Single-Shot Qubit Readout in Circuit QED using Parametric Amplification

Joonas Govenius

Advisers:

Christopher Eichler
and
Prof. Andreas Wallraff

February 20, 2012

Contents

1 Introduction

This thesis is about measuring the state of a superconducting two-level system (qubit) using another superconducting circuit that acts as a nearly quantum limited microwave amplifier. As described in Chapter 4, this goal was achieved with 85 % fidelity, which is high for the chosen readout method. Before defining fidelity or the method precisely, however, a brief overview of what is special about the field of superconducting circuits in general and the presented experiments in particular is given in this introductory chapter. In addition this chapter provides references to other related experiments in the field.

Superconducting circuits offer an exciting opportunity to build experimental solid state realizations of a wide variety of quantum mechanical models. What is special about superconducting circuits is that they are macroscopic and still accurately modeled by a few effective particles. This is possible because the excitation spectrum in a superconductor is gapped in the sense that the continuum of quasi-particle excitations starts from a non-zero energy, unlike in a normal metal. What remains below the gap energy are the collective excitations which form a discrete spectrum even for macroscopic samples and can be manipulated by the choice of sample geometry. At sufficiently low excitation energies the continuum of quasi-particle excitations can be ignored altogether because of the large energy detuning to the levels of interest. This is different from atomic as well as quantum dot systems where the quasi-particle excitations themselves are used. In that case the system size must be small because the spacing of the quasi-particle energy levels vanishes as the system becomes larger.

Circuit quantum electrodynamics (cQED) is a particularly interesting application of superconducting circuits where qubits are coupled to microwave resonators, much like atoms are coupled to optical cavities in traditional cavity quantum electrodynamics experiments [Blais04, Wallraff04, Girvin09]. The qubits in cQED consist of small superconducting islands coupled by capacitances and Josephson junctions in such a way that their excitation spectrum is anharmonic, often so much so that only the ground and first excited state are relevant, making them effective two-level systems. The microwave resonators on

the other hand are co-planar waveguides terminated by impedance mismatches that act as mirrors for the microwave photons inside the quasi one-dimensional resonators. In cQED the coupling between the qubit and the resonator can easily reach the strong coupling regime, i.e. a regime where the coupling rate at the single excitation level exceeds all decoherence rates. This is interesting because it allows the systems to dramatically affect each other even in the low energy regime where quantum mechanics is essential for explaining the observed behavior. Experimentally, this regime is easy to reach in cQED because of the macroscopic size of the qubits and the one dimensional confinement of the microwave photons that lead to large dipole moments and large electric fields per photon.

One of the problems currently limiting cQED experiments is that usually high fidelity non-destructive single-shot qubit measurement is not available. Recently progress towards such measurements was demonstrated in [Vijay11] and in [Mallet09]. Typically, however, only averaged quantities can be reliably determined after repeating the same experiment many times because of the poor signal-to-noise ratio in individual measurements, see e.g. [Wallraff05, Bianchetti09]. Besides the obvious disadvantage of making experiments slower this has the more dire consequence of making feedback based on measurement results impossible. This is because quantum mechanically a measurement result is not in general deterministic even if the input state is always the same. Hence the averaged measurement result may not contain any useful information at all, unlike in a classical deterministic process with additional unbiased noise. In particular, many quantum error correction schemes require feedback based on high fidelity single-shot measurement [Nielsen00, Fowler09].

Dispersive read-out is a common tool used for measuring the qubit in cQED. This continuous measurement technique entangles the qubit state with the coherent amplitude of a measurement signal reflected from or transmitted through the resonator. The measurement is in principle quantum non-demolition (QND) because it is performed in the dispersive limit where the frequency detuning of the qubit from the resonator is large and the measurement power is low [Blais04]. This means that it projects the qubit state once but does not cause further transitions between the measurement basis states. In principle the signal to noise ratio is only limited by the unavoidable vacuum noise in the signal quadratures and by the measurement time limited by qubit relaxation [Gambetta07]. However, staying in the dispersive regime limits the signal power to a few photons per second per hertz, which is an order of magnitude less than the additional noise added by state of the art commercial microwave amplifiers. Therefore, there is room for greatly increasing the rate of information gain from this type of measurement without causing additional back-action on the qubit or resonator by simply using a better amplifier.

In the experiments presented here a superconducting Josephson parametric amplifier (paramp) was used as a pre-amplifier before a commercial high electron mobility transistor (HEMT) amplifier. The noise added by the paramp is nearly quantum limited and, since it also has a high gain, it makes the noise added by the later amplification stages insignificant. With this setup the qubit state

can in fact be determined in individual dispersive measurements with a time resolution of tens of nanoseconds as shown recently in [Vijay11] and also in this thesis. Therefore the setup allows high fidelity single-shot qubit readout even with a fairly short qubit relaxation time as discussed in Chapter 4. With the time resolution of the measurement being small compared to the qubit relaxation rate, it is also possible to directly investigate the QND nature of the measurement as described in Section 4.6.

It should also be mentioned that other single-shot measurement schemes with similar fidelities have been proposed and implemented. The most relevant ones for cQED are the non-QND high power readout discussed in [Boissonneault10] and the nearly QND Josephson bifurcation amplifier (JBA) based readout schemes [Siddiqi04, Mallet09]. The two are similar in that they both entangle the initial qubit state with a metastable state of a non-linear resonator but in the JBA the non-linearity is much stronger and hence the power necessary for bifurcation much smaller. Finally, for so called flux and phase qubits, which are not commonly used in current cQED experiments, non-QND single-shot measurement with high fidelity has been possible essentially from the start [Chiorescu03, Martinis02]. However, all of the non-QND schemes leave the measured system in a classical state with large entropy, which means that restoring the system back to a known state is slow. Hence these other schemes—with the exception of that demonstrated in [Mallet09]—are not ideal for feedback experiments, which are arguably the most important use case for single-shot measurement.

2

Circuit Quantum Electrodynamics

Circuit quantum electrodynamics is a solid state alternative to the more traditional field of cavity quantum electrodynamics. In cavity quantum electrodynamics real atoms are coupled to the electromagnetic field modes of optical cavities via the electric dipole moments of the atoms. The basic idea is that, although the interaction between an atom and an individual photon is extremely weak due to the small dipole moment of the atom and the small field produced by a single photon, it is possible to observe the distinctly quantum mechanical effects of this coupling by confining the photon in a cavity. Roughly speaking, the cavity gives the photon many chances to interact with the atom before leaking out of the system. For a review of cavity quantum electrodynamics experiments see [Walther06].

In circuit quantum electrodynamics (cQED) the cavities are replaced by superconducting resonators and the atoms by superconducting qubits as shown in Figure 2.1 [Blais04, Wallraff04, Girvin09]. Perhaps the most important advantage of this system is the fact that most parameters can be chosen either during circuit design and fabrication or during the experiment by applying external magnetic or electric fields. This allows for example easy access to the strong coupling regime where the resonator-qubit coupling exceeds all dissipation rates. A solid state system also offers certain practical advantages such as having the qubit at a fixed position in the resonator and the use of microwave signals, which are in general easier to produce, easier to accurately guide, and safer than optical signals. On the other hand, the fundamental disadvantage of solid state systems is that the qubits and resonators are subject to significant uncontrolled electric and magnetic field fluctuations because they are surrounded by imperfect solid dielectrics rather than vacuum. On the practical side the smaller frequencies also mean that, in order to resolve effects at the level of a single energy quantum $h\nu$, one must keep the temperature T well below $h\nu/k_B$. For typical resonance frequencies ν of the order of 10 GHz this implies that T must be much smaller than 0.5 K. In order to meet this condition, all experiments for this thesis were

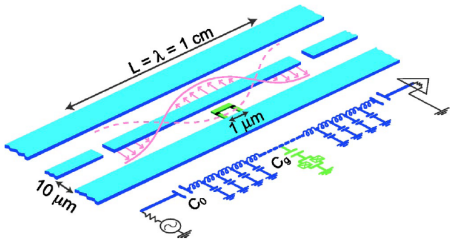

Figure 2.1: The basic building blocks of cQED experiments: a superconducting co-planar waveguide resonator (light blue) and a charge qubit (green). The superconductors are deposited on a dielectric such as sapphire and measured near 20 mK in vacuum. The pink curves and arrows represent the distribution of electric field produced by the 1^{st} harmonic of the resonator that couples to the charge qubit in the center. As explained in [Blais04], the system can be well modeled as the lumped element circuit shown in dark blue and green. Note that in the experiments here the qubit is placed near the end of the resonator so that it couples to the fundamental mode. The figure is from [Blais04].

performed in a dilution refrigerator with a base temperature of approximately 30 mK.

There are several papers and thesis from various groups in the field that discuss and derive detailed properties of cQED experiments [Blais04, Girvin09, Koch07, Schuster07]. This section therefore only presents the models and results that are most directly relevant to dispersive qubit measurements and the experimental setup shown in Figure 4.1. The interested reader is directed to the referenced material for more details and derivations.

2.1 Transmission Line Resonator

The transmission line resonator shown in Figure 2.1 consists of a coplanar waveguide with constant 50 Ω impedance everywhere except at the ends where the center pin has gaps. These gaps, referred to as input and output capacitors, act as mirrors due to the impedance mismatch and define the physical length and therefore—together with the phase velocity—also the electrical length of the resonator for any given wavelength. At wavelengths where the electrical length is an integer or half integer, the reflections from the capacitors form standing waves inside the resonator. This is very similar to the standing waves that form in coaxial cables between (un)intentional impedance mismatches.

