Rapid, High Fidelity, Single-Shot Readout

Theodore E. Walter

Master Thesis

Quantum Device Lab, ETH Zurich

Supervisor: Philipp Kurpiers

Professor: Andreas Wallraff

September 30, 2014

Abstract

Single-shot measurements are fundamental to the success of quantum information processing and circuit quantum electrodynamics (cQED) experiments. Improving readout fidelity has been an active area of research in both fields since their inceptions. However, there has been little effort to improve the absolute speed of the measurements. In this thesis the practical and theoretical aspects of single-shot measurements are studied with the aim of maintaining a high level of fidelity, as well as, understanding and improving the speed of the measurement process. After first discussing the optimization of each step of our measurement procedure, each parameter of the measurement model is investigated in detail in order to illuminate the relationships and limitations. Our experimental results have aligned well with the theoretical model, allowing us to predict an optimal set of system parameters for rapid, high fidelity, single-shot readout. With the optimized parameters we hope to create a cQED chip with readout capabilities of greater than 95% fidelity in only 40 ns from the time of state preparation to state determination.

Contents

1 Introduction and Motivation

Accurate determination of a qubit state is quintessential for quantum computation and circuit quantum electrodynamic experiments (cQED). To date, nearly all experiments related to high fidelity readout are concerned with extending the qubit decay time for improvements. In January, 2014, the Martinis group claimed to achieve the surface code threshold with 99.8% fidelity in 140 ns [20]. The high fidelity can most likely be attributed to the use of a Purcell filter, improving the effective lifetime of the qubit by a factor of $100/\kappa$. While this is indeed a good path for higher fidelity, this has less impact on the absolute time necessary for readout. The 140 ns is really only "fast" in the relative sense to the qubit decay time of roughly 10 μs . This thesis will attempt to answer what influences the speed of readout, and what level of fidelity one could expect in only 40 ns?

Improving readout speed, as defined as the time from state preparation to state determination, may find applications in feedback based experiments and quantum algorithms. In feedback experiments for example, there is often a race against the decoherence times of ancillary qubits. A direct improvement of the measurement speed should directly improve the results of the feedback. Also, with faster readout times, one may be able to increase the complexity of quantum algorithms for other improvements and functionality [16].

In the Wallraff lab, our interest in measurement speed is related to a future Bell Test experiment, in which the speed of the measurement is directly related to the cost and complexity of the experimental setup [3]. This experiment will attempt to close the locality loophole of the Bell Test [25]. Therefore, the measurement of the two qubit states must be finished before a light signal can propagate from one qubit to the other. On top of speed, however, high fidelity is also necessary in order to close the detection loophole [19][1]. Both features have yet to be achieved simultaneously in an experiment.

Finally, in my opinion, the most important motivation is to get an increased understanding of the measurement process with hopes that it will lead to better design principles. By attempting to get the fastest measurement possible with high fidelity, we will need to push our current measurement theory to its limits and better understand all of the variables and constraints. We then hope that the answers we find in this more heavily constrained environment will also work in a setting where timing is not a limitation. This is exactly why, for example, automotive companies spend large amounts of money on race-car divisions, even though the actual consumer products they produce for profit do not have the necessity to be the fastest.

In this thesis I will first give a brief introduction to single-shot measurements, I will explain our experimental setup and then discuss the techniques and procedures for practically optimizing the readout fidelity. After this, I will delve into optimizing the qubit parameters, and explain the conclusions of the thesis.

2 Experimental Setup and Single-Shot basics

Though a large portion of what is discussed in this thesis is quite general for quantum measurements, the focus will be on superconducting circuits with coplanar wave guide resonators coupled to Transmon qubits. For a more complete description of this system I refer the reader to the paper from Koch et. al [13]. In this section, I will first briefly overview the experimental context, I will then give a detailed explanation of our experimental setup, as well as the various steps we take to perform a single-shot measurement and extract our measurement fidelity.

2.1 General Context

This thesis can be put in a larger context of cQED and a theoretical setup is depicted in Figure 1. An input line is capacitively coupled to a resonator with decay rate κ_{in} and bare resonant frequency ω_{res}. Coupled to the resonator with strength, g, is a qubit which has its own charge-line for single qubit gate pulses. The output of the resonator is coupled to the output line with decay rate κ_{out}. After exiting the resonator, there will be some amplification chain and down-conversion scheme before the signal is detected and recorded.

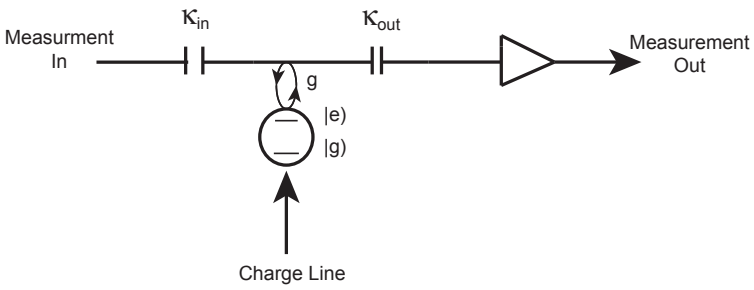

Figure 1: A basic cQED system with resonator coupled to a qubit. The input and output of the resonator are capacitively coupled to the outside world. The output signal will then proceed to amplification and down-conversion before being measured.

This system can most generally be modeled using a Jaynes-Cummings Hamiltonian given by

$$\hat{H}_{\text{JC}} = \hbar \sum_{\text{j}} \omega_{\text{j}} |j\rangle \langle j| + \hbar\omega_{\text{res}} \hat{a}^\dagger \hat{a} + \left(\hbar \sum_{\text{i}} g_{\text{i,i+1}} |i\rangle \langle i+1| \hat{a}^\dagger + \text{H.c.} \right) \tag{1}$$

where $|i\rangle$ are the "atomic states", $\hat{a}$, $\hat{a}^\dagger$ are the annihilation and creation operators, respectively, for photons in the resonator and g_{ij} is the coupling between the resonator and the i, j transition of the qubit. Only nearest level coupling has been taking into account as a simplification in the large $E_{\text{J}}/E_{\text{C}}$ Transmon limit [13].

The Jaynes-Cummings Hamiltonian can only be numerically solved and one is tempted to further simplify this equation down to the common dispersive Hamiltonian given by

$$\hat{H}_{\text{Disp}} = \hbar \frac{\tilde{\omega}_{\text{ge}}}{2} \hat{\sigma}_{\text{z}} + \hbar(\tilde{\omega}_{\text{res}} + \chi \hat{\sigma}_{\text{z}}) \hat{a}^\dagger \hat{a} \tag{2}$$

In this equation the qubit transition frequency, $\tilde{\omega}_{\text{ge}}$, and resonator frequency, $\tilde{\omega}_{\text{res}}$, have been renormalized, only the first two levels of the "atom" are taken into account, $\hat{\sigma}_{\text{z}}$, and χ is the dispersive shift. This formula is derived from second order perturbation of Equation 1 under the assumptions that $g/\Delta_{\text{i}} \ll 1$, where Δ_{i} is the difference between

the i^{th} qubit transition frequency and the resonator frequency.

The Dispersive Hamiltonian can be analytically solved and allows one to predict χ, as well as other system parameters, for example, the energy levels, anharmonicity etc. The interaction between the qubit and the resonator causes the resonant frequency shifts, χ, dependent on the state of the qubit. This is directly seen in the $\tilde{\omega}_{\mathrm{res}} + \chi\hat{\sigma}_{\mathrm{z}}$ term of Equation 2. This shift causes changes in the In-phase and Quadrature (I, Q) components, or correspondingly the amplitude and phase, of a signal transmitted through the resonator. These differences will be used to determine the state of the qubit during a single-shot measurement.

2.2 Qubit Parameter Table and Schematics

Throughout the thesis, many experimental results will be reported and the system parameters for the respective setups are shown in Table 1. Figure 2, on the next page, shows the physical schematics of the circuit chips used in the experiments.

Parameter	Chip 1	Chip 2	Chip 3	Optimized Chip
$\Delta = \omega_{\mathrm{qb}} - \omega_{\mathrm{res}}$	-1.76 GHz	-1.800 GHz	373 MHz	430 MHz
ω_{res}	7.722 GHz	7.712 GHz	7.202 GHz	7.8 GHz
$g/2\pi$	70 MHz	250 MHz	50 MHz	24 MHz
$T_{1\mathrm{eg}}$	3.8 μs	3.1 μs	0.35 μs	3 μs
$T_{1\mathrm{fe}}$	3.0 μs	-	-	-
α	374 MHz	290 MHz	440 MHz	480 MHz
$\kappa/2\pi$	2.2 MHz	2.2 MHz	20 MHz	15 MHz

Table 1: Qubit Parameter Table. Important parameters for the three chips used during this thesis, as well as, the optimized parameters found for the results of the thesis presented in Section 5.1. The T_{1eg} value given for the optimized chip is a theoretically derived value approximately $0.8 * 1/\gamma_{\mathrm{purcell}}$ defined in equation 16. The resonators were all highly asymmetric with $\kappa_{\mathrm{out}}/\kappa_{\mathrm{in}} \approx 100$.

2.3 Experimental Setup

The complete experimental setup can be seen in Figure 3. The setup was similar for all measured chips with the exception of the specific dilution refrigerator which was different for Chip 3. The chip box has been left empty, symbolizing that this is different and one should refer to Figure 2 for the chip specific designs. In Figure 3, the red section represents the up-conversion process for the resonator input signal, the green section is the up-conversion for the charge-line pulses and qubit tuning. The black section includes the output amplification chain and the purple part is the analog-down conversion and ends at the input of the FPGA card. The AWG creates the qubit pulses and also triggers the FPGA to start and stop recording its input as well as when the measurement generator should turn on and off. The orange section is used to drive the amplifier and minimize this drive signal from leaking back into the chip by splitting the pump tone and passing it through a variable phase and attenuator for destructive interference of the reflected wave [7].

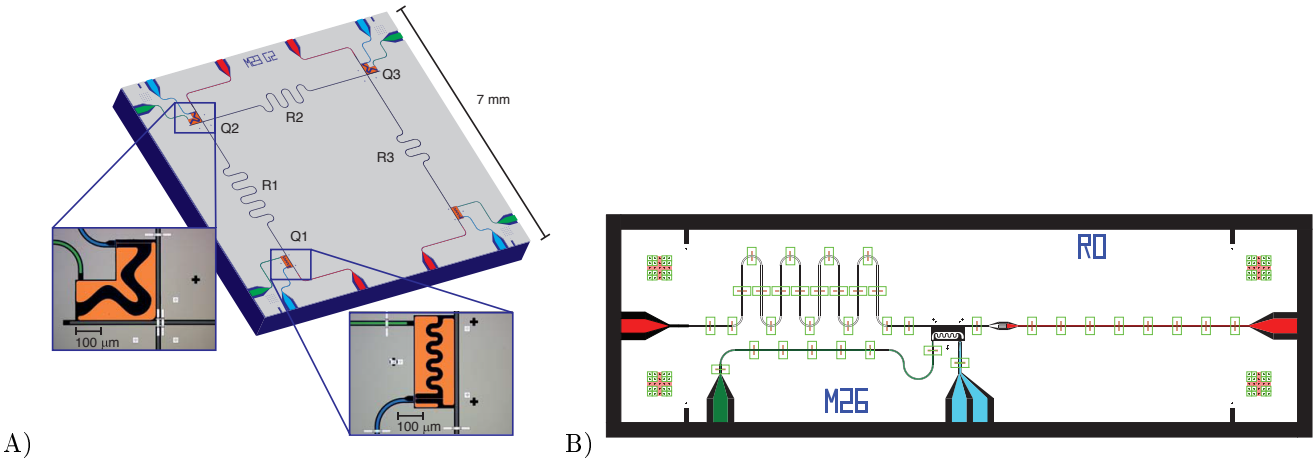

Figure 2: Schematics of the three chips used in this thesis. A) The first two chips have an identical mask, however Chip 1 had the flux-lines shorted. The charge-lines are green, the flux lines blue and the measurement line is red. The chips were produced from Mask 23 and the figure is presented in [24]. We used qubit two (Q2) and resonator one (R1) for Chip 1 data, and qubit one (Q1) and resonator one (R1) for Chip 2 data. B) The third chip was produced on Mask 26[17]. The green rectangles on the lines are air bridges used to stabilize the ground plane. The color coding is the same as in A)

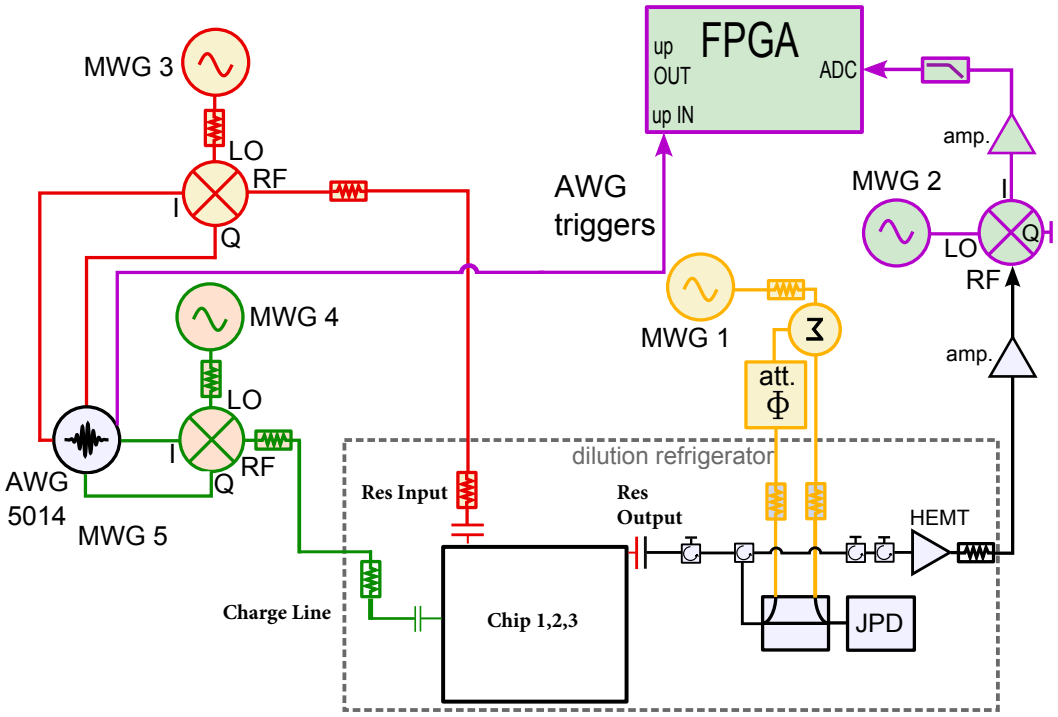

Figure 3: Overview of experimental setup. Red components are for resonator input, green for charge line input, black for resonator output and amplification, purple is for analog down-conversion and FPGA recording. The orange components are for driving the JPD as well as drive leakage cancellation. The chip box is the location of the various chips used in this thesis and shown in Figure 2.

