 - p_0 - p_1 \end{aligned} \tag{F.3}$$

If one has knowledge of all thee signals V_0, V_1 and V_2 (see Section 7.3.4 for information on measuring V_2), the three populations can be obtained using F.3.

Appendix G

Algorithms

G.1 Compiled randomized benchmarking

G.1.1 Finding the optimal gate sequence

In compiled randomized benchmarking each set of Cliffords is compiled such that the total number of pulses sent is as small as possible. This is done by comparing all the possible decompositions for each Clifford in the set of Cliffords with each other, and determining the combination that results in the least amount of total pulses.

Given a tuple of Cliffords $\left(C^1_{\alpha_1},\ldots,C^n_{\alpha_n}\right)$, where n is the number of qubits and α_i is the Clifford number for qubit i, a particular decomposition of the Cliffords is given by $\left(\left(G^1_1,...,G^1_{m_1}\right),...,\left(G^n_1,...,G^n_{m_n}\right)\right)$, where m_i is the number of gates in the decomposition of Clifford $C^i_{\alpha_i}$, and G^i_j is gate j of the Clifford decomposition of C_i. The algorithm used for finding the minimum number of gates for a particular tuple of Clifford decompositions is a recursive algorithm. The algorithm determines all possible ways in which the gates can be ordered, and finally chooses the sequence having the smallest length.

To explain this algorithm, let us denote $\boldsymbol{\beta} = \left(\beta_1,\ldots,\beta_n\right)$ as the indices of the next possible gates. The corresponding gates are given by $\boldsymbol{G_\beta} = \left(G^1_{\beta_1},\ldots,G^n_{\beta_n}\right)$. If the indices where $\beta_i = m_i + 1$, there is no next gate, and so G_{β_i} does not exist and should not be added to $\boldsymbol{G_\beta}$.

1. Start with an empty sequence of gates G_{seq} and gate indices $\boldsymbol{\beta} = \left(\beta_1,...,\beta_n\right) = (1,...,1)$
2. Determine the set of distinct gates in $G_{\boldsymbol{\beta}}$.
3. For each gate g in the set of distinct gates perform the following steps:
   (a) Append g to G_{seq}.
   (b) Determine all indices i for which the next gate $G^i_{\beta_i}$ equals g.
   (c) Copy gate indices $\boldsymbol{\beta}$ to $\boldsymbol{\beta}^{\text{new}} = \left(\beta^{\text{new}}_1,...,\beta^{\text{new}}_n\right) = \left(\beta_1,...,\beta_n\right)$.
   (d) For all indices i for which $G^i_{\beta_i} = g$, increase the gate index $\beta^{\text{new}}_i = \beta_i + 1$.
   (e) Go to step two using the new gate indices $\boldsymbol{\beta}^{\text{new}}$.

4. If $G_{\boldsymbol{\beta}}$ is empty, then G_{seq} is a particular sequence of pulses that would result in all Cliffords being applied.

5. When all possible sequences are combined, choose the sequence with the minimum number of gates in G_{seq}

This algorithm determines the least amount of gates necessary to perform all Cliffords in one particular tuple of decompositions. However, each Clifford has on average 38 different decompositions, and so for n qubits the total decomposition combinations is approximately38^n, meaning this algorithm would have to be repeated 38^n times to find the global minimum number of pulses to perform all Cliffords. And in fact this would only determine the optimal compilation for one particular set of Cliffords; to find the average gates per Clifford tuple one would have to perform this on all possible combinations of Cliffords and then determine the average gates per Clifford tuple. This problem therefore scales exponentially with the number of qubits, and would seem impossible for as few as $n = 3$ qubits, as it would require $24^3 * 38^3 = 7.6 * 10^8$ different decomposition combinations. Nevertheless the average gates per Clifford has been calculated exactly for up to $n = 5$ qubits. This has been done using several optimization methods, which will be discussed below.

G.1.2 Optimizing the gate compilation algorithm

Determining the average gates per Clifford tuple for $n = 5$ qubits would initially result in $6.3 * 10^{14}$ different combinations of Clifford decompositions. For each of these combinations the algorithm must find the least amount of gates necessary to perform all gates in that particular set of Clifford decompositions. It is clear that this is computationally not viable. Luckily several optimizations can be performed which can drastically reduce this number by many orders of magnitude. These optimizations will be discussed in this section.