Each such mode of electromagnetic radiation in the resonator can be thought of as a harmonic oscillator with the field quadratures of the radiation corresponding to the position and momentum of a mechanical oscillator. As shown in standard quantum optics text books, it is also possible to use this analogue to treat each mode quantum mechanically and call the quantized excitations

photons [Scully97]. In the remainder of this thesis only the fundamental mode at $\omega_r/2\pi \approx 7.17\,\mathrm{GHz}$ will be considered since interactions with the qubit at $\omega_{01}/2\pi \approx 6.13\,\mathrm{GHz}$ become weaker with increasing frequency detuning.

Mathematically, the evolution of an isolated resonator mode is described by a Hamiltonian

$$H_r = \hbar\omega_r a^\dagger a \tag{2.1}$$

where a and $a^\dagger$ are the lowering and raising operators for the mode and obey $\left[a, a^\dagger\right] = 1$. The basis that diagonalizes this Hamiltonian is called the Fock basis $\{|n\rangle \,|n \geq 0\}$, where the energy quanta n are called photons. The lowering operator acts on the Fock basis states in the usual way:

$$a\,|n\rangle = \begin{cases} n\,|n-1\rangle & \text{if } n > 0 \\ 0 & \text{if } n = 0. \end{cases}$$

One can also define the quadrature operators

$$x = \frac{a^\dagger + a}{2},\; y = i\frac{a^\dagger - a}{2}$$

which obey $[x, y] = i/2$ and are proportional to the quadratures of the electric (or magnetic) field produced by the mode. These quadratures can also be defined with an arbitrary overall phase ϕ by replacing a by $e^{i\phi}a$ in the definitions above.

The lowering operator is often called *the* mode operator because coherent states

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \alpha^n\,|n\rangle\,/\sqrt{n!}$$

are eigenstates of a and coherent states are the closest analogue to classical states in the sense that both x and y have equal and minimal uncertainty around their expectation values. Furthermore, these expectation values are the real and imaginary parts of $\langle a\rangle = \langle x\rangle + i\,\langle y\rangle = \alpha$.

Finally, the coupling of the resonator to the outside world is ideally only due to the capacitive coupling to the transmission lines connected to the ends of the resonator. By making the capacitors highly asymmetric—as is the case for the experiments in this thesis—one of the capacitors can be considered an input and the other an output capacitor. The much larger output capacitor causes much stronger coupling to the transmission line modes which ensures that the photons leaking out of the cavity at rate κ mostly go to the paramp rather than the attenuators connected to the input. On the input side the low coupling is not a problem because the input signals are known coherent signals that can be made essentially arbitrarily strong. Details of the modes in the semi-infinite transmission lines and the so called input-output theory that relates them to the resonator mode are discussed in Chapter 3 on parametric amplification. For a more complete review of transmission lines, resonators, and input-output theory see the online appendices of [Clerk10].

2.2 Transmon

The qubit used in the experiments is called a transmon and it consist of two superconductors connected by a Josephson junction with Josephson energy E_J and a large parallel capacitance that makes the charging energy E_C small compared to E_J [Koch07]. More precisely, $E_J(\phi_{ext})$ is the effective Josephson energy of a small SQUID loop formed by a pair of junctions and it can be tuned between the sum and difference of the two individual Josephson energies by adjusting the magnetic flux ϕ_{ext} threading the SQUID loop. This allows tuning of the qubit transition frequency from close to zero to a maximum design value set by the maximum $E_J(\phi_{ext}=0)$ as seen below. More detailed explanation of all the features described below can be found in [Koch07], which originally introduced the transmon.

Ignoring coupling to other systems, this system of two superconductors can be described by a Hamiltonian of the form

$$H_q = -4E_C(n-n_g) - E_J(\phi_{ext})\cos(\varphi) \tag{2.2}$$

where $E_C \ll E_J$ is the charging energy set by the large capacitance, n_g is an offset charge determined by surrounding electric fields, n is the operator for the number of cooper pairs, and φ is the operator for the gauge invariant phase difference across the effective junction. The first two transition energies found by solving this Hamiltonian are

$$\begin{aligned}\omega_{01} &\approx \sqrt{8E_CE_J(\phi_{ext})} - E_C + \delta C_{01}(n_g)\\ \omega_{12} &\approx \omega_{01} + \alpha + \delta C_{12}(n_g)\end{aligned}$$

where $\alpha \approx -E_C$ within about 20 % for $E_J/E_C \gtrsim 15$ and $\delta C_{ij}(n_g)$ is referred to as charge dispersion for the $i \to j$ transition. What is special about the transmon compared to its predecessor called Cooper Pair Box (CPB) is that $|\delta C_{ij}(n_g)|$ decreases exponentially as E_J/E_C is increased and therefore the transmon can be made insensitive to low frequency stray electric fields.[1] This is important since n_g typically cannot be controlled accurately in solid state systems and fluctuations in ω_{01} set the dephasing time of the qubit. On the other hand E_C also sets the anharmonicity α—and therefore the minimum pulse length required for addressing only the $|0\rangle \to |1\rangle$ transition—so evidently there is a compromise to be made. Fortunately, α scales only linearly while $|\delta C_{ij}(n_g)|$ scales exponentially with E_C so a good compromise can usually be found.

For cQED another crucial feature of the transmon is its strong electric dipole moment at microwave frequencies that allows strong coupling to the oscillating electric field of the resonator when placed between the center pin and the ground plane, as shown in Figure 2.1. This is in spite of the fact that the charge dispersion terms—and therefore the DC dipole moment—are exponentially small. This can be understood from the fact that the eigenfunctions of the Hamiltonian

[1] In fact the CPB is described by the same Hamiltonian but it is operated in the $E_C \gg E_J$ regime where only two eigenstate of the number operator n are relevant.

in Eq. 2.2 are in fact highly sensitive to n_g but in such a way that the spacing of their corresponding eigenvalues stays nearly constant. This means that fast changes in n_g—as produced by high frequency photons in the resonator—indeed cause transitions between qubit states while adiabatically changing n_g does not lead to dephasing (or transitions).

An important practical consideration for current experiments is that due to strong interactions with the environment the qubit spontaneously relaxes from $|e\rangle$ to $|g\rangle$ on a typical time scale of $T_1 \sim 100\,\mathrm{ns} \ldots 10\,\mu\mathrm{s}$. Due to a combination of the charge insensitivity of the transmon and the short T_1 times, the dephasing time $T_2^\star$ that describes how long superpositions of $|e\rangle$ and $|g\rangle$ stay coherent is typically limited by the theoretical limit $2T_1$ [Houck09]. The cause of the short relaxation time is not well understood but a recent discovery of reproducible relaxation times of tens of microseconds in an equivalent but three dimensional setup have shown that the current limitations are at least not intrinsic to the Josephson junctions themselves [Paik11].

2.3 Jaynes-Cummings Model

The traditional Jaynes-Cummings model describes the coupling of a two-level system to a single resonator mode. However, due to the small anharmonicity α of the transmon it is essential to use at least a three level system to quantitatively predict the effects of the coupling. Hence a generalized Jaynes-Cummings Hamiltonian will be used here [Koch07]:

$$H = \hbar \sum_j \omega_j |j\rangle \langle j| + \hbar\omega_r a^\dagger a + \hbar \sum_{i,j} g_{i,j} |i\rangle \langle j| \left(a + a^\dagger\right)$$
$$\text{where } g_{i,j} = 2\beta e V_{rms}^0 \langle i|n|j\rangle ,$$

β is a ratio of capacitances set by geometry, and V_{rms}^0 is the root-mean-square of the voltage produced by a single photon. This Hamiltonian is obtained by adding H_r and H_q from Eqs. 2.1 and 2.2 together with a dipole interaction term $2\beta V_{rms}^0 n \left(a + a^\dagger\right)$ that couples the voltage operator of the resonator mode and the charge operator of the qubit.

For this thesis the relevant parameter regime is the dispersive one where hybridization of the qubit and resonator states is small. More precisely, in the dispersive regime the detuning $\Delta = \omega_{ij} - \omega_r$ of the qubit transition frequency $\omega_{ij} = \omega_j - \omega_i$ from the resonator frequency ω_r must be large compared to the magnitude of the interaction term for the transition $0 \to 1$ and $1 \to 2$. For a coherent or Fock state the magnitude of the interaction term is set roughly by $g_{ij}\sqrt{\langle n \rangle}$. Furthermore, since $g_{12} \approx \sqrt{2} g_{01}$, the requirement for the validity of the dispersive approximation is often expressed as $\langle n \rangle \ll n_{crit}$, where $n_{crit} \equiv \Delta^2 / 4 g_{01}^2$ [Blais04]. If this is the case, the system is well described by the

dispersive Hamiltonian

$$\begin{aligned} H &= \frac{\hbar\omega'_{01}}{2}\sigma_z + \hbar\left(\omega'_r + \chi\sigma_z\right)a^\dagger a \\ \chi &\equiv \frac{g_{01}^2}{\Delta} - \frac{1}{2}\frac{g_{12}^2}{\Delta+\alpha} \end{aligned} \tag{2.3}$$

where $\sigma_z \equiv \left(|1\rangle\langle 1| - |0\rangle\langle 0|\right)/2$ and the interaction has led to Lamb shift type renormalization of the qubit frequency $\omega'_{01} = \omega_{01} + g_{01}^2/\Delta$ and the resonator frequency $\omega'_r = \omega_r - g_{12}^2/2\left(\Delta+\alpha\right)$, as well as an interaction term $\hbar\chi\sigma_z a^\dagger a$. What is special about this regime is that the interaction term commutes with both the qubit and resonator parts of the Hamiltonian and therefore does not mix the qubit and resonator states.[2] As described later, this has an important effect on the nature of the qubit measurement that can be performed using this interaction.