The parametric amplifier used in our experiments was a Josephson Parametric Dimer (JPD) and is fully described in Eichler et. al. [?]. It is composed of a pump signal and two coupled SQUID arrays, which act as coupled oscillators for parametric amplification. It can amplify in either a phase-sensitive or insensitive mode in which only a single quadrature of the signal is amplified or both quadratures are amplified equally. Unless otherwise specified, the JPD was calibrated in a phase-insensitive mode with roughly 20 dB gain.

2.4 Single-shot Measurements Procedure

A single-shot measurement is one where the eigenstate of the qubit must be determined in a single experimental run. In order to determine an optimal measurement procedure for this task, the qubit is prepared in its possible eigenstates and transmission statistics are gathered. From the statistics a decision boundary is calculated, which divides the ground and excited state signals. With this threshold value, one can determine the probability of correctly identifying the qubit state and quantify this with the fidelity metric, formally defined momentarily.

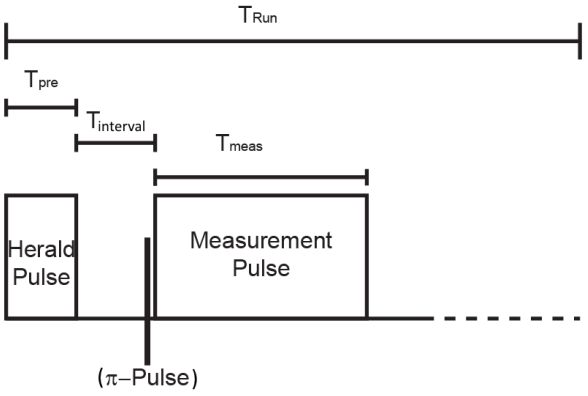

Figure 4: Single-shot measurement pulse scheme. The experiment begins with a Heralding Pulse, followed by a time T_{interval}, and then an optional π-pulse. After this, the measurement pulse is applied and afterward there are no pulses for the rest of the 25 μs to reinitialize the qubit to the ground state.

Our measurement procedure follows the measurement pulse scheme depicted in Figure 4. Each run is 25 μs and begins with a heralding measurement pulse of length T_{pre}. This first pulse is followed by some optimized interval time, T_{interval}, which will be discussed in more detail in Section 3.2. After which, a $\pi-$pulse is applied to the qubit if we wish to gather excited state statistics, otherwise no pulse is applied [2]. Following the possible $\pi-$pulse, the actual measurement pulse is applied for roughly 3-5 μs to get a full quantum trajectory measurement. After which, the remaining time is left for the qubit to return to the ground state for another repetition of the experiment. The FPGA, which acts as our detector at the end of the output line, is continuously running from the beginning of the Heralding pulse to the end of the Measurement Pulse. It records the integrated signal of both the I and Q quadrature at its input every 10 ns [14].

The two pulses applied to the resonator are identical in all but their time length. The herald pulse is used for measuring the initial state of the qubit and the result from it can be used for post-selection, that is removal of data sets where the qubit was not properly initialized to the ground state. The measurement pulse is longer, simply to see the full trajectory of the qubit.

The results of two different experimental runs are shown in Figure 5, where the Q signal component is plotted as a function of time. In order to produce this graph a 20-point box car filter has already been applied to the raw

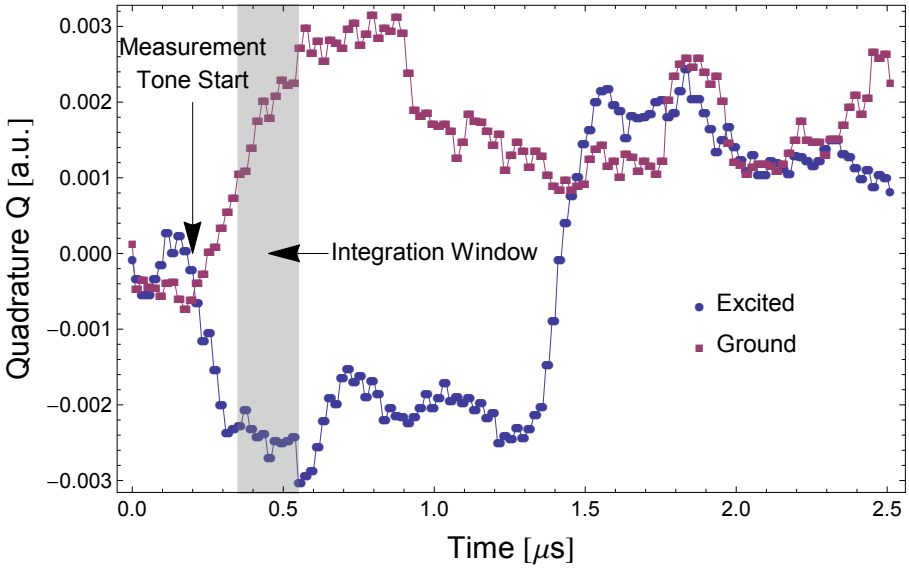

Figure 5: Example Quantum trajectories. The Quadrature component of the transmitted signal is recorded as a function of time for a qubit prepared in the ground and excited state. The data has been 20-point boxcar filtered to remove noise. The data points are not averages and the lines only connect the points and are not fits. The data acquired from Chip 2.

data set, the use of filters will be explained in Section 3.3. The data was recorded from Chip 2.

There are a number of fascinating things which can be gleaned from quantum trajectories, but for our purposes there are only a few main features to see [9]. The first thing is that there are distinct differences between the ground and excited state trajectories and that the difference in trajectories is time dependent. Secondly, one can see a quantum jump of the excited state to the ground state at around 1.4 μs. Of course these events are probabilistic and in this setup the measured T_1 time is roughly 3 μs, but this particular excited state decayed much earlier. Finally, one can see in this case that the level of fluctuations is large compared to the difference of the two trajectories.

We not only look at the the Q component, but also the I component of the signal

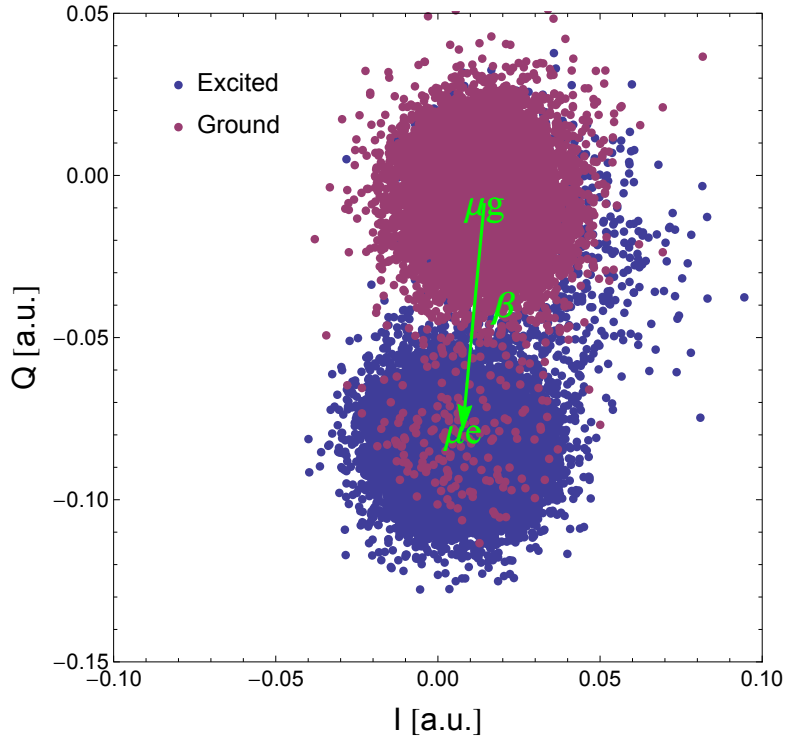

Figure 6: IQ scatter plot. Over 10,000 traces of the qubit prepared in the ground and excited state are taken. From each trace of the I and Q component the signal is integrated and form an (I,Q) pair which is plotted here. The mean of the ground and excited state traces have been depicted by μg and μe and a vector β is also shown which is the vector difference of the two mean vectors defined from the origin. Data taken from Chip 2.

as well. From these trajectories we integrate each component over some identical optimal time to get an (I,Q) pair representing the state of the qubit. The determination of the optimized integration window will be discussed in Section 3.3 and 3.4. Repeating these runs many times, and keeping the integration time identical, one can build up a scatter plot of (I,Q) pairs as seen in Figure 6.

From the IQ scatter plot one sees that the ground and excited state have some mean and standard deviation. In Section 4.5, it will be further clarified why these two states split in the IQ plane and what the split depends on. Also, in the picture is defined a vector β which is the difference between the two means in the IQ plane and will be more formally discussed in Section 4.5.

If one repeats the measurement procedure with the qubit in an unknown state, then if the derived IQ pair is closer to the ground state mean values, it is most likely in the ground state. Conversely, if the value is closer to the excited state means, then the qubit is in the excited state. In fact, the entire information about the state of the qubit can be projected onto a single axis. Projecting the data on an axis perpendicular to β results in the mean values of the states becoming identical, and therefore they contain zero information about the state of the qubit. However, if we project the data along an axis defined by β, then the mean values of the two states become maximally different. This rotation in the IQ plane allows one to reduce the data to a single dimension without loss of information. For Figure 6 one can rotate the data three degrees counter clockwise so that β is parallel to the Q axis. Then by projecting the data onto this β =Q' axis one can produce a histogram for the two states as seen in Figure 7. If one instead looks at the components of the vectors along the axis perpendicular to this, the two state histograms would be exactly on top of each other.

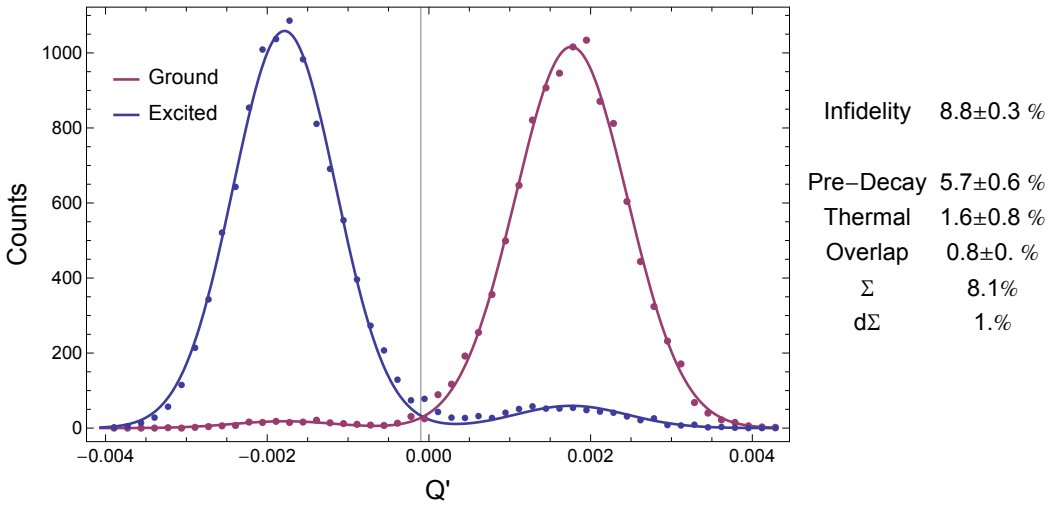

Figure 7: Histograms of IQ data projected on the axis through β of Figure 6. The lines are fits to the data points and the terms on the right are calculated from the fit. See text for details.

These histograms contain the totality of statistics related to the measurement protocol and useful information can be extracted from them. The first thing that will be ascertained is the threshold value, which is depicted in

Figure 7 as the vertical line near the origin. This value is the decision boundary, data points found below this line will be identified as the qubit being in the excited state, and for points above it, the qubit is identified as being in the ground state. Determining the threshold value is done by simply finding the point at which the percentages of ground states below and excited states above are at a minimum. This percentage of misidentified runs is an exact calculation of the Infidelity, defined as the complement of fidelity.