The first and simplest optimization is from the simple observation that the optimal gate compilation for a certain tuple of Cliffords $\left(C^1_{\alpha_1}, \ldots, C^n_{\alpha_n}\right)$ is the same as any permutation of those Cliffords. Therefore we only have to determine the optimal compilations for the tuples $\left(C^1_{\beta_1}, \ldots, C^n_{\beta_n}\right)$ where $\beta_1 \leq \beta_2 \leq \cdots \leq \beta_n$. This already reduces the amount of calculations by an exponential amount (a factor 81 times less computations when $n = 5$.

The second optimization is to keep track of the minimum number of gates so far found that can compile a given tuple of Cliffords. At each stage of the algorith it checks if the gates that are so far in G_{seq} plus all the distinct gates left in $G_{\boldsymbol{\beta}}$ is equal to or greater than the minimum number of gates found so far. If this is the case, it will know that it cannot find a better combination of gates using the sequence G_{seq}. It can therefore stop this sequence and start with the next sequence. The minimum number of gates can initially be placed at 5 gates, as the 5 primitives method proves that there is always a decomposition of an arbitrary number of Cliffords into 5 gates. This optimization places an upper bound on the number of gates, and results in a massive decrease in computation time. This is especially the case because as the minimum number of gates decreases, so does the frequency at which the algorithm stops its current sequence increase.

The third optimization relies on the fact that it is more likely that decompositions with fewer gates result in an optimal gate compilation. Therefore the decompositions of all Cliffords have been arranged in ascending number of gates. When comparing the decompositions of the Cliffords the first decompositions compared are those with the fewest gates. The probability of finding the optimal gate compilation is therefore high, and even if it did not find the optimal gate compilation, it is likely that the minimum gate length found so far will be low. This optimization is especially effective when combined with the second optimization.

The fourth optimization places a lower bound on the number of gates. For a given tuple of Cliffords $\left(C^1_{\alpha_1},\ldots,C^n_{\alpha_n}\right)$, the lower bound is found by looking at the optimal lengths previously found for all $n-1$ Clifford subsets. This requires knowledge of the lengths of all the $n-1$ Cliffords. Since the length of the tuple of n Cliffords can never be less than any of the lengths of the $n-1$ Clifford subsets, the maximum of these lengths therefore places a lower bound on the optimal number of gates. This means that if during the algorithm ever finds one sequence of gates for any tuple of decompositions whose length is equal to this lower bound, it will know it has found the optimal gate compilation for the tuple of Cliffords, and may abort all further search. This is in contrast to the second optimization, where an upper bound was found, in which case only the particular sequence of gates could be aborted. As the number of qubits increases, it becomes more and more likely that the lower bound is equal to five. In this case the lower bound is equal to the upper bound set by the 5 primitives method, and so it can immediately be concluded that the optimal gate compilation is the 5 primitives method. This optimization that places a lower bound is probably the best of all optimizations used, and results in a massive gain in computation time, by many orders of magnitude.

The fifth optimization is the most complicated optimization, and is based on separating all decompositions with three gates or less from those composed of four gates. In this optimization first all combinations are tested using only decompositions with three or less gates. This greatly reduces the average number of decompositions per Clifford, from 38 to 7. Especially as the number of qubits increases, this results in drastically less comparisons in total. Now it is not always the case that the minimum number of gates is found using only up to three gates per decomposition: sometimes the optimal gate compilation requires one of the decompositions to have four gates. However, we only need to include four gate decompositions when the lower bound is four or less and when the minimum gate compilation found using only three gates or less is equal to five or more. Only in this case could there be a four gate decomposition that could outperform gate decompositions with only three gates or less. We furthermore know that if there is such a four gate decomposition that results in the optimal gate compilation, its length must be equal to four, as the 5 primitives method places the upper bound at 5 gates. We therefore also know that only one Clifford can have a four gate decomposition, and all other Cliffords must be subsets of these four gates. We can therefore simply loop over each of the four gate decompositions, and test whether all other Cliffords are subsets of these four gates. This changes the comparison from scaling exponential with the number of qubits to scaling linear with the number of qubits.