[2]More precisely the interaction term does not mix the already slightly hybridized qubit and resonator states with the renormalized frequencies. These states are known as dressed states.

3

Parametric amplification

A parametric oscillator has an externally modulated system parameter, e.g. a string with periodically varying tension or an LC oscillator with a periodically varying $L(t)$. Such external parametric pumping can transfer energy into or out of the system like external forcing of a regular damped harmonic oscillator. However, while the response of a regular forced oscillator is linear in different forcing terms, a parametric oscillator responds non-linearly to the combination of parametric pumping, regular external forcing, and initial oscillations. In particular, the parametric pumping transfers energy into existing oscillations in one quadrature of oscillations at half of the pump frequency and can therefore be used for amplification of this signal quadrature.

When a parametric oscillator is used as an amplifier, the input signal appears as an additional external forcing term that the oscillator should be sensitive to while still absorbing energy from the parametric drive. This is indeed the case in a certain regime of parametric drive power as will be sketched mathematically in this section. However, it is also conceptually useful to think of the parametric drive as acting like an effective negative damping force that competes with the real damping. Once the negative damping becomes stronger than the real damping, the zero amplitude state of the resonator becomes unstable and the system bifurcates. In this high parametric drive regime the initial conditions determine which of the two bistable states the oscillator will reach.[1] In this picture, one can define the time averaged amplitude as an order parameter and think of the parametric oscillator as undergoing a second order phase transition as the strength of the parametric drive relative to the damping is increased. It is then clear from general properties of second order phase transitions that the system should act like an amplifier near the bifurcation point since the susceptibility to external perturbations near a critical point always diverges.

[1] This is called "spontaneous symmetry breaking" because the system steady state violates the symmetry of the Hamiltonian. Only two stable steady state solutions emerge because the parametric drive acts as negative damping only for oscillations in one quadrature, for the other quadrature it in fact acts like additional positive damping.

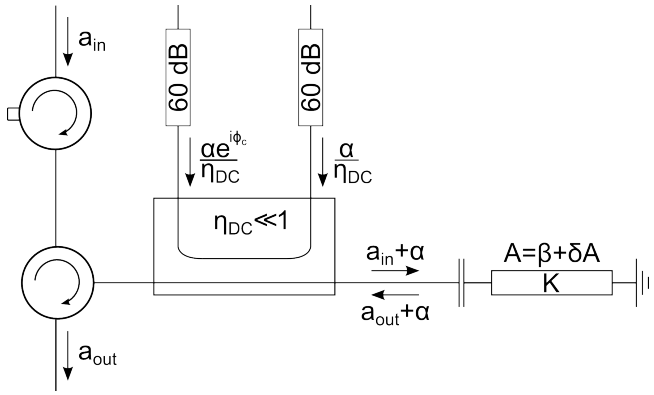

Figure 3.1: Paramp based on a non-linear cavity with Kerr constant K. The lower left component is a circulator that separates the paramp input and output with roughly -20 dBm leakage in the undesired direction. The component in the middle is a 20 dBm ($\rightarrow \eta_{DC} = 0.01$) directional coupler that adds the pump tone α to the paramp input and subtracts it from the output for an appropriately chosen ϕ_c. This is helpful for minimizing leakage of the strong pump tone into the input port as well as for preventing saturation of other amplifiers in the detection chain. The upper left component is an isolator that provides an additional 20 dB of isolation between the input and the strong pump tones. The (distributed) large attenuators provide thermal isolation and dissipate the pump power leaving the directional coupler.

Furthermore, the correlation time of the fluctuations also diverges which implies that the bandwidth in which the oscillator responds to external perturbations also vanishes near the bifurcation point.

3.1 From Kerr Nonlinearity to Squeezing

The quantum mechanical model of a paramp based on a resonator with a Kerr non-linearity is shown in Figure 3.1. The parametric oscillator in this model is the fundamental resonator mode A, which has a resonance frequency ω_0 that can be modulated using the Kerr effect and a large amplitude pump tone.[2] The additional components shown in Figure 3.1 are there for technical reasons as discussed in Appendix A.

Quantum mechanically the Kerr effect is described by a four-photon term in the Hamiltonian H_r for the fundamental resonator mode:

$$H_r = \hbar\omega_0 A^\dagger A + \frac{\hbar K}{2} A^\dagger A^\dagger A A \tag{3.1}$$

[2]The dependence of the refractive index $n = n_0 + n_2 E^2$ on electric field is called the Kerr effect [Scully97]. Note that the E^2 dependence implies that the parametric drive indeed occurs at twice the resonator frequency even though the pump tone is at the resonator frequency.

where A is the lowering operator of the resonator mode and where the Kerr constant $K < 0$ arises naturally in superconducting transmission line resonators from the non-linear kinetic inductance of the Cooper pairs [Yurke06]. In addition, the magnitude of the non-linearity can be greatly increased and the resonance frequency ω_0 made tunable by incorporating pairs of parallel Josephson junctions in the center conductor of the resonator and controlling the magnetic flux through these so called SQUID loops. This and very similar type of parametric amplifiers were recently used in [Eichler11] and [Vijay11]. Similar Josephson junction based amplifiers were also studied already much earlier by Yurke [Yurke89, Yurke96].

How the Hamiltonian in Eq. 3.1 leads to parametric amplification as well as the effect of internal cavity losses is described in detail in [Yurke06]. Here an alternative simpler inspection of the steady state behavior of a lossless resonator is presented to convince the reader of the feasibility of using the system as a paramp. Furthermore, only the so called degenerate mode of operation is examined where the signal is exactly at the pump frequency and only one signal quadrature is amplified. This is not only theoretically simpler than the non-degenerate mode but also practically the more relevant case for single qubit measurements.

Consider the case of the resonator mode A being capacitively coupled to a continuum of transmission line modes. Also assume that the transmission line modes in steady state are described by

$$|\psi_{tl}(\omega)\rangle = \begin{cases} N\left[\mathcal{D}\left(\alpha e^{-i\omega t}\right)|0\rangle_\omega + |\delta(\omega)\rangle\right] & \text{if } \omega = \tilde{\omega}_0 \\ N\left[|0\rangle_\omega + |\delta(\omega)\rangle\right] & \text{otherwise} \end{cases}$$

where $|0\rangle_\omega$ is the ground state of mode ω, $\mathcal{D}(\alpha') = \exp\left(\alpha' a^\dagger - (\alpha')^\star a\right)$ displaces the ground state to a coherent state α', $\int d\omega \langle \delta(\omega) | \delta(\omega) \rangle \ll 1$, and $N \approx 1$ is a normalization coefficient. In other words, a sharp mode at the pump frequency $\tilde{\omega}_0$ is mostly in a large amplitude coherent state $\alpha e^{-i\tilde{\omega}_0 t}$ while other modes are mostly in their ground states, except for small fluctuations $|\delta(\omega)\rangle$.

Let $a^\dagger(\omega)$ and $a(\omega)$ be the raising and lowering operators for the continuum of transmission line modes. As bosonic operators they obey $\left[a(\omega), a^\dagger(\omega')\right] = \delta(\omega - \omega')$. Later it will also be convenient to consider the set of displaced mode operators

$$\delta a(\omega) = a(\omega) - \delta(\omega - \tilde{\omega}_0) N \alpha e^{-i\omega t} \mathcal{I}(\omega)$$

where $\mathcal{I}(\omega)$ is the identity operator on mode ω and will be absorbed into α from now on. These displaced operators also satisfy $\left[\delta a(\omega), \delta a^\dagger(\omega')\right] = \delta(\omega - \omega')$ and are independent of the pump in the sense that they act only on the fluctuations $|\delta(\omega)\rangle$. More precisely,

$$\int_\omega^{\omega+\epsilon} d\omega' \langle \psi_{tl}(\omega') | \delta a(\omega') | \psi_{tl}(\omega') \rangle = \int_\omega^{\omega+\epsilon} d\omega' N^2 \langle \delta(\omega') | a(\omega') | \delta(\omega') \rangle \quad \forall \omega, \epsilon$$

and similarly for other normally ordered moments involving $\delta a(\omega)$ and $\delta a^\dagger(\omega)$.

The Hamiltonian for the combined system including the interaction between the resonator mode and the transmission line is

$$\begin{aligned} H =& H_r + H_a + H_{int} \\ \text{where } H_a =& \int d\omega \hbar\omega a^\dagger(\omega)\, a(\omega) \\ \text{and } H_{int} =& \int d\omega \hbar\sqrt{2\pi\kappa}\left[A^\dagger a(\omega) + \text{H.c.}\right]. \end{aligned} \tag{3.2}$$

Note that A is dimensionless but $a(\omega)$ has units of $\sqrt{\text{Hz}}^{-1}$ and therefore κ has units of Hz. The factor of 2π is added so that κ correspond to that used in most cQED papers, e.g. [Blais04].