Fidelity is defined as

$$F = 1 - P(g|e) - P(e|g) \tag{3}$$

where $P(i|j)$ is the probability of identifying the qubit in the i state when it is really in the j state and $i \neq j$. Fidelity is the metric of choice for measuring the "quality" of a measurement procedure and it is related to the ability to correctly identify a qubit state. It is not a proper percentage, however, as fidelity ranges from -1 to 1, but it is symmetric about zero and a negative fidelity means you have switched the names of the states. In this particular case, our Infidelity can be seen to be 8.8%, resulting in a fidelity of 91.2%.

The rest of the information which can be extracted from this plot is used to identify the sources of the Infidelity. In order to extract the sources, a fit is applied to the histograms. The model used for the fit takes the following form and assumptions

$$\begin{cases} C_{\mathrm{e}}(Q') & = A_{11}N(\mu_{\mathrm{e}}, \sigma_{\mathrm{e}}) + A_{12}N(\mu_{\mathrm{g}}, \sigma_{\mathrm{g}}) \\ C_{\mathrm{g}}(Q') & = A_{21}N(\mu_{\mathrm{e}}, \sigma_{\mathrm{e}}) + A_{22}N(\mu_{\mathrm{g}}, \sigma_{\mathrm{g}}) \end{cases} \tag{4}$$

where C_{i} is the state dependent counts as a function of Q', N represents a normal distribution with mean μ_{i}, and standard deviation, σ_{i} [11]. The two equations are fit simultaneously to both data sets allowing for a better extraction of μ_e, μ_g, σ_e, σ_g. Therefore, we assume that the Gaussians on the right of the threshold are identical in mean and standard deviation, but differ in amplitude. We assume the same for the Gaussians on the left of the threshold. In total, then, we have four amplitudes, two means and two standard deviations as fitting parameters. The other three terms in Figure 7 can now be gleaned from the fit results.

The Pre-decay term is defined as $A_{12}/(A_{12} + A_{11})$ and includes a few possible physical processes. It includes qubits which are excited by the $\pi-$pulse correctly, but decay before the measurement integration begins. Qubits which are thermally excited after the heralding pulse, so that the $\pi-$pulse returns the qubit to the ground state. It also includes failed $\pi-$pulsed qubits and finally a small portion of excited state trajectories which decay near the very beginning of a long integration time.

Analogously, the Thermal term is defined as $A_{21}/(A_{21} + A_{22})$ and represents qubits which are thermally excited after the heralding pulse.

The Overlap term is defined as the integral of the ground state normal distribution function below the threshold, plus the integral of the excited state normal distribution above the threshold. The sum of the three terms is given

by the Σ term and the total error is given by the $d\Sigma$ term. These three terms describe the dominate physical process which cause Infidelity as can be seen from $\Sigma \approx$ Infidelity in Figure 7.

The fit model, however, does not directly take into account qubit decay or excitation during the measurement integration, which would lead to more counts in the middle between the two distributions. This is because the trajectory will begin with one mean value, and switch to the other state mean value at some time during integration. This results in an averaged mean value between the two state means, depending on the relative time of the state jump compared to the integration time [11]. A more thorough model including this possibility was found to actually reduce the usefulness of the fit, as the number of fit parameters grows too large, leading to arbitrary fitting capabilities with larger errors in the fitted values. On top of that, there was not a great increase in the amount of information gained from such a fit as one could not reliably separate the different process further than was done using Equation 4. Finally, the model also does not include the possibility of state mixing, which will be explained in more detail in Section 4.6. These shortcomings are the reason the data points and the fit do always agree, especially in the middle between the ground and excited state peaks, as well as sometimes on the tails.

3 Practical Optimizations

Before one can consider optimizing the measurement for speed, one must make sure all experimental aspects of the measurement process are already near their optimum for readout fidelity. In this section I will discuss the practical optimizations of the experimental process to make sure one is capable of extracting every percent of fidelity during a measurement, independent of the system parameters. This is akin to calibrating and fixing all other variables in a experimental setup. First, there will be a brief introduction to the trace distance, followed by optimizing the heralding process, the filtering, the integration start time, and finally the measurement power and amplification gain.

3.1 Trace Distance

As stated previously we will use fidelity as our metric for the quality of a measurement procedure. As seen in the introduction to single-shot measurements, the procedure for calculating the fidelity is dependent on one's ability to rotate the IQ scatter plot and identify the optimal threshold value. This procedure is resource intensive and prone to error, so instead the trace distance between the two IQ scatter sets is calculated. The trace distance is defined as

$$D(P,Q) = \frac{1}{2}\sum_{\mathrm{x,y}} |P(x,y) - Q(x,y)| \tag{5}$$

where P and Q are probability distributions. This metric is continuous and bounded from zero to one. It is zero when the two probability distributions are identical or $P(x,y) = Q(x,y)\ \forall x,y$. The trace distance is one if and only if the distributions do not overlap at all, or $P(x,y) = 0$ when $Q(x,y) \neq 0$ as well as $Q(x,y) = 0$ when $P(x,y) \neq 0$

for all x and y. It is therefore related to the similarity between two arbitrary probability distributions and calculates their relative overlap. It can be shown that this value is an upper bound to our definition of measurement fidelity, ie $F(e,g) \leq D(e,g)$ [16]. Because the trace distance is rotation independent and threshold independent, it provides a better way to compare different experimental procedures, especially when the difference in the extracted fidelity is very small. When we compare the trace distance instead of the fidelity, we remove the possibility that the extracted fidelity is higher in one case due to a more optimal rotation and threshold value. This should leave only the experimental procedures as the source of Infidelity. In all of the data which proceeds, however, the actual fidelity is shown after the rotation and thresholding procedure, but our qualitative conclusions of comparing different procedures were driven by the trace distance calculations.

3.2 Heralding, Post-Selection and T_{interval}

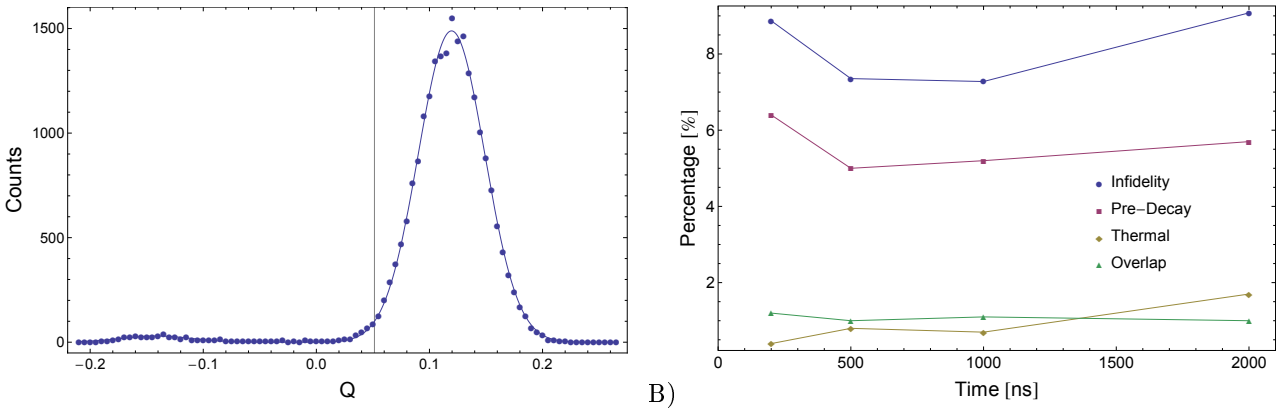

Figure 8: A) Heralding Plot. The heralding pulse data is 4 point box car filtered and totaled before projection and binning. The points represent the data and the blue line is a double Gaussian fit for the ground state and thermally excited states of the qubit. The gray vertical line is the threshold value, below which, the data sets will be removed from analysis because they are most likely improperly initialized qubits. **B) T_{interval} optimization.** The results of single-shot measurements after varying T_{interval}. Data acquired from Chip 2.

In order to minimize the Thermal term discussed in the Section 2.4, an optimized post-selection process is used. The results of the traces taken during the herald pulse time are first 4-point box car filtered to remove the noise from the down conversion process which will be discussed in Section 3.3. Then the data trace is summed, optimally rotated, as explained in the previous section, and histogramed. The resulting histogram is then fitted using only the C_{g} formula in Equation 4 since we do not prepare qubit in the excited state before the herald pulse. The herald plot for Chip 2 can be seen in Figure 8A. From the fit we extract a threshold value defined as the point where the cumulative distribution function of the large ground state peak reaches 1%. It is represented by the vertical line. Points below this line are traces which are removed from further analysis. In this case we filtered out 3.9% of the total taken traces, and this is a pretty typical value. In the Kelvionx and Triton fridges we routinely got 3-5% Thermal excitations.

The hope is that after this heralding measurement, the qubits are prepared in the ground state before applying

the $\pi-$pulse and beginning the measurement. The interval time between the herald pulse and the measurement pulse must be optimized for maximal correlation between the state of the qubit at the end of the herald pulse and at the time of the measurement pulse. If the interval is too short, then photons from the herald pulse will still populate the resonator and we expect these photons to leak into our excited state measurements and over-exaggerate the Pre-Decay value. Also, we expect that the remaining photons will AC-Stark shift the qubit and reduce the accuracy of our $\pi-$pulses [21]. However, if we wait too long then we expect the post-selection filtering to be less efficient and the Thermal value should rise back to its equilibrium value. This result is verified in Figure 8B and one can see that there is likely a global minimum between 500 ns and 1000 ns for this sample. We chose 700 ns for the optimum T_{interval} for this chip.

3.3 Optimized Filtering

Filtering is a useful mathematical tool for extracting the relevant signal out of the underlying noise. It is essential in cases were the bare signal is weak compared to the level of the noise in a measurement. With the use of linear filters, it can be shown that a mode-matched filter is optimal for maximizing the signal to noise ratio (and therefore the fidelity, as we will be explained in Section 4.2) for the case of a known signal buried in white Gaussian noise [15]. A mode-matched filter is one where the measured signal is filtered with the theoretically expected signal. For example, given a noisy digital pulse, the optimal filter would simply be a boxcar filter, as it matches the theoretical step function of the "true" signal.

One can either use the theoretical model presented later in this thesis, or the average signal of many runs, to determine the "expected" signal to be used as the filter in the single-shot measurement experiment. We expect the excited state signal to have a strong exponential decaying factor due to the decay of the state back to the ground level and this suggests an exponential filter as the optimum linear filter for our measurements.

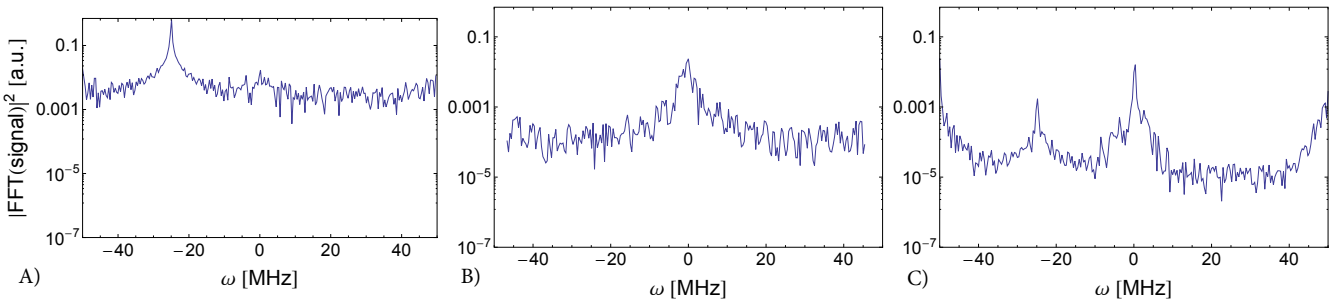

Figure 9: Noise Filtering. Fourier Transform of A) unfiltered signal, B) signal after optimal box car filtering, and C) signal after mode-matched filtering. Data taken with Chip 2

After repeated experiments comparing a mode-matched filter to an optimized boxcar filter it became clear, however, the boxcar filter was marginally better for our measurement setup. The reason for this can be seen in Figure 9 where the Fourier Transform (FFT) of the unfiltered signal is shown. The gigantic peak of noise around the -25 MHz range dwarfs the true signal located around 0 MHz. The origin of this noise peak is a consequence of

our two stage down conversion. After the signal is routed out of the cryostat, it passes various amplifying and filter components, it reaches an analog down conversion process which converts the high frequency GHz signal down to a 25 MHz signal. The 25 MHz signal then enters the FPGA and is sampled at the Nyquest frequency of 50 MHz, in this case, and is therefore digitally down converted to a DC signal. This process, however, shifts all of the DC and low frequency noise at the input of the FPGA down to -25 MHz. This was done purposefully to separate the signal from this high power noise for easy filtering. Any filter with very small frequency components at -25 MHz will completely remove this noise. A 4-point box-car filter is then a very ideal filter as it is identically zero at -25 MHz.

Figure 9B) and C) are the signal after box-car filtering and mode-matched filtering respectively. One can see that the box-car filter effectively removes all of the -25 MHz noise leaving the true signal around 0 MHz the largest peak. The mode-matched filtered signal at 0 MHz is a bit sharper and smoother than the box-car filtered data. Unfortunately, the -25 MHz peak significantly persists through the filter and edge-effects also reduce the quality. Though the noise peaks are still smaller then the true signal, the overall signal to noise ratio is lower then the box-car filtered case, leading to slightly worse fidelity. We tested a hybrid filter by combining the box-car filter and the mode-matched filter, but the results were not conclusively better, so for simplicity we opted to apply an optimized box-car filter to our measurements.