Using these 5 optimizations the average gates per Clifford tuple has been calculated exactly for $n = 5$ qubits within two hours. Furthermore, the average gates per Clifford has

number of qubits	average gates per Clifford
1	1.875
2	2.925
3	3.521
4	3.874
5	4.137

Table G.1: The average gates per Clifford after compilation. Values up to n=5 have been calculated exact, while values starting from n=6 have been determined using random sampling.

number of qubits	1	2	3	4	5	6	7	8	9
gates per Clifford	1.875	2.925	3.521	3.874	4.137	4.380 (12)	4.570 (15)	4.721 (10)	4.808 (14)

Table G.2: The average gates per Clifford after compilation. Values up to n=5 have been calculated exact, while values starting from n=6 have been determined using random sampling.

been approximated for up to $n = 10$ Cliffords using random sampling. The results are shown in Table G.1.

- Change set to the correct term
- Talk about constraints: Single pulse sequence, markers

Bibliography

[1] R Barends, HL Hortensius, T Zijlstra, JJA Baselmans, SJC Yates, JR Gao, and TM Klapwijk. Contribution of dielectrics to frequency and noise of nbtin superconducting resonators. *Applied Physics Letters*, 92(22):223502, 2008.

[2] Rami Barends. *Photon-detecting superconducting resonators*. PhD thesis, Delft University of Technology, 2009.

[3] Lev Samuel Bishop. *Circuit Quantum Electrodynamics*. PhD thesis, Yale University, 2010.

[4] V Bouchiat, D Vion, P Joyez, D Esteve, and M H Devoret. Quantum coherence with a single cooper pair. *Physica Scripta*, 1998(T76):165, 1998.

[5] A Bruno, G de Lange, S Asaad, KL van der Enden, NK Langford, and L DiCarlo. Reducing intrinsic loss in superconducting resonators by surface treatment and deep etching of silicon substrates. *Applied Physics Letters*, 106(18):182601, 2015.

[6] Jerry M. Chow. *Quantum Information Processing with Superconducting Qubits*. PhD thesis, Yale University, 2010.

[7] AD Córcoles, Easwar Magesan, Srikanth J Srinivasan, Andrew W Cross, M Steffen, Jay M Gambetta, and Jerry M Chow. Demonstration of a quantum error detection code using a square lattice of four superconducting qubits. *Nature communications*, 6, 2015.

[8] L DiCarlo, JM Chow, JM Gambetta, Lev S Bishop, BR Johnson, DI Schuster, J Majer, A Blais, L Frunzio, SM Girvin, et al. Demonstration of two-qubit algorithms with a superconducting quantum processor. *Nature*, 460(7252):240–244, 2009.

[9] Arkady Fedorov, Lars Steffen, Matthias Baur, MP Da Silva, and Andreas Wallraff. Implementation of a toffoli gate with superconducting circuits. *Nature*, 481(7380):170–172, 2011.

[10] Jiansong Gao, Miguel Daal, Anastasios Vayonakis, Shwetank Kumar, Jonas Zmuidzinas, Bernard Sadoulet, Benjamin A Mazin, Peter K Day, and Henry G Leduc. Experimental evidence for a surface distribution of two-level systems in superconducting lithographed microwave resonators. *Applied Physics Letters*, 92(15):152505–152505, 2008.

[11] Kurtis L. Geerlings. *Improving Coherence of Superconducting Qubits and Resonators*. PhD thesis, Yale University, 2013.

[12] Blake R Johnson, Marcus P da Silva, Colm A Ryan, Shelby Kimmel, Jerry M Chow, and Thomas A Ohki. Demonstration of robust quantum gate tomography via randomized benchmarking. *arXiv preprint arXiv:1505.06686*, 2015.

[13] E Knill, D Leibfried, R Reichle, J Britton, RB Blakestad, JD Jost, C Langer, R Ozeri, S Seidelin, and DJ Wineland. Randomized benchmarking of quantum gates. *Physical Review A*, 77(1):012307, 2008.

[14] Jens Koch, Terri M. Yu, Jay Gambetta, A. A. Houck, D. I. Schuster, J. Majer, Alexandre Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf. Charge-insensitive qubit design derived from the cooper pair box. *Phys. Rev. A*, 76:042319, Oct 2007.