The Hamiltonian in Eq. 3.2 together with the specified state of the transmission line modes leads to driving as well as damping of the resonator mode. Hence it is reasonable to assume—and is verified later—that the steady state of the resonator mode is a large amplitude coherent state $\beta e^{-i\tilde{\omega}_0 t}$ with additional small fluctuations around this value described by the operator $\delta A \equiv A - \beta e^{-i\tilde{\omega}_0 t}\mathcal{I}$, where $\beta \in \mathbb{C}$ and $\mathcal{I}$ is identity and will be absorbed into β. In terms of δA the resonator Hamiltonian is

$$\begin{aligned} H_r =& \hbar\tilde{\omega}_0 \delta A^\dagger \delta A + \frac{\hbar K}{2}\left(\beta^2 \delta A^\dagger \delta A^\dagger e^{-2i\tilde{\omega}_0 t} + \text{H.c.}\right) \\ &+ \hbar\left(\omega_0 + \frac{K}{2}|\beta|^2\right)\left(\beta \delta A^\dagger e^{-i\tilde{\omega}_0 t} + \text{H.c.}\right) \\ &+ C_0 + O(|\beta|) \end{aligned}$$

if we set the pump frequency to $\tilde{\omega}_0 = \omega_0 + K|\beta|^2$ and lump all constant terms into $C_0 \in \mathbb{R}\mathcal{I}$. The terms on the last line can be ignored as constant or small and the term on the second line can be thought of as an additional β dependent classical drive $H'_{int}(\beta) = \left(\hbar\Omega'(\beta)\,\delta A^\dagger e^{-i\tilde{\omega}_0 t} + \text{H.c.}\right)$, where $\Omega'(\beta) \equiv \left(\omega_0 + \frac{K}{2}|\beta|^2\right)\beta$. Hence the four photon Kerr term of Eq. 3.1 combined with a strong coherent drive is seen to finally result in the standard squeezing Hamiltonian for the fluctuations:

$$H_r - H'_{int}(\beta) = \hbar\frac{\hbar\omega'_{01}}{2}\sigma_z + \hbar(\omega'_r + \chi\sigma_z)\, a^\dagger a \tilde{\omega}_0 \delta A^\dagger \delta A + \hbar\frac{\lambda}{2}\left(e^{2i\phi}\delta\tilde{A}^\dagger \delta\tilde{A}^\dagger e^{-2i\tilde{\omega}_0 t} + \text{H.c.}\right)$$

where $\lambda \equiv K|\beta|^2$ and $\phi = \arg\beta$.

3.2 Input-Output Relations and Gain

The Heisenberg equations of motion for the Hamiltonian in Eq. 3.2 are

$$\begin{aligned}\frac{dA}{dt} &= \frac{1}{i\hbar}[A, H] \\ &= \frac{1}{i\hbar}[\delta A, H_r] + i\sqrt{2\pi\kappa}\int d\omega a(\omega) \\ \frac{da(\omega)}{dt} &= \frac{1}{i\hbar}[a(\omega), H] + \frac{\partial}{\partial t}a(\omega) \\ &= -i(\omega - \omega_0)a(\omega) - i\sqrt{2\pi\kappa}A.\end{aligned}$$

These equations are solved in the so called input-output theory of Gardiner and Collett[Gardiner85].[3] The idea is to write the above differential equations in integral form and solve for $A(t+\Delta t) - A(t)$ for an interval Δt, which is small enough to assume the change in A to be small but also large enough to make the Markov approximation that $\kappa(\omega)$ is constant in the relevant frequency band. A particularly thorough discussion of the details of the calculation and the associated approximations is found in [Cohen-Tannoudji98] in the similar derivation of the master equation for the reduced density matrix of a resonator mode. Here the result from [Gardiner85] is used without proof to write down the equation that governs the damped, driven, and squeezed evolution of the intra-cavity field:

$$\frac{dA}{dt} = \frac{1}{i\hbar}[A, H_r] - \frac{\kappa}{2}A - \sqrt{\kappa}a_{in}(t) \tag{3.3}$$

Here κ describes the rate of photon decay from the resonator and $a_{in}(t)$ is the field incident at the resonator input at time t due to the superposition of all the bath modes as defined by

$$a_{in}(t) = \frac{1}{\sqrt{2\pi}}\int_0^\infty d\omega e^{-i\omega(t-t_0)}a(\omega, t_o). \tag{3.4}$$

Note that $a_{in}(t)$ is fully determined for all time t by initial conditions of the bath modes at some early time t_0. Similarly, one can define an output field

$$a_{out}(t) = \frac{1}{\sqrt{2\pi}}\int_0^\infty d\omega e^{-i\omega(t-t_1)}a(\omega, t_1) \tag{3.5}$$

which is fully determined by final conditions of the bath modes at some late time t_1. Taking advantage of the time reversal symmetry of the system the two can also be related to each other by

$$a_{out}(t) = a_{in}(t) + \sqrt{\kappa}\delta A(t). \tag{3.6}$$

Only the deviations of the input and output fields from the pump fields are relevant for amplification. These deviations are described by $\delta a_{in/out}(t)$ and are defined exactly as $a_{in/out}(t)$, except that $a(\omega)$ is replaced by $\delta a(\omega)$ everywhere

[3] See also [Walls94] or the online appendices of [Clerk10] for a helpful review of input-output theory.

in Eqs. 3.4, 3.5, and 3.6. Rewriting Eq. 3.3 in terms of δA and $\delta a(\omega)$ then finally gives the desired equation of motion for the intra-cavity fluctuations:

$$\begin{aligned}\frac{d\delta A}{dt} &= \frac{1}{i\hbar}\left[\delta A, H_r - H_d'\right] - \frac{\kappa}{2}\delta A - \sqrt{\kappa}\delta a_{in}(t) \\ &\quad + \left[-i\Omega'(\beta) + i\tilde{\omega}_0\beta - \frac{\kappa}{2}\beta - \sqrt{\frac{\kappa}{2\pi}}N\alpha\right]e^{-i\tilde{\omega}_0 t}\end{aligned}$$

where the term on the second line corresponding to driving must vanish for the assumed steady state. This is indeed the case if

$$\left(i\frac{K}{2}|\beta|^2 + \frac{\kappa}{2}\right)\beta = -\sqrt{\frac{\kappa}{2\pi}}N\alpha.$$

The remaining terms in the Hamiltonian then describe the squeezed, damped, and driven dynamics of the fluctuations:

$$\frac{d\delta A}{dt} = -i\tilde{\omega}_0\delta A + \lambda e^{i(2\phi-\pi/2)}\delta\tilde{A}^\dagger e^{-2i\tilde{\omega}_0 t} - \frac{\kappa}{2}\delta A - \sqrt{\kappa}\delta a_{in}(t) \tag{3.7}$$

Solving Eq. 3.7 and the equivalent equation for $\delta A^\dagger$ in steady state gives

$$\begin{aligned}\delta X(\phi) &= \frac{2\sqrt{\kappa}}{2\lambda-\kappa}\delta x_{in}(\phi) \\ \delta Y(\phi) &= -\frac{2\sqrt{\kappa}}{2\lambda+\kappa}\delta y_{in}(\phi) \\ \text{where } \delta X(\phi) &\equiv \frac{e^{i(\phi-\pi/4)}\delta A^\dagger + e^{-i(\phi-\pi/4)}\delta A}{\sqrt{2}} \\ \text{and } \delta Y(\phi) &\equiv i\frac{e^{i(\phi-\pi/4)}\delta A^\dagger - e^{-i(\phi-\pi/4)}\delta A}{\sqrt{2}}\end{aligned}$$

and $\delta x_{in/out}(\phi)$ and $\delta y_{in/out}(\phi)$ are the analogously defined quadrature operators, referenced to the pump phase ϕ. The solutions show that the resonator responds more and more strongly to $\delta x_{in}(\phi)$ and less and less strongly to $\delta y_{in}(\phi)$ as the pump strength λ approaches $\kappa/2$. At $2\lambda = \kappa$ the response to $\delta x_{in}(t)$ diverges, which corresponds physically to the bifurcation point where two steady state solutions begin to emerge.

For amplification the quantity of most direct interest is the output field, rather than the intra-cavity field found above. Fortunately, Eq. 3.6 allows the output field to be determined from the intra-cavity field δA and the input field δa_{in}. The result is:

$$\begin{aligned}\delta x_{out}(\phi) &= \sqrt{G}\delta x_{in}(\phi) \\ \delta y_{out}(\phi) &= \frac{1}{\sqrt{G}}\delta y_{in}(\phi)\end{aligned}$$

where $G = \left[(2\lambda+\kappa)/(2\lambda-\kappa)\right]^2$ is the gain in power for the amplified quadrature [Clerk10].

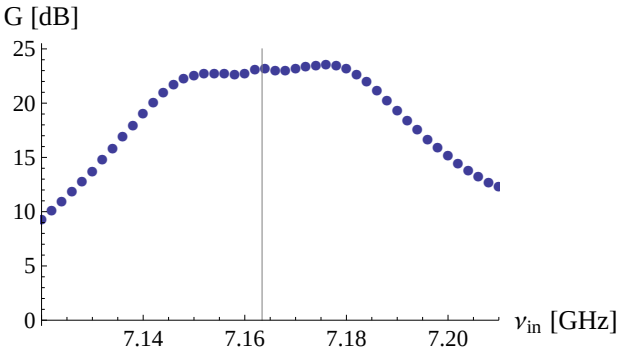

Figure 3.2: Experimentally measured power gain G versus signal frequency curve for a pump frequency of 7.1634 GHz (gray vertical line). Note that the gain bandwidth product is roughly $\sqrt{G}D = 200 \times 50\,\text{MHz}$, which exceeds the theoretical prediction $\kappa_p/2 \approx 600\,\text{MHz}$ by more than an order of magnitude. The dependence on input power is shown in Figure 3.3.