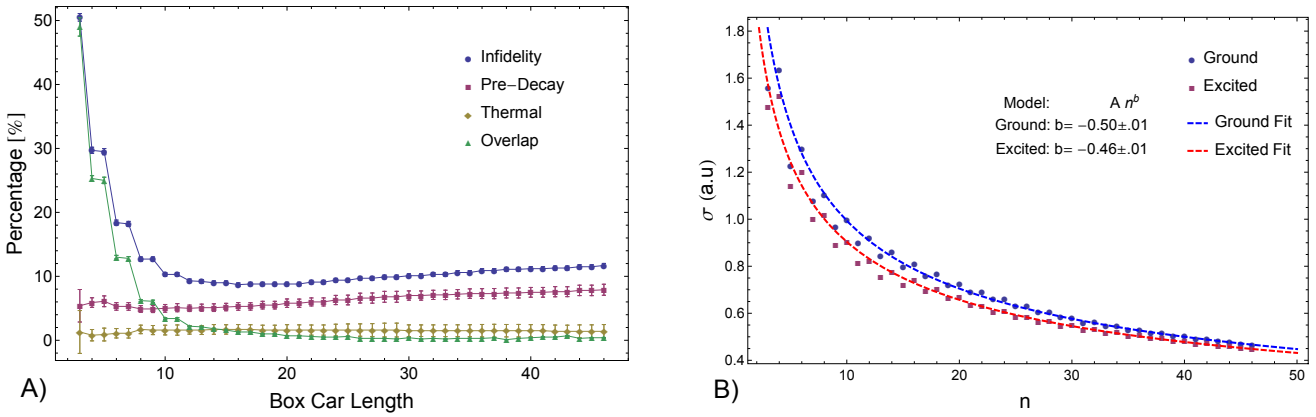

Figure 10: Optimized box-car length. A) Plot of single-shot results as a function of box-car filter length. B) Plot of the standard deviation of the ground and excited state traces as a function of the box-car filter length. Data taken from Chip 2.

Applying a box-car filter is identical to a normalized integration of the signal, where each point is summed and weighted by the total number of points in the sum. Since it is a type of average, on one hand we expect as more points are added to the filter, the effective standard error of the signal to drop, causing a reduction in the Overlap term between the two state Gaussians. On the other hand, as more points are added to the filter from later times, we expect a greater increase of decay during the integration, which will inflate the Pre-Decay term and lower the fidelity. Therefore, there should be an optimal number of filter points which balances these two things. It is well known that the standard error is inversely proportional to the square root of the number of data points so we also plotted the standard deviation of the ground and excited state Gaussians as a function of the box-car length. All

of these expectations are reflected in our experimental results shown in Figure 10. The excited state trace is seen to have a dependence of $b = -0.46$ instead of -0.5, this is likely due to corruption from decay.

3.4 Optimum Starting Time

The analysis of the previous section regarding optimized integration time for the box-car filter is performed in parallel with finding the optimal time to start integrating. In Figure 10 the data was obtained by fixing the integration start time for comparison purposes. However, if we allow the initial time of integration to be variable while searching for the optimum box-car filter we expect qualitatively similar, but quantitatively better results in particular cases. For example, in Figure 10 the optimal boxcar will likely have the most separated points in the middle of the integration window. This means that we will not apply the 2-point box car filter on the two optimal points, but instead on two earlier points with worse separation. Figure 11 corroborates these expectations as one can see the initial points have a much better fidelity.

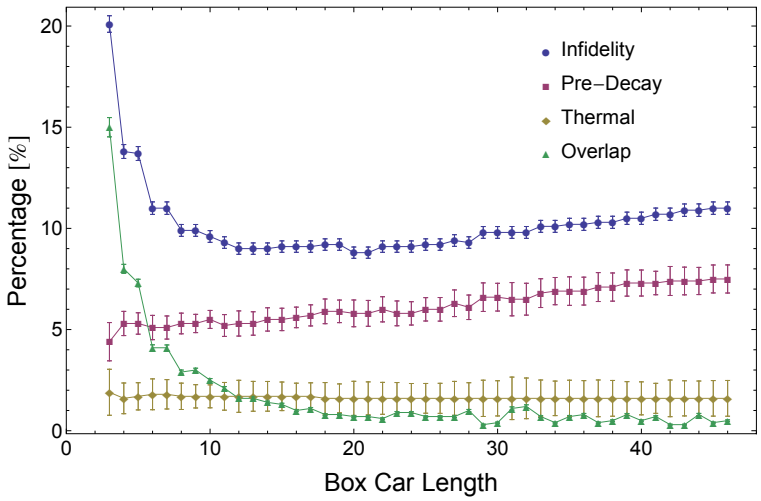

Figure 11: Optimized start time and box-car length. Same procedure as in Figure 10A, however this time the integration start time is variable for each box-car length. Data taken from Chip 2.

What is not seen in the Figure is the trend of the integration start time. Short box-car filters start later, since these points have the largest separations in time (see Section 4.5), while larger box-car filters start earlier. The two point box car begins roughly 100 ns after the measurement tone begins, while the 45 point box-car starts at the beginning of the measurement tone. The optimal filter, in this case, is the 20 point box-car beginning 50 ns after the measurement tone. This suggests that, for this sample, it takes roughly 50 ns for the mean of the ground and excited state traces to separate far enough apart to exceed the noise level and be useful. Integration prior to this will result in more noise than signal and explains why integration should not begin exactly when the measurement pulse is applied.

3.5 Measurement Power and Parametric Amplifier Gain

The final two practical parameters to optimize are the measurement drive power and the amplifier gain. One expects that increasing the drive power, results in larger output signal and therefore a better signal to noise ratio. This should result in a reduction of the Overlap term, as well as possibly, slightly earlier integration start time, leading to

a small reduction in the Pre-Decay term. However, as the power is continued to increase, at some point we expect to drive transitions between the states causing an increase in the Pre-Decay and Thermal terms. This may also result in a large increase in the amount of data filtered out during the post-selection process. These expectations will be better understood in Section 4.6. However, from a practical standpoint one simply increases the power until the fidelity drops, or one is unhappy with the amount of data they must filter out. The results of such an optimization are shown in Figure 12A and support our expectations.

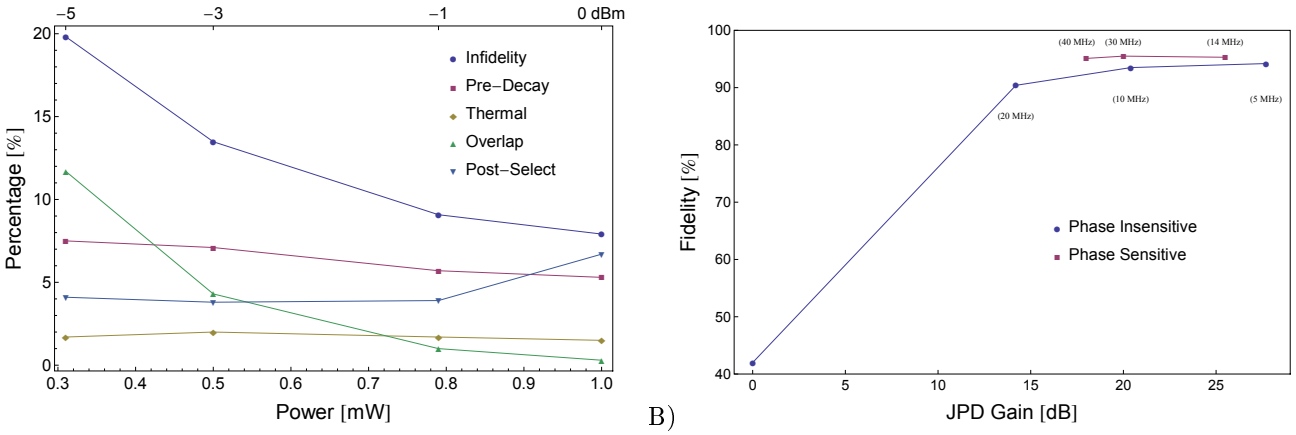

Figure 12: Optimizing the measurement drive power. A) Single-shot results after 16 point box-car filter. B) Fidelity versus JPD gain and JPD amplification mode. The numbers in parenthesis are the respective bandwidth of the gain peak. Data acquired using Chip 2.

As for the amplifier gain, we expect that an increase in gain should also improve our signal to noise ratio, causing the standard deviation of the data points to reduce, thereby also decreasing the Overlap term and improving the fidelity. From theory, however, we expect that there is a gain-bandwidth product that will be constant and will limit the maximum amount of gain possible [5]. Also in this theory, we expect that the phase sensitive mode of amplification should outperform the phase insensitive mode by at least a half photon of noise. The ramifications of the extra noise in the phase in-sensitive mode will be better understood in Section 4.3. The results shown in Figure 12B substantiate these expectations. One can also see that once a certain level of amplification is reached, further increases do not significantly affect the fidelity. This would be the point were the overlap term has roughly reached 0% and one is completely limited by the Pre-Decay and Thermal terms. For simplicity, we used the phase insensitive mode with roughly 20 dB gain unless otherwise stated.

4 Qubit/Resonator Optimizations

The practical optimizations are essential, but they offer little insight into how various design parameters influence the measurement outcome. Usually, the best one can say is that decay dominates infidelities, and so improvements of the qubit lifetime need to be made. Theory, however, can give us more insight into how qubit specific parameters

may affect our measurements. In this section, I will first introduce the various system parameters relevant for optimization, I will then introduce the signal-to-noise ratio (SNR) and follow this with an in-depth analysis of each parameter of the SNR focusing on their limitations and relations.

4.1 Qubit Parameter Introduction

The first system parameter is the resonator decay rate, κ. The total decay rate of the resonator is given by the sum of the two ports[9], κ_{in} and κ_{out}, of the resonator which are depicted in Figure 1

$$\kappa = \kappa_{\text{in}} + \kappa_{\text{out}} \tag{6}$$

We will not consider other decay channels in the theory, assuming they are much smaller than these two ports. All other parameters discussed will follow the Transmon review paper of Koch et. al. [13] and values without subscripts, imply the ground state value.

The next system parameter is the qubit resonator coupling. Each level couples differently to the resonator and is generally given by the formula

$$g_{\text{ij}} = g_{\text{ji}} = \sqrt{i+1}g \tag{7}$$

where g is the ground to first excited state coupling to the resonator and $i < j$, are the qubit levels. As stated previously only nearest level coupling is allowed so $j = i + 1$. With this, the partial dispersive shifts are defined as

$$\chi_{\text{ij}} = \frac{g_{\text{ij}}^2}{\Delta_{\text{ij}}} \tag{8}$$

Due to Transmon level repulsion between neighboring states, the bare dispersive χ is given by

$$\chi = \chi_{01} - \frac{\chi_{12}}{2} = \frac{g^2}{\Delta}\left(\frac{\alpha}{\Delta+\alpha}\right) \tag{9}$$

The qubit detuning parameter Δ_{ij} is then given by

$$\Delta_{\text{i,i+1}} = \omega_{\text{i,i+1}} - \omega_{\text{res}} \tag{10}$$

where $\omega_{\text{i,i+1}}$ is the qubit transition frequency between the i^{th} and $i^{\text{th}} + 1$ state. For simplicity we will assume that $\Delta_{12} = \Delta + \alpha$ where α is the qubit anharmonicity and $\alpha < 0$.

There is a final parameter called the readout detuning, δ, and in our case we will define it from the renormalized resonator frequency

$$\delta = \omega_{\text{readout}} - \omega_{\text{renorm}} \tag{11}$$

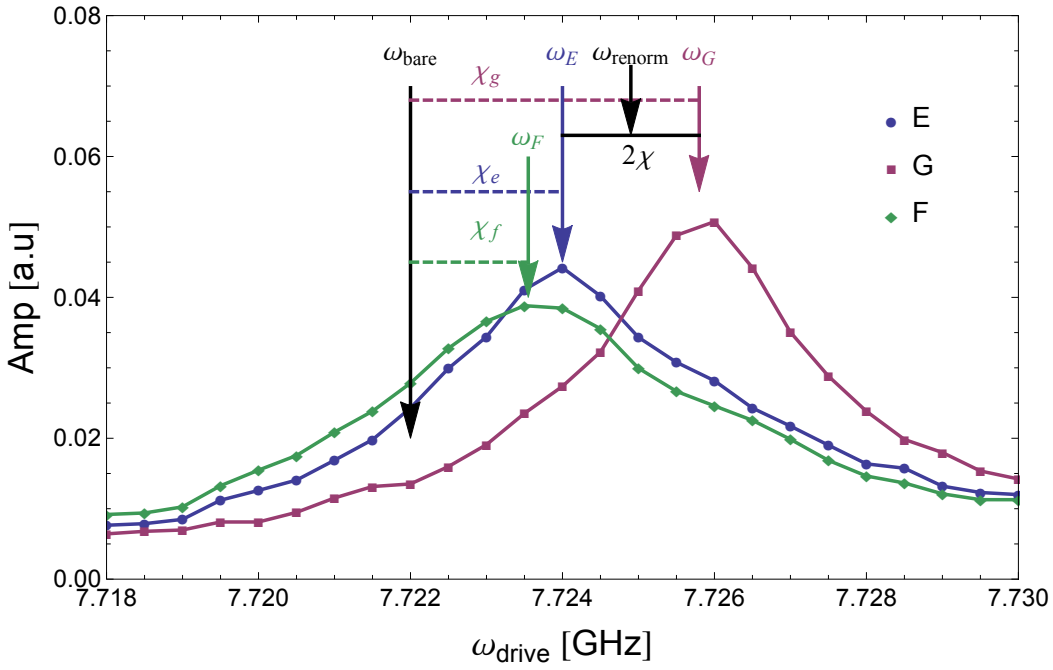

Figure 13: Dispersive shifts of the resonator due to the state of the qubit. χ_{i} are individual dispersive shifts of each Transmon level with respect to the bare resonator frequency. χ, however, is defined between the ground and excited state shifts and the renormalized resonator frequency. The lines simply connect the data points and are not fits. Each point is an average of over 4,000 thousand runs. More details in the text.