[15] Erik Lucero, Julian Kelly, Radoslaw C Bialczak, Mike Lenander, Matteo Mariantoni, Matthew Neeley, AD OâĂŹConnell, Daniel Sank, H Wang, Martin Weides, et al. Reduced phase error through optimized control of a superconducting qubit. *Physical Review A*, 82(4):042339, 2010.

[16] Easwar Magesan, Jay M Gambetta, and Joseph Emerson. Scalable and robust randomized benchmarking of quantum processes. *Physical review letters*, 106(18):180504, 2011.

[17] Easwar Magesan, Jay M Gambetta, Blake R Johnson, Colm A Ryan, Jerry M Chow, Seth T Merkel, Marcus P da Silva, George A Keefe, Mary B Rothwell, Thomas A Ohki, et al. Efficient measurement of quantum gate error by interleaved randomized benchmarking. *Physical review letters*, 109(8):080505, 2012.

[18] John M Martinis and A Megrant. Ucsb final report for the csq program: Review of decoherence and materials physics for superconducting qubits. *arXiv preprint arXiv:1410.5793*, 2014.

[19] Benjamin A. Mazin. *Microwave Kinetic Inductance Detectors*. PhD thesis, California Institute of Technology, 2004.

[20] F Motzoi, JM Gambetta, P Rebentrost, and Frank K Wilhelm. Simple pulses for elimination of leakage in weakly nonlinear qubits. *Physical review letters*, 103(11):110501, 2009.

[21] Yu Nakamura, Yu A Pashkin, and JS Tsai. Coherent control of macroscopic quantum states in a single-cooper-pair box. *Nature*, 398(6730):786–788, 1999.

[22] Michael A Nielsen and Isaac L Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010.

[23] I Nsanzineza and BLT Plourde. Trapping a single vortex and reducing quasiparticles in a superconducting resonator. *arXiv preprint arXiv:1405.0256*, 2014.

[24] PJJ OâĂŹMalley, J Kelly, R Barends, B Campbell, Y Chen, Z Chen, B Chiaro, A Dunsworth, AG Fowler, I-C Hoi, et al. Qubit metrology of ultralow phase noise using randomized benchmarking. *Physical Review Applied*, 3(4):044009, 2015.

[25] Britton Plourde, C Song, TW Heitmann, MP DeFeo, and Kang Yu. Microwave response of vortices in superconducting thin films of re and al. 2009.

[26] Matthew D. Reed. *Entanglement and Quantum Error Correction with Superconducting Qubits*. PhD thesis, Yale University, 2013.

[27] D Ristè, S Poletto, M-Z Huang, A Bruno, V Vesterinen, O-P Saira, and L DiCarlo. Detecting bit-flip errors in a logical qubit using stabilizer measurements. *arXiv preprint arXiv:1411.5542*, 2014.

[28] Jeremy M Sage, Vladimir Bolkhovsky, William D Oliver, Benjamin Turek, and Paul B Welander. Study of loss in superconducting coplanar waveguide resonators. *Journal of Applied Physics*, 109(6):063915, 2011.

[29] JA Schreier, Andrew A Houck, Jens Koch, David I Schuster, BR Johnson, JM Chow, Jay M Gambetta, J Majer, L Frunzio, Michel H Devoret, et al. Suppressing charge noise decoherence in superconducting charge qubits. *Physical Review B*, 77(18):180502, 2008.

[30] Adam P. Sears. *Extending Coherence in Superconducting Qubits: from microseconds to milliseconds*. PhD thesis, Yale University, 2013.

[31] Sarah Sheldon, Lev S Bishop, Easwar Magesan, Stefan Filipp, Jerry M Chow, and Jay M Gambetta. Characterizing errors on qubit operations via iterative randomized benchmarking. *arXiv preprint arXiv:1504.06597*, 2015.

[32] Gabriel Vasilescu. *Electronic noise and interfering signals: principles and applications*. Springer, 2006.

[33] J Wenner, R Barends, RC Bialczak, Yu Chen, J Kelly, Erik Lucero, Matteo Mariantoni, A Megrant, PJJ OâĂŹMalley, D Sank, et al. Surface loss simulations of superconducting coplanar waveguide resonators. *Applied Physics Letters*, 99(11):113513, 2011.