This mode of operation where one signal quadrature is amplified while the other is squeezed is called degenerate parametric amplification.[4] Note that without the addition of additional noise—as is the case in non-degenerate parametric amplification—the squeezing of the other quadrature is required by the Heisenberg uncertainty relation. Mathematically the uncertainty relation is enforced by the requirement that both the input and output modes obey the bosonic commutation relations $[x_{out}(\phi), y_{out}(\phi)] = [x_{in}(\phi), y_{in}(\phi)] = 1$. See [Clerk10] for a detailed discussion of this and other topics on limits of quantum measurement.

3.3 Gain-Bandwidth Product

So far only the steady state behavior of the paramp has been considered. This on its own does not say much about the expected performance in qubit measurements where the microwave signal is correlated with the initial qubit state only for a short time due to qubit relaxation. In order to quantify how quickly the paramp can respond to changes in the input signal one needs to know the bandwidth over which signals are amplified. This can be done by Fourier transforming Eq. 3.3, which results in a frequency dependent gain for the different frequency components. The bandwidth over which the gain is appreciable in this simple model is $D = \kappa/2\sqrt{G}$ [Clerk10]. The fact that the product of the bandwidth D and the amplitude gain $\sqrt{G}$ is fixed by the decay rate $\kappa_p/2\pi \approx 200\,\text{MHz}$ explains why a low quality factor was chosen for the paramp resonator.

[4] Other common terms for this mode of operation are phase sensitive and phase non-preserving amplification. The corresponding terms for the opposite case where both quadratures are amplified equally are non-degenerate, phase insensitive, and phase preserving amplification.

In practice, however, the measured gain versus signal frequency curve substantially deviates from this prediction at many bias points as shown in Figure 3.2. Neither the observed shape, the large bandwidth, nor the fact that the added noise remains low can be explained by the simple model presented here, in [Clerk10], or in [Yurke06]. In fact, Yurke already observed a similar gain curve in 1996 in a Josephson junction based microwave amplifier and made similar remarks about the unexpectedly large gain-bandwidth product [Yurke96]. Finally note that at other bias points—namely when little external magnetic field is applied through the SQUID loops to tune the bare resonance frequency ω_0—the gain curve does have the expected shape with roughly the expected gain-bandwidth product.

3.4 Dynamic Range

The limited dynamic range of input signal powers that the paramp can amplify was overlooked so far. It is clear that the range is limited since the system is already near the bifurcation point and a signal that is too strong will act as an additional pump tone that pushes the system into the bistable regime. An example of the saturation of the output power versus input signal power is shown in Figure 3.3, which shows a reduction of 1 dB in gain at $\langle n \rangle \approx 50$ and saturation of the output power at $\langle n \rangle \approx 200$.[5] Here $\langle n \rangle$ translates to a power in absolute units according to $\langle n \rangle \times \kappa_{source} \times \hbar\omega_{signal}$, which is the power radiated by a resonator in a coherent state with a decay rate $\kappa_{source} \approx 2\pi \times 4.5\,\mathrm{MHz}$ and mean photon number $\langle n \rangle$ at frequency $\omega_{signal}/2\pi = 7.166\,\mathrm{GHz}$. $\langle n \rangle$ is the relevant scale here since these are the parameters of the resonator that the measured qubit was coupled to.

[5] The so called 1 dB compression point is a standard measure for linearity.

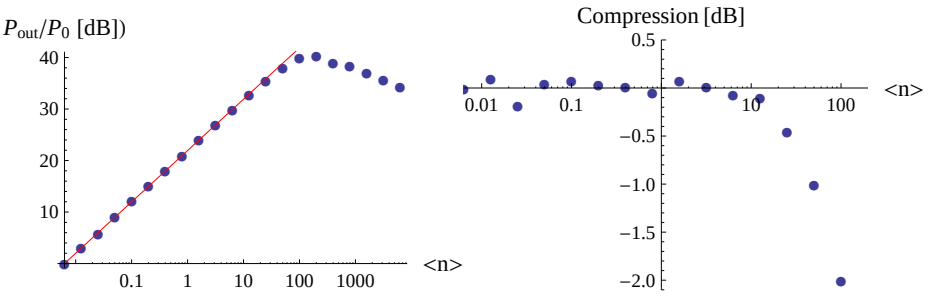

Figure 3.3: Dynamic range of the paramp at $\omega_{signal}/2\pi = 7.166\,\text{GHz}$ and the same pump settings as in Figure 3.2, i.e. $G = 23\,\text{dB}$ and 50 MHz bandwidth. (left) Dependence of output power on the input power on a log-log plot. Output power is in arbitrary units and the input power is expressed as the mean photon number that relates to the power as $P_{in} = \langle n \rangle \times \kappa_{source} \times \hbar\omega_{signal}$ with $\kappa_{source} \approx 2\pi \times 4.5\,\text{MHz}$. Blue points are measurements, red line is a fit to the first eight points. (right) Difference between the linear fit and the measurements showing 1 dB compression in gain at $\langle n \rangle \approx 50$.

4 Dispersive Measurement

The interaction term in Eq. 2.3 can be thought of as a qubit state dependent shift $\chi\sigma_z$ on the resonance frequency ω_r', as suggested by the choice of factoring. The qubit state can therefore be inferred by measuring the transmission through the resonator at some frequency ω_m. Since the interaction term commutes with the qubit Hamiltonian, the qubit will stay in the measured state indefinitely in the absence of additional relaxation mechanisms that make $T_1 < \infty$. This is why the measurement is called quantum non-demolition (QND), which means that the measurement projects the qubit to a stationary state with respect to the ideal Hamiltonian and does not cause further qubit transitions [Clerk10]. This means that in the absence of relaxation, repeating the measurement always returns the same value as the initial measurement.

This section discusses what kinds of noise affect the measurement and what choices related to measurement pulse shapes and filtering of the output signal must be made. The simplest case of a noisy binary signal and phase-insensitive measurement with infinite bandwidth has already been discussed in detail in [Gambetta07]. This section presents some of those results but for practical application of the results it is necessary to also understand the unavoidable effects of analog filtering of the signal due to the resonator and the phase-sensitive finite bandwidth detection chain. The ensemble averaged effects of the resonator have been calculated and experimentally measured in [Bianchetti09] for continuous and step pulse measurements. Here the emphasis will be on the character of the mean amplitude in single shot traces and on the use of a two step measurement pulse that minimizes the signal rise time. These aspect are relevant when the signal-to-noise ratio is large enough that qubit decay starts to dominate the measurement errors before the signal reaches its steady state amplitude.

The experimental setup used for the measurements is shown in Figure 4.1 and the numerical values of the parameters are listed in Table 4.1. The values for the qubit-resonator system are typical for current cQED experiments, except

cQED parameter	Value
$\omega_r'/2\pi$	7.1655 GHz
$\kappa/2\pi$,$1/\kappa$	4.5 MHz, 35 ns
$\omega_{01}'/2\pi$	6.2750 GHz
$\alpha/2\pi$	450 MHz
$\lvert\delta C_{01}\rvert/2\pi$	140 kHz
$g_{01}/2\pi$	76 MHz
$\chi/2\pi$	2.2 MHz
n_{crit}	34
T_1	300 ± 10 ns
$T_2^\star$	610 ± 20 ns

Paramp parameter	Value
$\omega_{pump}/2\pi$	7.1634 GHz
G	23 dB
3 dB bandwidth	50 MHz
1 dB compression point	$20 \times \kappa \times \hbar\omega_{pump}$
Added noise	$\sim 2-3$ photon/s/Hz

Table 4.1: Numerical values of the most important sample parameters. Definitions are given in the text.

for the low T_1. The values for the paramp on the other hand are extraordinary compared to commercial HEMT amplifiers which have a large bandwidth but add at least 20 photons/s/Hz of classical noise into the signal.

4.1 Signal to Noise Ratio

Consider an output mode $a_{out}(t)$ that is in a coherent state $|\alpha(t)\rangle$ but $\alpha(t)$ is unknown. The signal to noise ratio (SNR) then describes how quickly one can reliably determine $\alpha(t)$ by measuring $a_{out}(t)$. The definition is:

$$\mathrm{SNR}(t) = \frac{\left|\frac{1}{\sqrt{T}}\int_t^{t+T} dt' \alpha(t')\right|^2}{\left\langle\left|\frac{1}{\sqrt{T}}\int_t^{t+T} dt' \left[a_{out}(t') - \alpha(t')\right]\right|^2\right\rangle} \tag{4.1}$$

where $a_{out}(t)$ is the measured noisy signal.[1] Here the numerator corresponds to the integral of the coherent part of the signal amplitude while the denominator is a measure of the variance in the integrated signal due to noise. The fluctuations in the measured values $a_{out}(t)$ are always at least as strong as the so called vacuum fluctuations required by quantum mechanics but they typically also include classical non-Heisenberg noise, which is introduced by imperfect components but is in principle avoidable.[2]

Because they are both unavoidable and spectrally white, the vacuum fluctuations also provide a useful scale for expressing the total noise present in $a_{out}(t)$.

[1] In general, one can also apply more complicated filter functions than the box car filter in Eq. 4.1 in order to increase the SNR for specific types of noise and specific types of time varying signals.

[2] The vacuum fluctuations can also be squeezed so that they are small in one quadrature and large in the other but this is not the case for the signal sources relevant here.