These equations are approximations derived within certain limits. The limits are given by $E_{\mathrm{J}}/E_{\mathrm{C}} \gg 1$, $g/|\Delta| \ll 1$ and $g/|\Delta_1| \ll 1$, where E_{J} is the Josephson energy and E_{C} is the charging energy of the qubit.

An experimental result is shown in Figure 13, where the amplitude of the output as a function of the input signal frequency is shown for the first three levels of the Transmon. Each data point is averaged from over 4,000 runs. The data was taken from Chip 1 and the bare resonator frequency was calculated since one can not turn off the coupling, g. The shift of the resonance frequency of the resonator is referred to as the dispersive pull or shift, χ, due to the presence of the qubit. Various definitions of this dispersive shift can be found in literature and two are shown in the figure. χ_i is defined from the bare resonator frequency and is a dispersive shift unique to each Transmon level. While in the context of a pure 2-level qubit a bare χ is defined from the renormalized qubit frequency in the middle of the resonance peaks of the ground and excited state given by Equation 8.

4.2 The Signal to Noise Ratio

So far, we have only used the fidelity as a measure for the quality of a measurement protocol. However, another useful metric is the signal to noise ratio (SNR). As described in Gambetta et. al. [11] the fidelity is a function of SNR. In the simplest case of no qubit decay the relationship is given by

$$F = \mathrm{erf}\left(\sqrt{\frac{\tau_{\mathrm{int}} SNR}{2}}\right) \tag{12}$$

where erf is the error function and $\tau_{\text{int}} = \frac{t}{T_1}$ is a dimensionless integration time normalized to the qubit decay time. The SNR is the ratio of the integrated signal power to the integrated noise after time τ_{int}. One can interpret the fidelity as an integrated SNR and it should be a monotone of SNR as well. Therefore, we can focus on maximizing the SNR knowing we are simultaneously maximizing the fidelity. Also, presented in the Gambetta paper is a simulated result of the necessary SNR after an optimal integration time and box car filtering for a qubit with decay. This provides a rough estimate of a target SNR for a desired level of fidelity because the simulation is closely related to our measurement protocol. The simulated results are shown in Figure 14 and taken directly from the paper. We found these numbers roughly aligned with our results.

The SNR defined by Gambetta et. al [9] is given by

$$SNR = \frac{\eta\kappa \left|\beta\right|^2 \cos^2(\theta_\beta - \phi)}{\gamma_1} \tag{13}$$

where η is the measurement efficiency, β the state discrepancy, $\gamma_1 = 1/T_1$ is the qubit decay rate, θ_β is the optimal rotation angle in the (I,Q) plane discussed in Section 2.4, and ϕ is the actual angle rotated. Each term will be discussed in the following sections, however, for simplicity we assume $\phi = \theta_\beta$ so that the cosine term is just 1. As we will see, this definition has all of the essentials necessary to model the measurement process and it will be the guiding equation for optimizing the system for speed.

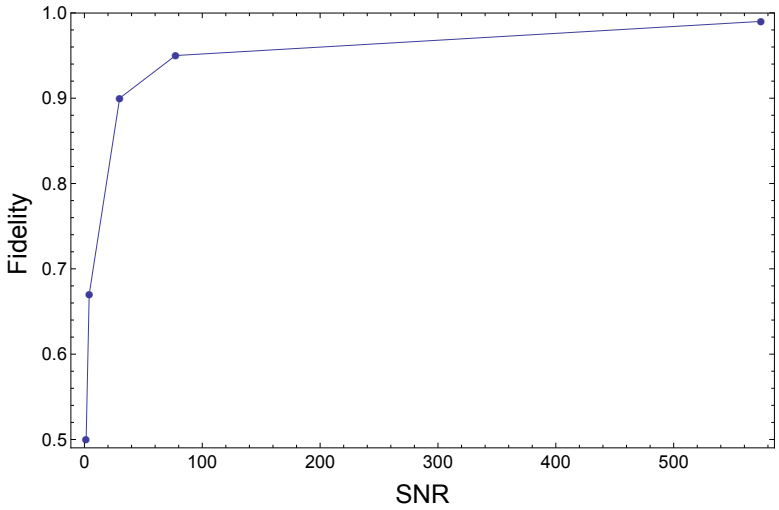

Figure 14: Fidelity versus SNR as presented in Gambetta et. al [11]. These are the results of a simulation for a qubit with decay and optimal box-car filtering.

4.3 η, The Detection Efficiency

The first term of the SNR is quite straightforward. The equation is given by

$$\eta = \frac{\kappa_{\text{out}}}{\kappa_{\text{in}} + \kappa_{\text{out}}} \frac{1}{N_{\text{noise}} + 1} \tag{14}$$

where N_{noise} is the effective number of noise photons in the measurement. The first term is related to the percentage of signal photons one can possibly observe. The model suggests that once a photon has entered the resonator, it has only two possible paths to exit. If we only measure at the output port, then in the limit $\kappa_{\text{out}} \gg \kappa_{\text{in}}$ this value approaches the maximum of 1. This term, for example is what significantly limits trapped ion experiments because the percentage of signal photons actually detected is small due to the tiny solid angle of the detector. One

dimensional super-conducting circuits do not suffer from this and the large limit is easily realizable. Reflection measurements are also a possibility, however this provides additional challenges in the experimental realization. For example, we attempted a few experiments measuring in reflection, but our circulators which isolate the input and output signal of the resonator from the input of the JPD were not strong enough. This caused a large increase in the amount of noise photons at the input of the JPD.

There is, however, concern with just how crude this factor is. In this asymmetric limit one must increase the drive power to maintain the same photon flux in the resonator as compared to the symmetric case. If there is any cross-coupling between the input and output ports, the increase in power at the input port will also increase the number of noise photons at the output port and increase our N_{noise}. Nevertheless, we used highly asymmetric resonators in our experiments. Further investigation of this cross coupling noise and resonator asymmetry should be pursued for a greater understanding of this term.

The second term of η is the efficiency of the amplification chain. The noise figure, defined as $\mathcal{F} = SNR_{\text{input}}/SNR_{\text{output}}$, for each microwave component describes the degradation of the SNR as it passes through the component. Cascaded components lead to a total noise figure give by

$$\mathcal{F}_{total} = \mathcal{F}_1 + \frac{\mathcal{F}_2 - 1}{G_1} + \frac{\mathcal{F}_3 - 1}{G_1 G_2} + \cdots \tag{15}$$

where $\mathcal{F}_{\text{i}}$ is the noise factor at the i^{th} stage of amplification and G_{i} is the linear scale (not decibel) gain of the i^{th} amplification stage [18]. This formula basically states that the first amplifier is the most important for the effective noise level of the system, i.e. N_{noise}.

With Equation 14 and 15 the results presented in Figure 12B can be further analyzed. We measured the effective noise level of our systems by taking power spectral density measurements of a coherent signal with and without the JPD turned on [8]. With the use of AC stark measurements we are able to convert drive power to photon flux [21]. The results show that there are roughly 3-4 noise photons in the Kelvinox and 4-5 photons in the Triton setup. Without the JPD on, making the HEMT the first amplification stage, we found more than 140 noise photons. This explains the large gap between the fidelity obtained with the JPD off and on in Figure 12B. It also explains why the phase sensitive mode, which is a half noise photon more efficient then the phase insensitive mode [5], results in a slightly higher overall fidelity.

This suggests that there is some room for improvement in our experimental setup to reduce the noise level at the input of the JPD by a few photons, since the theoretical limit is zero noise photons [5]. G. de Lange et. al have reported as good as 1 to 2 noise photons [6], but due to this parameter's relationship with SNR and the room for improvement, we would not expect a drastic increase in our fidelity or our measurement speed.

4.4 κ and γ_1, Purcell Decay

The linear relationship between κ and SNR is a bit misleading, as well as the notion that faster decay out of the resonator would lead to a faster measurement. First, one must remember that a larger κ will also result in a larger Purcell decay. The Purcell decay rate is given by [13]

$$\frac{1}{T_1} = \gamma_1 \leq \gamma_{\text{purcell}} = \kappa \frac{g^2}{\Delta^2} \tag{16}$$

If one is in a regime where the Purcell decay limits the T_1 time, substitution of Equation 16 into Equation 13 will cancel the linear term of κ in the SNR. SNR does not become completely independent of κ, however, because it will appear again in the β term as we will see in the next section. As will be shown, κ is a damping term which affects the time of peak SNR and therefore does impact the measurement time.

This Purcell decay is a large limiting factor of the SNR and this explains why the use of a Purcell filter, which has shown to increase the κT_1 product by a factor of 100, directly correlates to improved fidelity [20]. This allows one to get a significantly higher fidelity with a qubit which might have otherwise been quite bad. Sank et. al. show that the filtered Purcell decay becomes

$$\gamma_{\text{filtered}} = \frac{g^2}{\Delta^2} \left(\frac{\omega_{\text{qb}}}{2\Delta Q_{\text{f}}} \right)^2 \tag{17}$$

where Q_{f} is the quality factor of the filter. Q_{f} is relatively independent of the qubit parameters allowing one to greatly improve readout without affecting the qubit.

4.5 β, The State Discrepancy

All of the previous terms have been relatively independent of the actual qubit and resonator system, however, β is the term that holds all of these dynamics. It is defined as

$$\beta \equiv \mathcal{C}_{\text{e}} - \mathcal{C}_{\text{g}} \tag{18}$$

where $\mathcal{C}_i$ is the coherent state of the resonator field when the qubit is in state i [9]. As stated previously, this is the difference between the ground and excited state means in the (I,Q) plane.

In order to find a theoretical prediction of β, one starts with the Master equation of the system [9]

$$\dot{\rho}(t) = -\frac{i}{\hbar}[H_{\text{eff}}, \rho(t)] + \kappa \mathcal{D}\left[a\right]\rho(t) + \gamma_1 \mathcal{D}[\sigma]\rho(t) + \gamma_\phi \mathcal{D}\left[\sigma_z\right]\rho(t)/2 \tag{19}$$

where the first term describes the evolution of the system under the effective Hamiltonian, the second term is damping due to photon decay out of the resonator, the third is related to qubit decay and the last is related to qubit dephasing, γ_ϕ. $\mathcal{D}$ is the damping super-operator defined as $\mathcal{D}[X]\rho = X\rho X^\dagger - X^\dagger X\rho/2 - \rho X^\dagger X/2$. In order

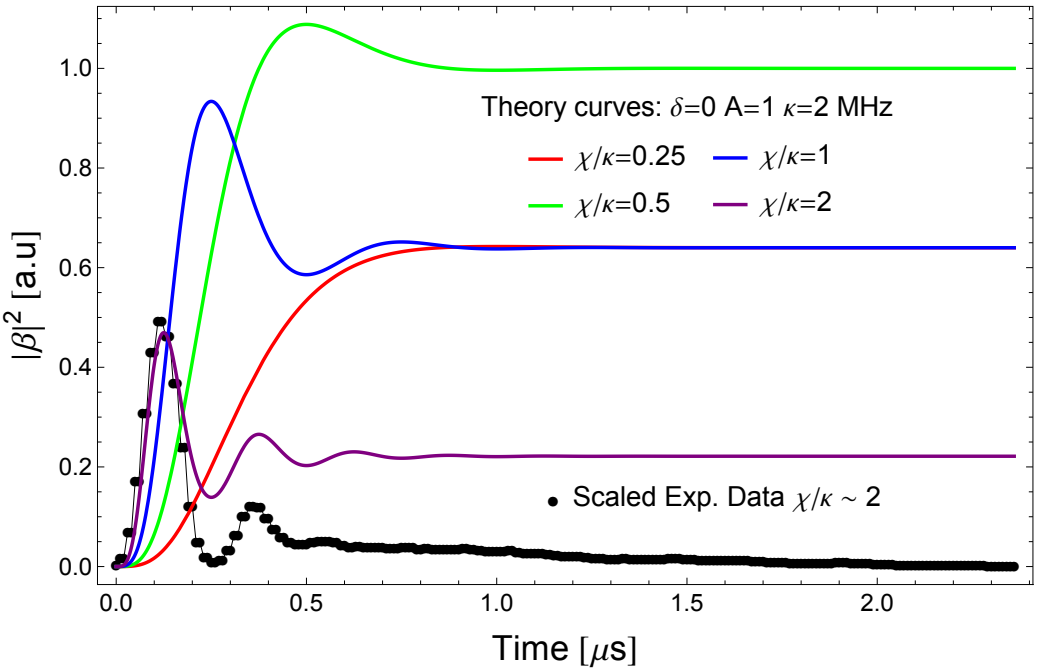

Figure 15: $|\beta|^2$ as a function of time for different ratios of χ/κ. The black points are calculated from Chip 2 data after a 20 point-box car filter and finding the difference between the mean of the ground state traces and the mean of the excited state traces. The experimental data was then scaled to the first maximum of the theory curve with similar values which was roughly $\chi/\kappa = 2$.

to solve this analytically, one must simplify the equivalent coupled differential equations by ignoring state coupling due to decay, γ_1, greatly simplifying the dynamics and resulting in the two coupled differential equations

$$\begin{aligned}\dot{\mathcal{C}}_{\mathrm{e}}(t) &= -i\epsilon_{\mathrm{drive}} - i\,(\delta+\chi)\,\mathcal{C}_{\mathrm{e}} - i\frac{\kappa}{2}\mathcal{C}_{\mathrm{e}} \\ \dot{\mathcal{C}}_{\mathrm{g}}(t) &= -i\epsilon_{\mathrm{drive}} - i\,(\delta-\chi)\,\mathcal{C}_{\mathrm{g}} - i\frac{\kappa}{2}\mathcal{C}_{\mathrm{g}}\end{aligned} \tag{20}$$

where $\epsilon_{\mathrm{drive}}$ is the measurement drive rate. This rate is related to the drive power as discussed in the next section.

For a simple analytic solution to Equation 20, one can use a step function for the drive term $\epsilon_{\mathrm{drive}} = A\Theta\,(t-t_0)$ where A is the amplitude of the pulse and Θ is the Heaviside step function. This results in

$$\beta = 2A\left(\frac{-1+e^{-2i\pi t(\delta+\chi)-\pi t\kappa}}{(\delta+\chi)-i\kappa/2} - \frac{-1+e^{-2\pi it(\delta-\chi)-\pi t\kappa)}}{(\delta-\chi)-i\kappa/2}\right) \tag{21}$$

which is roughly the difference between the solutions to two damped harmonic oscillators with resonant frequencies $\delta \pm \chi$ and damping factor κ. Now one can see why the ground and excited state means separate in time when the measurement pulse is applied. The drive acts as an external force on the oscillatory system and the resulting oscillations depend on the state of the qubit. The qubit acts as an external control to the relative harmonic frequency of that oscillator, by either changing its relative mass or spring constant.

This equation is plotted in Figure 15 with selected parameters and various χ/κ ratios. The first thing to notice

about the plots are the non-decaying steady states. As a result of neglecting the qubit decay, in order to solve the master equation analytically, we inevitably lost this feature in the results. Realistically, the excited state will always decay back to the ground state which should result in $\beta \to 0$ as $t \to \infty$. For comparison, an experimental result of $|\beta(t)|^2$ with roughly the same χ, κ and δ is also plotted. The experimental data is rescaled to have approximately the same amplitude as the equivalent theoretical curve at the first overshoot. The experimental data shows the expected decay of β over time due to excited state decay. β is calculated from the data by taken the difference between the mean of the excited state and ground state traces after filtering.

When χ/κ is one quarter the system acts like an "over-damped" oscillator, with an exponential rise to the steady state without any oscillations. χ/κ of one half is analogous to the "critical damped" case where beta reaches its maximum value and there is a single over shoot before returning to the steady state. Reducing κ further leads to the "under-damped" cases where significantly more oscillations occur before reaching equilibrium. Though the peak values again decrease from the "critical-damped" case, one should notice that the large overshoots happen earlier, signifying the possibility for a more rapid measurement. Also, one should not forget that A is identical for each case, but it is not yet clear if this is a valid assumption.

4.6 A, The Drive Amplitude

The most important thing to see in Equation 21 is that the drive amplitude, A, becomes a scaling factor. This is also the case if one uses a more realistic pulse, for example an exponentially rising pulse. This, in turn, implies that $SNR \propto A^2$ and suggests that by increasing the drive power, one can achieve arbitrary fidelity. This warrants further investigation in finding ways to maximize this value in order to achieve large increases in the SNR.

The drive amplitude is related to the steady state, mean photon flux of the resonator by

$$\bar{n}_{\pm} = \frac{A^2}{\frac{\kappa^2}{4} + (\delta \pm \chi)^2} \tag{22}$$

for the qubit in the excited (+) and ground (-) state [10]. Larger amplitude, implies a greater number of photons and therefore a larger signal, β.

However, there are limitations on A imposed by the system. The first limitation is due to state mixing. If one solves for the eigenstates of the dispersive Hamiltonian, Equation 2, the mixing coefficients of the dressed states become proportional to $|e\rangle = \cos\theta\,|1,n\rangle + \sin\theta\,|0,n+1\rangle$ and $|g\rangle = -\sin\theta\,|1,n\rangle + \cos\theta\,|0,n+1\rangle$, where $|e,g\rangle$ are the eigenstates labeled ground and excited in the single-shot measurement, $|i,n\rangle$ are the coupled Transmon levels and resonator Fock states, and $\theta = \frac{1}{2}\tan^{-1}\left(\frac{2g\sqrt{n+1}}{\Delta}\right)$ [4]. When the argument of $\tan^{-1}$ goes to infinity, one finds maximal entanglement between the coupled systems. At this point one can no longer disentangle the information about the state of the qubit and the resonator state, making a qubit state measurement protocol meaningless. This limit is reached as $\Delta \to 0$, which is one of the limits of the dispersive regime, i.e. $g \ll \Delta$. One can also see that as

n increases, the effective mixing increases. One will often find in literature the critical photon number defined as [10]

$$n_{\text{crit}} = \frac{\Delta^2}{4g^2} \tag{23}$$

Within the dispersive limits, one finds that χ is reduced from its predicted value in Equation 9 by a continuous factor relative to n/n_{crit} as shown in Figure 16.

The dispersive Hamiltonian of Equation 2 does not account for mixing with higher levels of the Transmon and one should expect similar mixing as discussed previously between other states at different values of Δ. If one looks at the form of χ there are effectively two infinite limits where it is likely to be unrealistic. They occur at $\Delta = 0$ and $\Delta = \alpha$. The first limit was discussed previously and the second is indicative of mixing between the $|e\rangle$ and $|f\rangle$ levels which would include terms like $a|1, n+1\rangle + b|2, n\rangle$. One can either, numerically solve the Jaynes-Cummings Hamiltonian in Equation 1 to see this mixing, or do master equation simulations. Indeed, in Figure 17 one can compare an experimental result of this e-f mixing of Chip 3, when it was tuned to roughly $\Delta \approx \alpha$, and compare it to the master equation simulation. The simulation strongly suggests e-f mixing. In this case, we expect the $|e\rangle$ level to be on the right side of the ground state and the $|f\rangle$ level on the left. This mixing could then cause problems with state identification and therefore our fidelity.

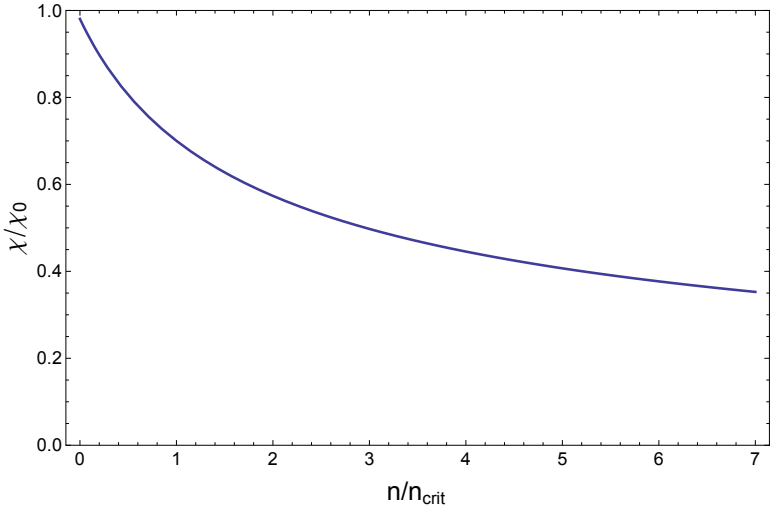

Figure 16: Effective χ with increasing photon flux. $\chi_0 = \frac{g^2}{\Delta}\left(\frac{\alpha}{\Delta+\alpha}\right)$ and $n_{\text{crit}} = \frac{\Delta^2}{4g^2}$

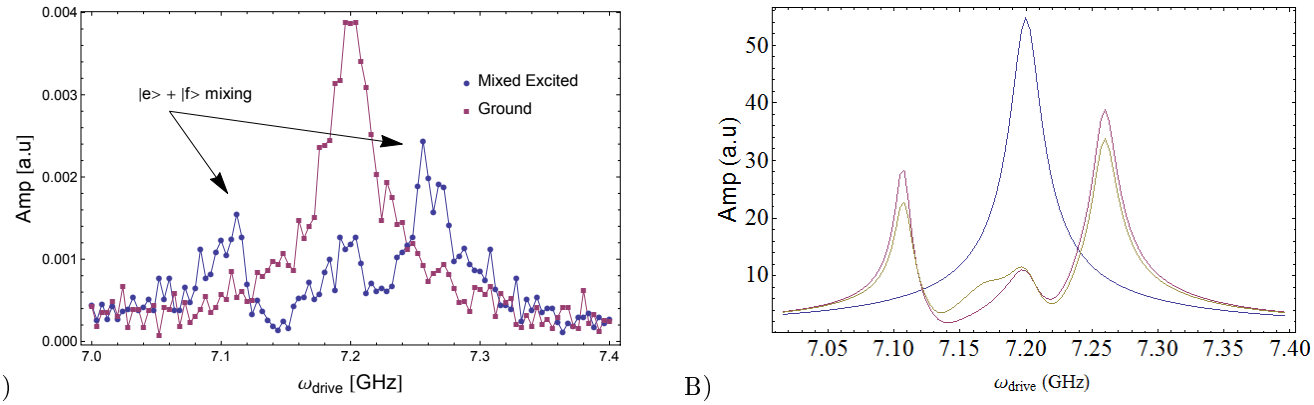

Figure 17: e-f mixing of states in the limit $\Delta \approx \alpha$ A) experimental results from Chip 3. B) master equation simulation results with identical system parameters and full Jaynes-Cummings Hamiltonian. The data is from Chip 3.

The state mixing problem can be avoided completely by simply remaining within the limits of the dispersive regime and selecting parameters such that n_{crit} is large. However, there is one last limitation to the drive amplitude that is the most constrictive for us. This is the dynamic range of the JPD [8]. As stated previously, the amplification provided by the JPD is crucial for obtaining a low effective N_{noise} in Equation 14. At some point, every amplifier will saturate and the input to output relation will no-longer be linear. In Figure 18 the dynamic range, marked by the 1 dB compression point, of the JPD in the Triton is plotted. The 1 dB compression point is not a hard limit, but it is clear that above 11 photons, the output gain is drastically reduced and this will lead to a much larger N_{noise}. The conversion from generator power to photon number was obtained from AC-Stark measurements [21].

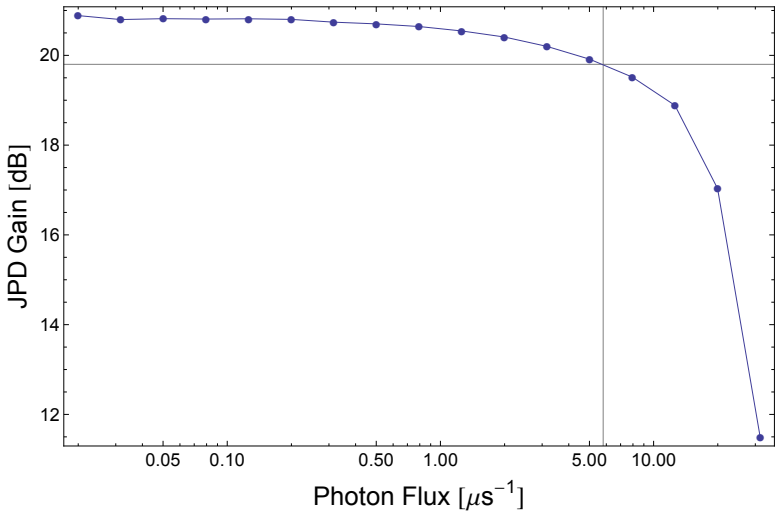

Figure 18: JPD dynamic range. The 1 dB compression point is depicted by the vertical line.

As discussed by Eichler et. al. [8], the dynamic range could be improved in the JPD by increasing the number of SQUIDs in the SQUID arrays. However, even if we removed these two constraints discussed on the drive power, there would always be limitations set by the physical properties of the resonator. At some point the resonator will become non-linear as the power increases, but in our experiments this limit was never reached.

5 Rapid, High fidelity, Readout

In this final section I will discuss the results for optimizing the system parameters for a rapid, high fidelity, readout, as well as a pulse shaping technique for increased measurement speed. I will then conclude with a summary and outlook for this project. ...

5.1 Results From Qubit Optimization

With all of the considerations of the previous sections combined, we have powerful predictive capabilities. Figure 19 shows the plot of the theoretical SNR as a function of time of Chip 2. The drive amplitude, $A/2\pi = 20\,\text{MHz}$ is determined from the JPD threshold of around 11 photons, while η is set to 0.2 to align with our noise measurements. This plot is in good agreement with the best measured fidelity of $>95\%$ in roughly 160 ns of Chip 2 (phase-sensitive amplifier). The graph also shows the peak separation of the two states will occur around 100 ns, also agreeing with

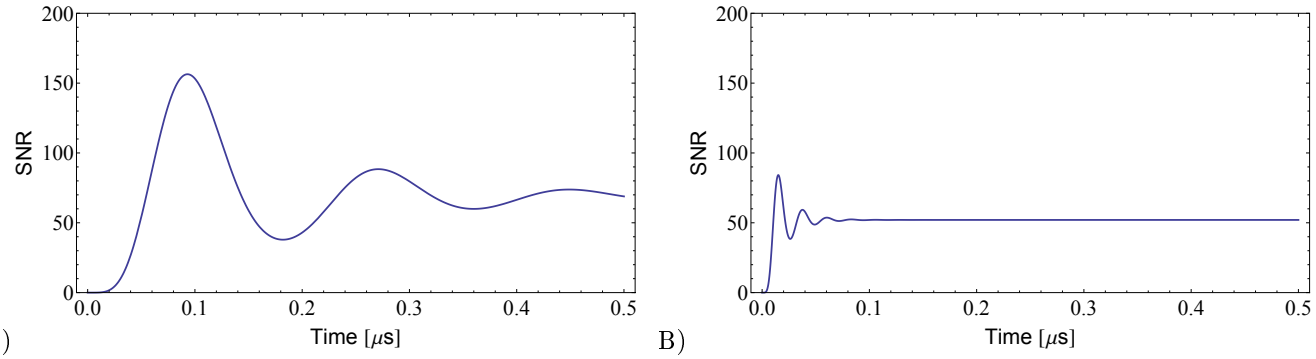

Figure 19: Theoretical SNR calculated from Equation 13 A) using Chip 2 parameters and A/2π=20 MHz and η=0.2 B) using Chip 3 parameters and A/2π=140 MHz and $\eta = 0.2$.

the experimental results in Figure 15.