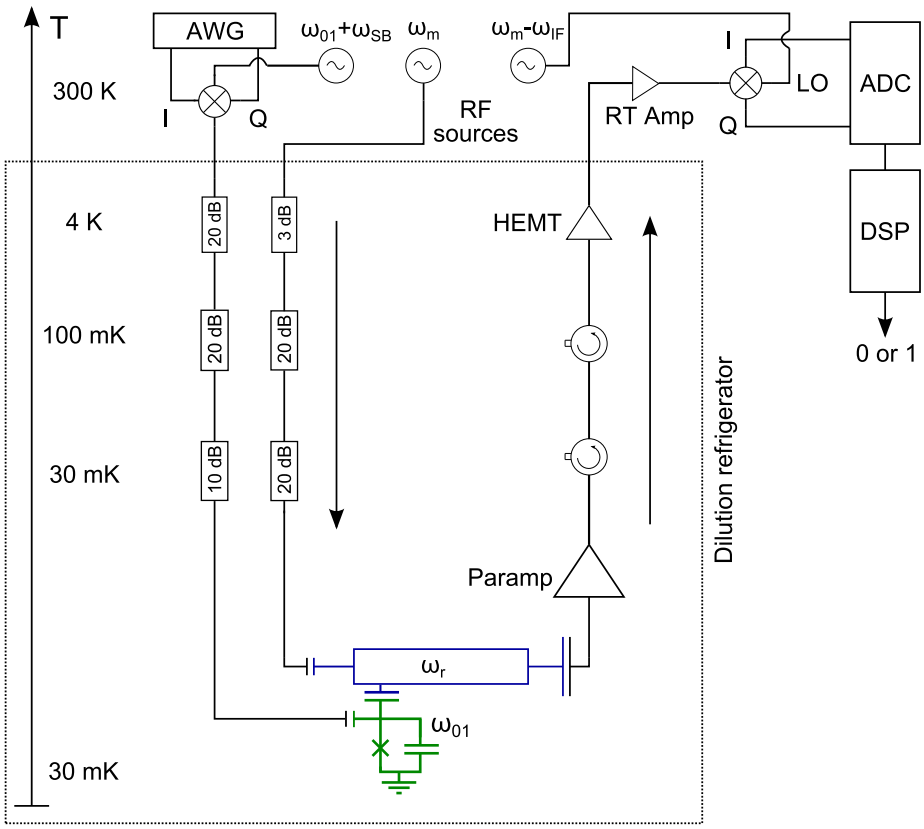

Figure 4.1: Experimental setup. The qubit (green) is coupled to a transmission line resonator (blue) at a base temperature of 30 mK. The qubit and resonator can be excited independently via two input RF lines. In the simpler step pulse measurements, only the qubit line is connected to an IQ mixer that up-converts a pulse generated by an arbitrary waveform generator (AWG) while the resonator input can only be turned on and off. The resonator output is connected to a paramp, described separately in Figure 3.1. The attenuators and isolators are necessary to prevent thermal radiation leaking in via the cables. At room temperature the signal is down-converted using an IQ mixer and digitized using an analog-to-digital converter (ADC). The sampled signal is then processed digitally (DSP) and finally reduced to a single bit indicating the best guess for the initial qubit state. Note that in measurements with more complicated measurement pulses the resonator input is also connected to an up-converting IQ mixer and an AWG.

In particular, the $1/\sqrt{T}$ normalization chosen for the integrals in Eq. 4.1 makes the denominator for a pure coherent state $|\alpha(t)\rangle$ exactly one, independent of T. For example for the vacuum state $|0\rangle$:

$$\left\langle 0 \left| \left| \frac{1}{\sqrt{T}} \int_t^{t+T} dt' a_{out}(t') \right|^2 \right| 0 \right\rangle = \frac{1}{T} \left\langle 0 \left| \int_t^{t+T} dt' \int_t^{t+T} dt'' \delta(t'-t'') \right| 0 \right\rangle$$
$$= 1$$

The numerator on the other hand grows linearly with T and corresponds to a power in units of photons. Therefore the noise added by the vacuum fluctuations is defined as $1\,\mathrm{photon/s/Hz} = 1\,\mathrm{photon}$. This also implies that SNR goes to zero for $T \to 0$ and vice versa for $T \to \infty$ regardless of the strength $\alpha > 0$ of the signal.

It is important to realize that the degradation of SNR due to amplifiers is determined largely by the first amplifier since the noise added by later (good) amplifiers does not scale with the signal power they receive. Therefore, relative to the amplified initial noise in the signal, the noise contribution of the second amplifier is scaled by $1/G_1$, where G_1 is the power gain of the first amplifier. Another contribution to SNR comes from the damping of the signal before it reaches the first amplifier but, in practice, this is usually difficult to distinguish from added noise and can instead be thought of as a renormalization of the noise added by the amplifiers. However, in some cases it is important to realize that damping treats classical and vacuum fluctuations differently since the former can be damped while the latter cannot.

4.2 Steady State Response

Let us begin with the case of infinite T_1 and a measurement pulse at frequency ω_m with an amplitude envelope $i\epsilon_m(t)/\kappa$. As stated in Section 2.3, the resonator Hamiltonian is

$$H_r = \hbar(\omega_r' \pm \chi) a^\dagger a$$

depending on the qubit state. The expectation value $\alpha(t) \equiv \langle a(t) \rangle$ of the intra-cavity field then evolves according to Eq. 3.3:

$$\frac{d\alpha(t)}{dt} = -i(\omega_r' \pm \chi)\alpha(t) - \frac{\kappa}{2}\alpha(t) - i\frac{\epsilon_m(t)}{\sqrt{\kappa}} e^{i\omega_m t}.$$

where $\epsilon_m(t) \in \mathbb{C}$ is the envelope of the measurement pulse used to drive the resonator. In a frame rotating at the measurement frequency ω_m this becomes:

$$\frac{d\alpha(t)}{dt} = i(\Delta_{mr} \mp \chi)\alpha(t) - \frac{\kappa}{2}\alpha(t) - i\epsilon_m(t) \tag{4.2}$$

where $\Delta_{mr} = \omega_m - \omega_r'$. This describes the response of a classical damped harmonic oscillator at frequency $-\Delta_{mr} \pm \chi$ to a driving signal $i\epsilon_m(t)$. The

expectation value of the output signal is given by the solution to this equation and Eq. 3.6. Alternatively, one can Fourier transform the above equation and see that $\alpha_{out}(t) = \langle a_{out}(t) \rangle$ consists of $\epsilon_m(t)$ filtered by a Lorentzian filter centered at $\Delta'_{mr} = \Delta_{mr} \mp \chi$ with FWHM of κ:

$$\alpha_{out}(\omega) \equiv \mathcal{F}[\alpha_{out}(t)] = \frac{\sqrt{\kappa}\epsilon_m(\omega)}{\Delta'_{mr} - \omega + i\frac{\kappa}{2}}. \tag{4.3}$$

Now consider the response of this system to a step measurement pulse $\epsilon_m(t) = \epsilon_m \theta(t)$. After a few resonator decay times $2/\kappa$ the resonator reaches steady state and the SNR is determined by two factors. First, the magnitude of the signal is determined by the difference of the mean steady state values α_g and α_e of the transmitted coherent signals corresponding to the ground and excited states of the qubit. This is set on one hand by the phase difference $\arg(\alpha_e/\alpha_g)$, which depends on Δ'_{mr}/κ, and the amplitude, which—according to Eq. 3.5—depends on the intra-cavity amplitude and κ. The phase difference is maximized by choosing $\omega_m = \omega'_r$ as can be seen by plotting the phase of the transmitted signal

$$\arg(\mathcal{T}) = \arg\left(\frac{1}{\Delta'_{mr} - i\kappa}\right)$$

where $\mathcal{T}$ is the transmission coefficient for a Lorentzian filter. Note that the amplitude of $\mathcal{T}$ is not directly relevant as lower transmission can be compensated by a stronger drive amplitude ϵ_m.[3] Instead the signal amplitude is limited on one hand by the limited dynamic range of input powers that the paramp can handle shown in Figure 3.3, and on the other hand by the limited intra-cavity occupation allowed by the dispersive approximation, which requires that $\Delta \gg g_{ij}\sqrt{\langle n \rangle}$. For the parameters used in these experiments the best results were obtained for $\langle n \rangle \sim 20$, which is of similar magnitude as both limits and it is not clear whether one of them dominates the other.

The second factor in the SNR is the magnitude of the noise determined by the averaging time T. As described previously, in the limit $T \to \infty$ the SNR diverges and therefore the qubit could be measured with arbitrary precision for any $|\alpha_e - \alpha_g| > 0$. In practice, however, a limited T_1 implies that after some jump time τ_J with statistics $e^{-\tau_J/T_1}/T_1$ the signal no longer reflects the original state of the qubit if it was initially in the excited state. In order to maximize the fidelity between the output signal and the initial qubit state it is therefore necessary to choose a T that balances the derivatives of the contributions to infidelity coming from states misidentified due to relaxation and those due to noise around the correct mean value. This intuition is confirmed by the theoretical results in [Gambetta07] that show that for low SNR the timescale of the optimal linear filter (i.e. weighted average) is given by T_1, while for higher and higher SNR the filter gives more and more weight to the early values.

Despite initial optimism in [Blais04] and [Gambetta07], high fidelity dispersive single-shot measurements have not been demonstrated experimentally due

[3] Strictly speaking the ratio $|\alpha_e|/|\alpha_g|$ is still relevant but it is reasonable to require that $\max\left(\left|\langle n \rangle_e\right|, \left|\langle n \rangle_g\right|\right)$ satisfies the discussed limits on the intra-cavity field.

to limited T_1, limited signal power, and the fact that even the best commercial cryogenic amplifiers add at least 10 to 20 photons of noise to the signal. Since it is optimal to integrate for a time similar to T_1 in such low SNR situation, the steady state approach described in this section as well as the instantaneous jump model in [Gambetta07] are appropriate.