Agreement between theory and experiment can be further supported by extracting the SNR as a function of time from the single-shot measurements. This is quantitatively difficult in Equation 13 because of the conversion between the FPGA recorded voltages and the theoretical amplitude of β. The experimental SNR can be defined as

$$SNR_{\exp} = \left(\frac{\mu_e - \mu_g}{\sigma_e + \sigma_g}\right)^2 \qquad (24)$$

and it can be converted into the SNR used in Figure 14 by multiplying by the noise bandwidth, B, and the qubit decay time T_1 [22].

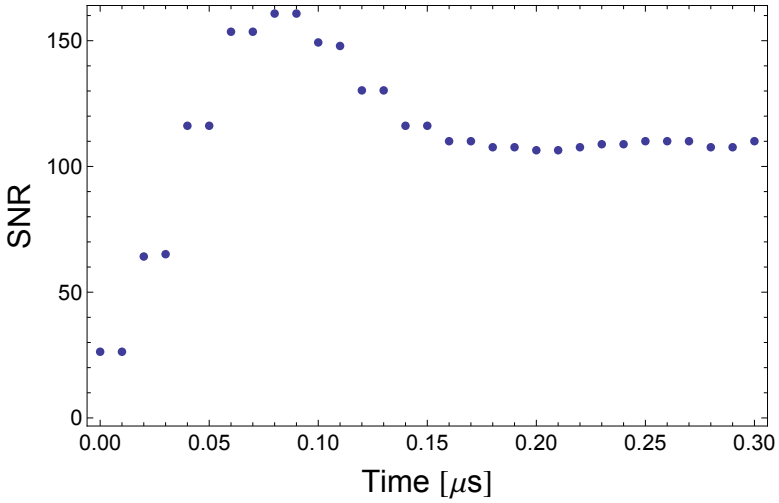

Figure 20: Experimental SNR. The experimental SNR defined in Equation 24 is plotted for Chip 2 data. The points are in good quantitative agreement with the theoretical results calculated from Equation 13.

The noise bandwidth includes the filter and the parametric amplifier bandwidths. This value of SNR is calculated as a function of time for the data from Chip 2 and is in good quantitative agreement with the theoretically calculated values in Figure 19A.

With confidence in the model and knowledge of the constraints of the variables, one is in a position to tune the variables for optimal readout and speed. With all of the equations combined one can use an algorithmic optimization routine to get an idea of how to maximize the SNR in a given amount of time. The results of the maximization routine will always tend towards the limits imposed by α and $\Delta \to \alpha$. To understand this we recall the SNR$\propto A^2$. Given that the power is limited by the input of JPD to roughly 10 photons, and from Equation 22 one realizes that they must maximize $(\delta \pm \chi)^2$ to increase A.

If $\delta \neq 0$, then the photon flux becomes asymmetric and the number of photons in the resonator will be dependent on the state of the qubit. For example if $0 < \delta < \chi$, then one will successfully increase $(\delta + \chi)^2$ and achieve a larger A. Unfortunately, it will decrease $(\delta - \chi)^2$, which will limit A more, and because we must use the same pulse for both states, $\delta = 0$ becomes the optimum and one must maximize χ. Looking at the form of Equation 9 we see that maximizing α while $\Delta \to \alpha$ leads to $\chi \to \infty$. As stated previously this is the regime of dispersive breakdown and state mixing. If we impose a further constraint that $|\Delta + \alpha| \gg g$, then the routine will still tend toward this limit and we are confident that within the straddle regime, $0 < \Delta < \alpha$, a globally optimal χ can be found. When Δ is constrained to regions outside of the straddle regime, the optimization routine always finds an optimum SNR that is worse, as seen in Figure 21.

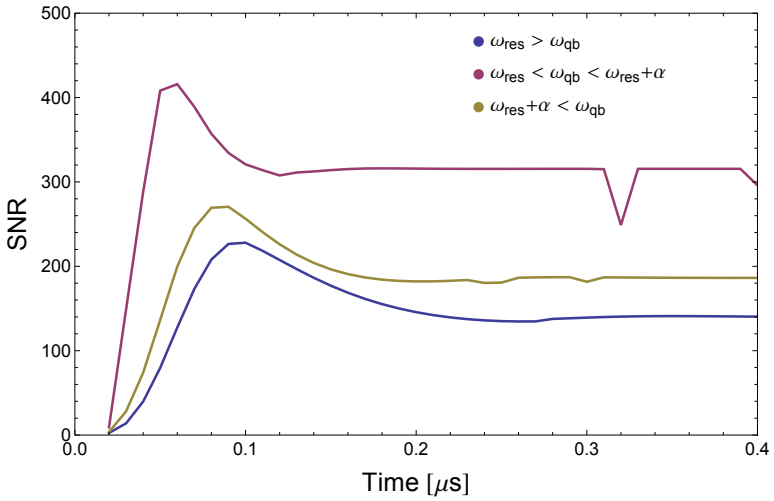

Figure 21: Optimized SNR of the three detuning regimes. The plots are the results of the optimization routine when limited to the three detuning ranges. The straddle regime, red curve, always finds parameters which produced the largest SNR results. A T_1 time of $3\,\mu$s was assumed given that the Purcell decay time was constrained to be above $3.5\,\mu$s.

α will be fixed based on Transmon limits, as well as the desired transition and resonator frequencies [13]. Specifically, the transition frequency scales proportional to $\sqrt{E_C E_J}$ while $\alpha \approx E_C$ and we should maintain $E_J/E_C \gg 20$. We desired a resonator around 8 GHz, and conservatively limited ourselves to roughly $\alpha = -480\,$MHz.

Once α, δ and Δ are fixed, one must optimize g and κ. This can be done by using the remaining constraints of the Purcell decay, n_{crit}, and the JPD dynamic range. In this regime it was easy to find parameters with a large n_{crit}, but it is more difficult to keep the Purcell decay rate and the photon flux within their limits. Therefore, these two limits are more constraining than was n_{crit}.

The results of the optimizing procedure are shown in Figure 22 and the values in The Qubit Parameter Table of Section 2.2. One can see from the curve that we expect an even greater fidelity than in Chip 2, in only 40 ns. Also, if not restricted by time, one can see that the Purcell limit is similar to that of Chip 2 and therefore we expect to be able to integrate over at least 150 ns or more, allowing for the possibility of fidelity greater than 99%. Furthermore, n_{crit} is well above the 11 photon limit imposed by the JPD and the mixing coefficients between the levels are calculated to be small. Specifically, 3 % mixing between $|g\rangle$ and $|e\rangle$, and approximately 20 % mixing between $|e\rangle$ and $|f\rangle$ are expected with a photon flux of ten. In comparison, Chip 2 has a calculated 15 % mixing between $|g\rangle$ and $|e\rangle$ and 17 % mixing with $|e\rangle$ and $|f\rangle$. As explained previously the mixing probabilities are numerically calculated by solving Equation 1.

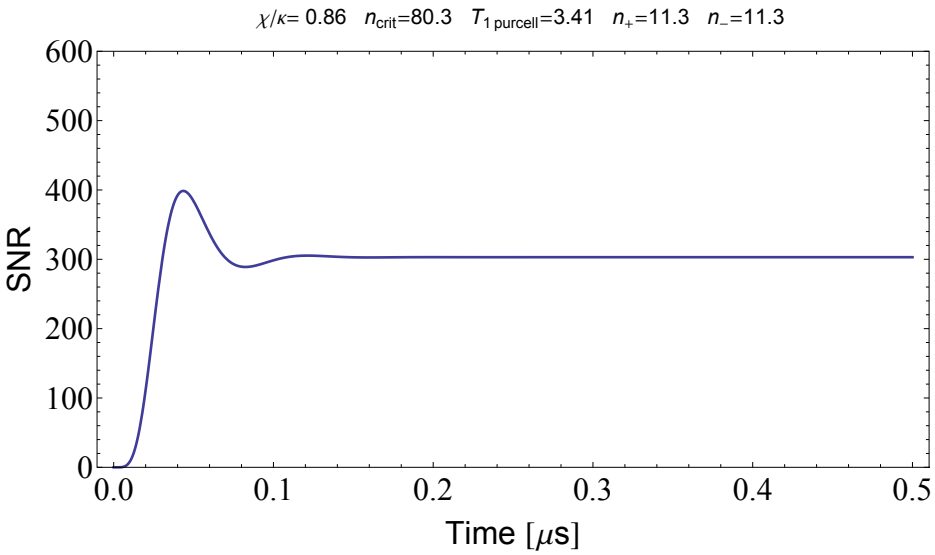

Figure 22: The theoretical SNR of the optimized qubit parameters. The parameters were found using the optimization routine. The theoretical χ/κ ratio, n_{crit}, the Purcell decay time and the photon flux of the resonator, $n_{\pm}$, are displayed at the top of the graph.

The optimized values are within the dispersive regime and we are confident the results are experimentally reproducible. We tested the theory on Chip 3 because it was already made for a previous experiment in the lab and it has values similar to our optimized ones. The theory plot in Figure 19 shows that we do not expect a high fidelity, but we expect that the maximized fidelity to be reached rapidly. The results of the single-shot measurements are shown in Figure 23. In this case, it only took 40 ns (4 point box-car) to produce the results and drive the Overlap term to 1%, successfully accomplishing a rapid readout. We can see the Infidelity is completely dominated by Pre-Decay, however, this is expected as the chip's parameters have a higher g and κ than our optimized one. This causes a huge decrease in the T_1 time due to Purcell decay which is calculated to be 450 ns. The measured T_1 is roughly 350 ns in agreement with theory.

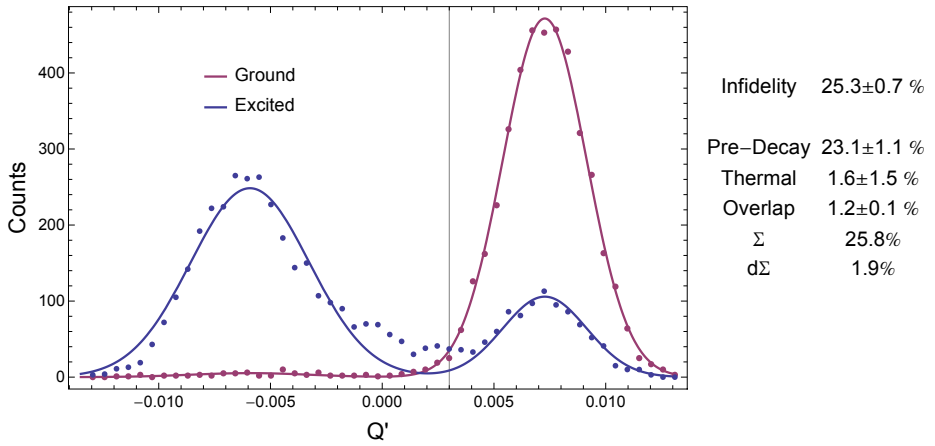

Figure 23: Single-shot measurement results of Chip 3. Only 40 ns of integration were necessary to reach this maximum fidelity.

Also, due to the increase of g, we are closer to the dispersive breakdown limit $g \ll \Delta + \alpha$ causing an increase in mixing, calculated to be 25 % between $|g\rangle$ and $|e\rangle$, 20 % between $|g\rangle$ and $|f\rangle$, and 36 % between $|e\rangle$ and $|f\rangle$. This also assumes a photon flux of 10. This mixing explains the large amount of data points off of the excited state Gaussian fit in the middle between the ground and excited state and on the left side as discussed in Section 2.4. Some of the offset in the middle is due to decay during integration, but it is likely more dominated by the mixing. One can see however, that due to the nature of the thresholding procedure, that even this large amount of mixing does not completely destroy the readout fidelity. The mixing is still small enough to keep a large portion of the excited state data points on the correct side of the line.

5.2 On-Hold Pulses

As a final attempt to improve measurement speed, we adjusted the measurement pulse shape. As depicted in Figure 4 our measurement pulses were approximately square pulses with constant amplitude. We then changed the pulse to have two sections, the first is tall and narrow, followed by a second longer weak pulse with a ratio of their amplitudes set to $t_{\text{pulse1}}\sqrt{\chi^2 + \kappa^2/4}$ where t_{pulse1} is the length of the first pulse. This shape is referred to as an On-Hold pulse [12]. This new pulse scheme is shown in Figure 24. The On-Hold pulse attempts to populate the resonator with photons, up to the desired photon flux, as fast as possible and then maintain this flux for the duration of the measurement. According to the theory of the previous section this new ϵ_{drive} will cause the initial rise of β to be steeper and slightly earlier than with the single square pulse. We therefore expect to see an improvement in the integration start time, as well as possibly a decrease in the Pre-Decay due to less decay occurring prior to integration.

We tested the On-Hold pulse on a fourth chip not included in Section 2.2. The qubit was not a standard Transmon, but a Tunable Coupling Qubit (TCQ) described in Srinivasan et. al [23], however, the results are given in relative terms of the system parameters and should be independent of the qubit type. This qubit had a much shorter lifetime than Chip 1 or 2, only 1.4 μs, which would make improvements due to the On-Hold pulse more apparent. Our single-shots without pulse shaping had an optimized integration start time around $0.8/\kappa$ and there was roughly 20% Pre-Decay with a total in-

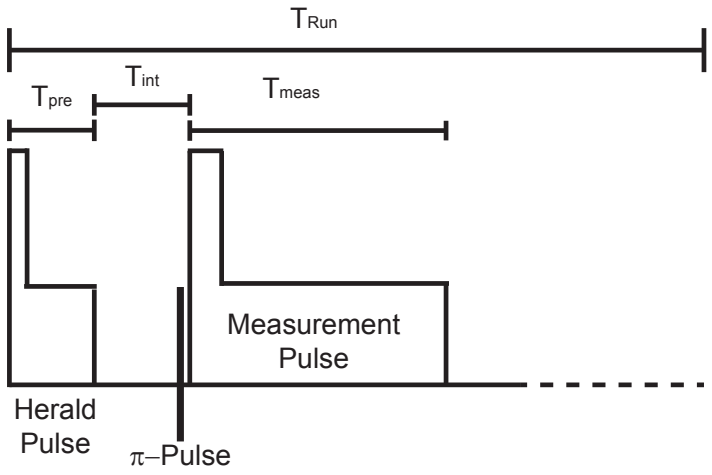

Figure 24: On-Hold Pulse Scheme. The new pulses begins with a very large amplitude for a short duration to quickly populate the resonator up to the desired photon flux. After which the amplitude is reduced to maintain the flux at the desired level. The ratio of the amplitudes is set to $t_{\text{pulse1}}\sqrt{\chi^2 + \kappa^2/4}$.

tegration time of $2.6/\kappa$. With the On-Hold

pulse, however, the integration start time was around $0.3/\kappa$ and the Pre-Decay was reduced to roughly 12%. The integration time of $2.6/\kappa$ was used again for comparison purposes. The On-Hold pulse shaping has not yet been fully optimized, but our initial results show a clear improvement in measurement speed as the integration start time was earlier.

5.3 Conclusion and Outlook

In conclusion, our experimental results are in agreement with our predictions. The outcomes of our practical optimizations steps of Section 3 as well as our more theoretical predictions presented in Section 5 all coincide with our intuitions and measurement results. We are confident that our understanding of the terms in the SNR model in Equation 13 and their limitations have allowed us to identify an optimal set of system parameters to perform rapid, high-fidelity, single-shot readout. Our results indicate that the measurement amplitude is the best parameter for improving the speed of the readout and this in turn is maximized by detuning the qubit into the straddle regime between $\Delta = 0$ and $\Delta{=}\alpha$. Then by designing the correct κ and g, the Purcell decay and state mixing can be tuned within acceptable limits for high-fidelity readout. The parameters indicate that we will be able to create measure the qubit state with greater than 95 % fidelity and only 40 ns of measurement time.

The next step of this project will be to create the chip with the optimized parameters and measure the actual fidelity and measurement time. If it is as expected, we would then like to investigate the use of a Purcell filter to further improve the fidelity [20]. If we can achieve the 100 factor increase of κT_1 as Sank et. al. predict and show, we may be able to achieve an incredible fidelity in a very rapid time and truly design a Formula One race-car qubit for single-shot readout.

Appendix: Other detuning regimes and F-level Mapping

If one is limited to negative detunings, where the qubit frequency is below the resonator frequency, one can perform f-level mapping for improved fidelity. In this regime $\chi_f \approx \chi_e$ as shown in Figure 13. This should allow an increase in fidelity due the decay properties of the Transmon and the thresholding procedure. Specifically, only decays to neighboring states are allowed [13], and since the $|e\rangle$ and $|f\rangle$ level are so similar in this regime, they will have similar means in the IQ plane. Therefore, a threshold between $|e\rangle$ and $|g\rangle$ will also work for between $|f\rangle$ and $|g\rangle$. Exploiting this, we can map $|e\rangle$ level excitations to the $|f\rangle$ level and create a "decay buffer" allowing for an effective increase in the qubit decay time. This will result in either allowing for a longer integration time to lower the Overlap term, or with a nearly identically integration time, a decrease in the number of decays during integration which will result in a reduction of the Pre-Decay term.

The results of testing this hypothesis are shown in Figure 25 using Chip 1 and identical start time and integration time. We can see that there is an improvement in fidelity. In this case 240 ns of integration was sufficient to drive the Overlap term to nearly zero without the use of the $|f\rangle$ level. Therefore, we would expect the improvement not in a longer integration time, but in a decrease of the Pre-Decay term, which is clearly the case in the measurement. The histogram has been plotted on a logarithmic scale to better see the effects of the "integration decay" which, as discussed previously, result in larger counts in the areas between the two curves. In this regime we expect nearly zero mixing between the states so the points in the middle are not related to this and the integration time is very long allowing for a greater chance of "integration decay". On can see in plot A) that the counts between the curves are larger than in plot B) on the right side (and therefore the wrong side) of the threshold and these are likely inflating the Pre-Decay term in plot A).

The optimization routine used in this thesis found that in this negative regime, it was better to have a large detuning and a large coupling, g imposed by the upper limit of g. This is why Chip 1 and 2 had such high fidelity. For detunings above the straddle regime, $\Delta > \omega_{res} + \alpha$, the results were similar in nature, however as can be seen in Figure 21, this range slightly outperforms the negative regime. This regime did not have as high values of g, stabilizing around 140 MHz.

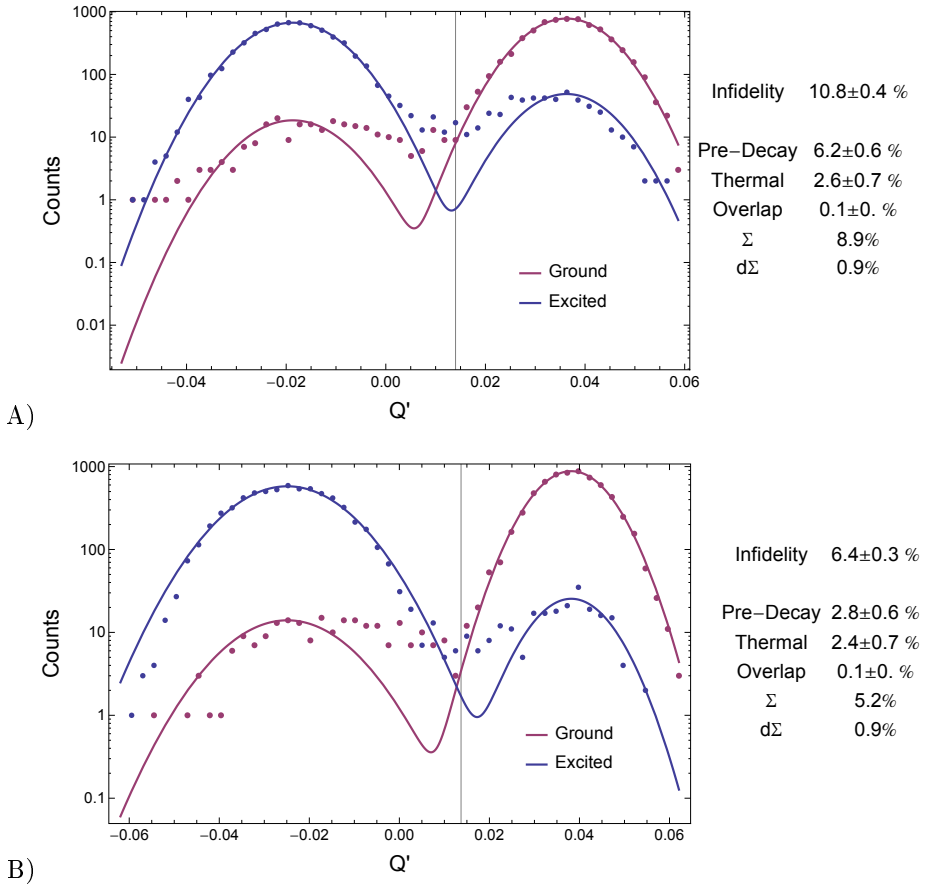

Figure 25: F-level mapping. A) Results of single-shot measurements from Chip 1. B) Results of single-shot measurements from Chip 1 using F-level mapping.

References

[1] Markus Ansmann, H. Wang, Radoslaw C. Bialczak, Max Hofheinz, Erik Lucero, M. Neeley, A. D. O'Connell, D. Sank, M. Weides, J. Wenner, A. N. Cleland, and John M. Martinis. Violation of Bell's inequality in Josephson phase qubits. *Nature*, 461(7263):504–506, September 2009.

[2] Matthias Baur. Quantum electrodynamics with superconducting circuits: Measurement of the cavity photon number using ramsey interference. Master's thesis, ETH Zurich, 08 2007.

[3] J. S. Bell. On the Einstein Podolsky Rosen paradox. *Physics (N. Y.)*, 1:195, 1964.

[4] A. Blais, R.-S. Huang, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf. Cavity quantum electrodynamics for superconducting electrical circuits: An architecture for quantum computation. *Phys. Rev. A*, 69(6):062320–14, June 2004.

[5] A. A. Clerk, M. H. Devoret, S. M. Girvin, Florian Marquardt, and R. J. Schoelkopf. Introduction to quantum noise, measurement, and amplification. *Rev. Mod. Phys.*, 82(2):1155–1208, Apr 2010.

[6] G. de Lange, D. Riste, M. J. Tiggelman, C. Eichler, L. Tornberg, G. Johansson, A. Wallraff, R. N. Schouten, and L. DiCarlo. Reversing quantum trajectories with analog feedback. *Phys. Rev. Lett.*, 112(8):080501–, February 2014.

[7] Christopher Eichler. *Experimental Characterization of Quantum Microwave Radiation and its Entanglement with a Superconducting Qubit.* PhD thesis, ETH Zurich, 01 2013.

[8] Christopher Eichler and Andreas Wallraff. Controlling the dynamic range of a josephson parametric amplifier. *EPJ Quantum Technology*, 1(1):2, 2014.

[9] J. Gambetta, A. Blais, M. Boissonneault, A. A. Houck, D. I. Schuster, and S. M. Girvin. Quantum trajectory approach to circuit qed: Quantum jumps and the zeno effect. *Phys. Rev. A*, 77:012112, 2008.

[10] J. Gambetta, A. Blais, D. I. Schuster, A. Wallraff, L. Frunzio, J. Majer, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf. Qubit-photon interactions in a cavity: Measurement-induced dephasing and number splitting. *Phys. Rev. A*, 74(4):042318, October 2006.

[11] J. Gambetta, W. A. Braff, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf. Protocols for optimal readout of qubits using a continuous quantum nondemolition measurement. *Phys. Rev. A*, 76(1):012325–11, July 2007.

[12] Joonas Govenius. Single-shot qubit readout in circuit qed using parametric amplification. Master's thesis, ETH Zurich, 02 2012.

[13] 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(4):042319, 2007.

[14] Christian Lang. Read-out strategies for multi-qubit states in circuit quantum electrodynamics. Diploma thesis, ETH Zurich and LMU Munich, 06 2009.

[15] S. Ng and M. Tsang. Optimal signal processing for continuous quantum-nondemolition qubit readout. *arXiv:1405.7262*, 2014.

[16] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, 2000.

[17] M. Pechal, C. Eichler, S. Zeytinoglu, S. Berger, A. Wallraff, and S. Filipp. Microwave-controlled generation of shaped single photons in circuit quantum electrodynamics. *arXiv:1308.4094*, 2013.

[18] David M. Pozar. *Microwave engineering.* John Wiley & Sons, Inc., 4th ed. edition, 2011.

[19] Rowe, Kielpinski, Meyer, Sackett, Itano, Monroe, and Wineland. Experimental violation of a bell's inequality with efficient detection. *Nature*, 409, 2001.

[20] D. Sank, E. Jeffrey, J. Y. Mutus, T. C. White, J. Kelly, R. Barends, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. Megrant, P. J. J. O'Malley, C. Neill, P. Roushan, A. Vainsencher, J. Wenner, A. N. Cleland, and J. M. Martinis. Fast scalable state measurement with superconducting qubits. *arXiv:1401.0257*, 2014.

[21] D. I. Schuster, A. Wallraff, A. Blais, L. Frunzio, R.-S. Huang, J. Majer, S. M. Girvin, and R. J. Schoelkopf. AC Stark shift and dephasing of a superconducting qubit strongly coupled to a cavity field. *Phys. Rev. Lett.*, 94(12):123602, April 2005.

[22] D. H. Slichter. *Quantum Jumps and Measurement Backaction in a Superconducting Qubit*. PhD thesis, University of California, Berkeley, 2011.

[23] S. J. Srinivasan, A. J. Hoffman, J. M. Gambetta, and A. A. Houck. Tunable coupling in circuit quantum electrodynamics using a superconducting charge qubit with a V-shaped energy level diagram. *Phys. Rev. Lett.*, 106(8):083601, Feb 2011.

[24] L. Steffen, Y. Salathe, M. Oppliger, P. Kurpiers, M. Baur, C. Lang, C. Eichler, G. Puebla-Hellmann, A. Fedorov, and A. Wallraff. Deterministic quantum teleportation with feed-forward in a solid state system. *Nature*, 500:319–322, 2013.

[25] Gregor Weihs, Thomas Jennewein, Christoph Simon, Harald Weinfurter, and Anton Zeilinger. Violation of bell's inequality under strict einstein locality conditions. *Phys. Rev. Lett.*, 81(23):5039–5043, Dec 1998.