On the other hand, the noise added by the paramp is comparable to the vacuum fluctuations so SNR can be high enough that the optimal measurement time is much shorter than T_1. Furthermore, the optimal measurement time can also be similar or short compared to the resonator rise time $2/\kappa$. For the parameters shown in Table 4.1 and a measurement power corresponding to $\langle n \rangle \approx 20$ this is certainly the case as can be seen in the measured signals shown later in Section 4.5. Therefore a more detailed understanding of the initial resonator dynamics will be necessary to understand the limiting factors on the measurement fidelity.

4.3 Non-Steady State Resonator Dynamics

The average resonator response to a step pulse is calculated and experimentally measured in [Bianchetti09] for different detunings and different initial qubit states. This is sufficient for measuring the ensemble averaged $\langle \sigma_z \rangle$ but for single-shot readout it is more useful to understand what the individual traces look like. When the qubit is prepared in the ground state, the response is indeed well described by the average response plus uncorrelated noise filtered to a certain bandwidth, but for the excited state traces this is not the case. When an excited state is prepared the deviations from the average are biased in a step like manner, i.e. the signal deviates systematically one way before some qubit decay time τ_J and the opposite way afterwards. If the SNR is low this is not significant but with the paramp these jumps can in fact be seen as shown later.[4] This is important because it implies that the information density per unit time for a single measurement trace is not directly proportional to the SNR calculated for the ensemble averaged signals. More concretely, the SNR reaches its maximum on a time scale set by T_1 but for high fidelity measurements there is little new information to be gained at times much later than $T_1 \log(F)$, where $F \leq 1$ is the measurement fidelity defined precisely later. This is because the misidentified traces at later times are due to early qubit decay which hence become unidentifiable from the ground state response on a time scale of $2/\kappa$.

The response to a step pulse in the ideal case $T_1 \to \infty$ can be calculated by solving Eq. 4.2 assuming that the resonator is initially in its ground state. The solution is

$$\alpha_{out}(t) = \frac{1 - e^{-\frac{\kappa t}{2}} e^{i\Delta'_{mr} t}}{\Delta'_{mr} + i\frac{\kappa}{2}} \epsilon_m \theta(t)$$

and is plotted in Figure 4.2 for a qubit state dependent detuning $\Delta'_{mr} = \pm\chi$. Evidently, the steady state for any detuning is reached on a timescale determined

[4] See Figure 4.5 for a sneak peek.

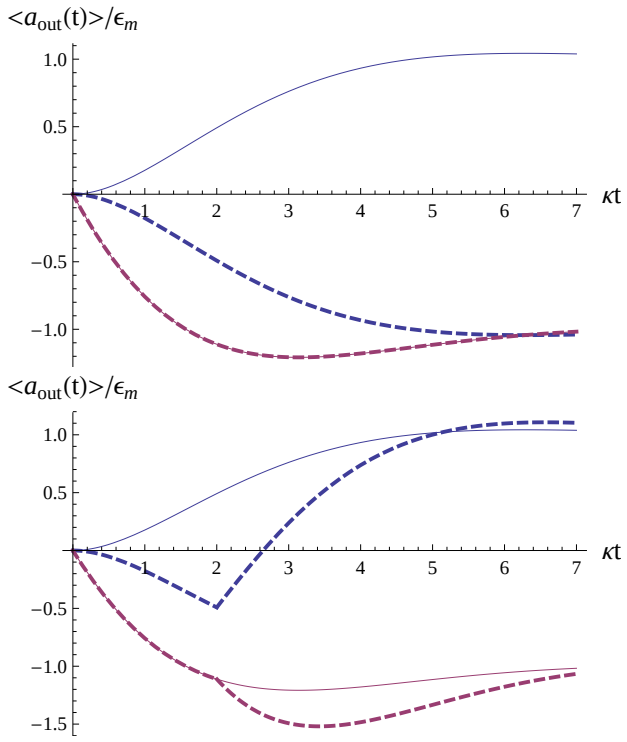

Figure 4.2: (top) The real (blue) and imaginary (red) parts of the expectation value of the output field of a resonator driven by a step pulse with a qubit state dependent detuning of $\pm\chi = \pm\kappa/2$ (solid/dashed) from the resonator frequency. The imaginary parts overlap perfectly for the two cases. (bottom) The same but with the detuning for the dashed (excited state) response switching sign at $t = 2/\kappa$ due to qubit decay.

by $2/\kappa$ but in addition the off-resonant cases show ringing at frequency $\Delta'_{mr} = \pm\chi$. Furthermore, as also noted in [Bianchetti09], the signal in the resonant case is in one quadrature for all times while in the off-resonant cases $\arg(\alpha_{out}(t))$ depends on t and the qubit decay time τ_J. Note, however, that for the case $\Delta'_{mr} = \pm\chi$ and for times $t < \tau_J$ the difference of the ground and excited state responses is still only in one quadrature. Since the transients entangled with the initial qubit state decay after τ_J on a time scale of $2/\kappa$ in any case, there is no advantage to choosing a measurement frequency that corresponds to one of the shifted resonance frequencies, even when using a phase-sensitive amplifier like the degenerate paramp.

4.4 On-Hold Sequence

Observe that the initial gradients of the quadrature signals in Figure 4.2 are the same. This feature is evidently common to all detunings since Δ'_{mr} contributes to $\frac{d}{dt}\alpha(t)$ in Eq. 4.2 only together with a factor of $\alpha(t)$, which initially grows linearly with $i\epsilon_m$. This means that the qubit dependent part of the signal initially grows only as $\epsilon_m \chi t^2$, which is especially detrimental for measurements with high SNR where the measurement fidelity is limited by qubit decays occurring at early times τ_J.

This observation suggests that it would be useful to populate the resonator quickly in order to entangle the qubit state faster with the output field. This can be done by prepending the measurement pulse with a strong square pulse of amplitude ϵ_p and width t_{on}. In the limit where $t_{on}\max(\kappa,\chi) \to 0$ and $\alpha(0) \equiv -i\epsilon_p t_{on}$ remains constant, this is equivalent to changing the initial conditions of the resonator from the vacuum state to the coherent state $|\alpha(0)\rangle$, independent of Δ'_{mr}. Eq. 4.2 implies that the initial difference in the gradients is then

$$\frac{d}{dt}\alpha_e(0) - \frac{d}{dt}\alpha_g(0) = -2i\chi\alpha(0) \tag{4.4}$$

regardless of the detuning Δ_{mr}.

Eq. 4.4 shows that the signal could grow arbitrarily quickly if arbitrarily large values of $|\alpha(0)|$ were achievable. This is demonstrated in Figure 4.3. However, the amplitude of the initial state is restricted for the same reasons that were previously evoked to argue why the steady state amplitude is restricted. Nevertheless, even when restricting $|\alpha(0)|$ to the ground/excited state steady state magnitude given by $\mathcal{A}_{g/e} = \epsilon_m/(\Delta'_{mr} + i\kappa/2)$, the initial scaling of the information containing part of the signal is improved from quadratic to linear with a gradient of $2\left|\mathcal{A}_{g/e}\right|\chi = 2\epsilon_m\left|\chi/(\Delta'_{mr} + i\kappa/2)\right|$. This is maximized by measuring at the bare resonator frequency where $\Delta'_{mr} = \pm\chi$.

With the amplitude $|\alpha_0|$ fixed, a non-trivial choice of the phase of the initial state still remains. A reasonable goal is to require that

$$\arg\left(\frac{d}{dt}\alpha_e(0) - \frac{d}{dt}\alpha_g(0)\right) = \arg\left(\mathcal{A}_e - \mathcal{A}_g\right).$$

This condition ensures that the degenerate paramp is able amplify the quadrature of the signal that is entangled with the qubit state both in the short and long time limits. In fact, if $\Delta'_{mr} = \pm\kappa$ is chosen, it is possible to satisfy the above condition for all $t < \tau_J$ as was the case for the initially empty resonator. This can be seen by choosing $\epsilon(t) \in \mathbb{R}$ and $\alpha(0) = -i\left|\mathcal{A}_g\right|$, which makes the initial conditions and the equations of motion given by Eq. 4.2 for $\alpha_e(t)$ and $-\alpha_g^\star(t)$ identical. This implies that $\alpha_e(t)$ and $\alpha_g(t)$ at any time $t < \tau_J$ differ only by the sign of the real part as shown in Figure 4.3. This again confirms that there is no fundamental disadvantage to using a phase-sensitive amplifier.

For more complex experiments that continue after the first measurement, it would also be useful to apply an off-pulse that displaces the resonator back to the vacuum state on a time scale faster than $1/\kappa$ after the qubit state has been

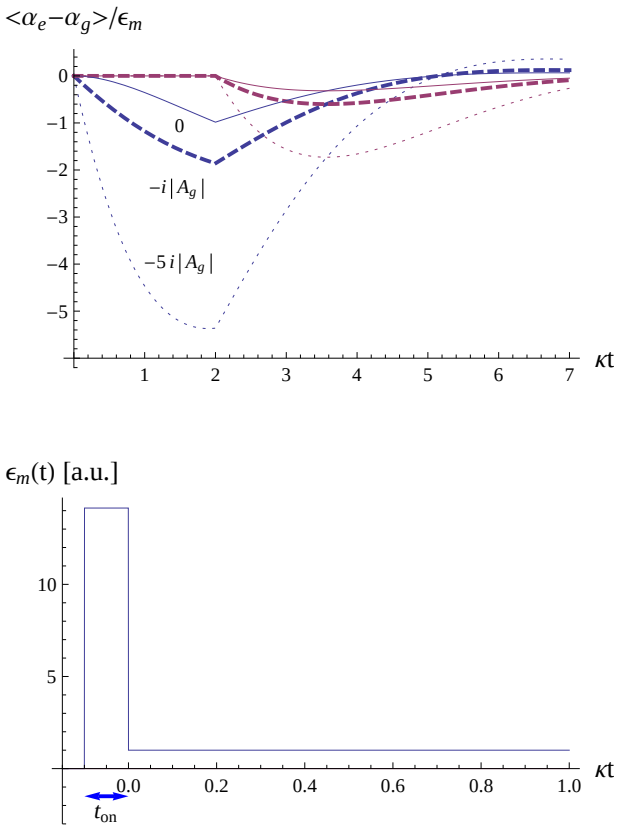

Figure 4.3: (top) The the real (blue) and imaginary (red) parts of the difference of ground and excited state responses of a resonator driven by a step pulse at a detuning $-\kappa/2$ until $\tau_J = 2/\kappa$ when the detuning switches sign due to qubit decay. The three cases shown correspond to initial states of $\alpha(0) \in \{0, -i|\mathcal{A}_g|, -5i|\mathcal{A}_g|\}$ (solid, thick dashed, thin dashed). (bottom) The pulse envelope $\epsilon_m(t) \in \mathbb{R}$ normalized to the steady state value $\epsilon_m(t \to 0)$ for $\alpha(0) = |\mathcal{A}_g|$, $\kappa t_{on} = 0.1$, and the values of κ and χ shown in Table 4.1. Note the different time scales in the plots.

reliably determined. However, determining the appropriate pulse amplitude and phase would require feedback from the measurement signal since the state of the resonator at times $t \gtrsim 1/\chi$ depends on the qubit state. Such feedback was not used in this thesis but could be a simple and practically relevant feedback experiment in the future.

In practice, the approximation that $t_{on} \max(\kappa, \chi) \to 0$ is not necessarily accurate due to the finite bandwidth of the electronics that generate the input signals. In the experiments presented here the measurement signal $\epsilon_m(t) e^{i\omega_m t}/\kappa$ was generated using the left sideband of a mixer that was driven by a local oscillator at $\omega_{LO} = \omega_m + \omega_{SB}$, where $\omega_{SB}/2\pi = 120\,\mathrm{MHz}$, and a Tektronix AWG5014 arbitrary waveform generator responsible for the I and Q modulation signals. In this setup the $1.2\,\mathrm{GS/s}$ sampling frequency of the arbitrary waveform generator sets a fundamental upper bound of 600 MHz on the bandwidth of the generated signals. This means that the rise time of $\epsilon_p(t)$ is at least some nanoseconds, which is not completely negligible compared to $1/\kappa \sim 1/2\chi \sim 35\,\mathrm{ns}$. This means that the resonator state will already become entangled with the qubit state during the on-pulse. This is not a fundamental problem for the scheme but it means that the linear scaling of the signal only kicks in on a time scale determined by t_{on}. Note, however, that even during the on-pulse the drive $\epsilon_m(t)$ remains real. Therefore the aforementioned symmetry for the $\Delta'_{mr} = \pm\kappa$ case is not violated and the signal remains in one quadrature.

Finally, note that the measurement pulses used for the data sets analyzed in this thesis unfortunately only included a modest on-pulse of approximately twice the amplitude of the hold-pulse and a length of 4 to 10 ns. Due to time constraints and technical problems with the subsequent samples it was not possible to investigate in detail how much the fidelity could be improved with a more aggressive on-pulse, such as that shown in Figure 4.3. Nevertheless, even this modest on-pulse reduced the number of misidentified traces by roughly a quarter compared to the usual step measurement pulse. Specifically, $1 - F$ was improved from approximately 0.2 to 0.15, where F is the fidelity defined below in Eq. 4.5.

4.5 Filtering

The goal of single shot measurement is to map the measurement trace to a single bit yes or no answer that should be maximally correlated with the initial state of the qubit. More precisely, the goal here is *defined* as maximizing fidelity

$$F \equiv 1 - P(\tilde{g}|e) - P(\tilde{e}|g) \tag{4.5}$$

where $\tilde{g}$ and $\tilde{e}$ indicate the possible answers given by the measurement, g and e indicate the true initial states of the qubit, and the prior probabilities are assumed to be $P(g) = P(e) = 1/2$. The operational meaning of this definition is that the probability of incorrectly identifying the qubit state given no prior information of it is $(1 - F)/2$. Maximizing F is therefore a reasonable goal but one should realize that it is not in general optimal if some prior knowledge of

the qubit state is available or the goal is to minimize something other than the total error probability.

Physically, the process of converting the resonator output signal $a_{out}(t)$ to a single bit answer consists of analog filtering, sampling of the analog signal using an analog-to-digital converter (ADC), digital signal processing, and finally assignment of the answer based on whether the processed signal exceeds some threshold s_{th}. Mathematically, this process can be decomposed as $\theta(f_d \circ f_a(a_{out}(t)) - s_{th})$, where f_a is a functional that maps the resonator output signal to vector of complex numbers and f_d is an arbitrary function that maps that vector to a scalar, which the step function θ finally maps to a single bit. Here f_a corresponds to the analog filtering and the signal digitization and f_d corresponds to the digital signal processing. This decomposition is useful because f_a can be described approximately as a phase-sensitive linear filter of bandwidth B that is unavoidable and hard to change. The digital signal processing f_d on the other hand can be specifically optimized for fidelity and is limited in complexity only by computational power.

Filtering of dispersive measurement signals is described in theory in [Gambetta07]. However, there the signal is assumed to consist of Gaussian white noise around a mean value that jumps *instantaneously* at the qubit decay time τ_J, hence ignoring the non-zero rise time of $a_{out}(t)$ due to $\chi < \infty$, the non-zero decay time of transients due to $\kappa < \infty$, as well as the unavoidable analog filtering due to f_a. Therefore the filtering schemes presented in [Gambetta07] are not directly applicable except in cases where T_1 is very long and SNR only moderate so that the values output by the proposed filters are not sensitive to fluctuations on time scales of $\max(1/\kappa, 1/B)$. Since the setup used for the measurements in this thesis have the opposite parameters, it is necessary to take these physical limitations into account.

This section begins with a discussion of how the raw data digitized by the ADC looks like and what preprocessing steps are applied to get the data into standard form before more complicated digital filtering. Linear filtering and experimental data processed by a numerically optimized filter are then discussed. Finally, the formally optimal digital filter is introduced and applied in a simplified form.

4.5.1 Analog Filtering and Digital Preprocessing

The raw digitized signal $f_a(a_{out}(t))$ for a single measurement trace consists of a vector of complex numbers, where the real and imaginary parts of each number correspond to the quadratures of the analog filtered signal sampled every 10 ns. In principle f_a has many contributions as shown in Figure 4.1: the paramp at 30 mK, circulators for thermal isolation, a cryogenic high electron mobility transistor (HEMT) amplifier at 4 K, warm amplifiers at room temperature, a frequency down converting IQ mixer, cables and attenuators, and finally the ADC. In practice, however, f_a is dominated by the paramp since it has a much

smaller bandwidth than all the other components[5] and high gain so that the SNR is dominated by noise at the paramp output. At the end, the effect of f_a can therefore be described approximately as a linear filter with a passband shown in Figure 3.2.

As explained in Chapter 3, the paramp needs to be pumped at the signal frequency. Most of this pump contribution is subtracted from the output signal using a directional coupler and a strong displacement tone as explained in Appendix A. Nevertheless, a necessary first step in the digital signal processing is to subtract the DC offset that remains in $f_a(a_{out}(t))$ even when $a_{out}(t)$ is in the vacuum state. Following this step the data is rotated in the IQ plane so that the amplified quadrature is aligned with the Q axis because the original angle is referenced to the arbitrary phase of the local oscillator used in the frequency down conversion step. The only non-trivial question here is the time scale on which these quantities should be determined. Here the same rotation angle was used for all traces taken over several minutes while the DC offset was updated with an exponential low pass filter with a time constant of the order of milliseconds.[6] Finally, the digital preprocessing also corrects a technical problem in the current setup that caused entire sets of traces to be rotated by exactly ± 90 or 180 degrees.[7] The combined effect of these preprocessing steps is shown in Figure 4.4 for a fixed time and will be considered part of f_a from now on. Furthermore, taking the imaginary part will also be considered part of f_a. A few individual traces and the mean responses after these steps are shown in Figure 4.5.

4.5.2 Numerically Optimized Linear Filter

A linear digital filter f_d is defined by a set of T coefficients $\tilde{f} = \left\{\tilde{f}_i\right\}$ and the rule:

$$f_d(s(t)) = \sum_{t'=t}^{t+T-1} \tilde{f}_{t'-t} s(t').$$

In addition $\sum \tilde{f}_i = 1$ is imposed here for consistent normalization of the filtered data.

The most straightforward way to find the optimal coefficients $\tilde{f}$ is to numerically optimize the resulting fidelity for a set of sample data points by varying the offset time t, the coefficients $\tilde{f}$, and the signal threshold s_{th}. In other words, one needs to find the global maximum of F in $\mathbb{Z} \otimes \mathbb{R}^T$, where $\mathbb{Z}$ corresponds to the

[5] With the exception of the relatively slow 100 MS/s ADC, which also has approximately 50 MHz bandwidth. It is however relatively easy to switch to a faster ADC if the paramp bandwidth is further improved in the future.

[6] This time constant is so long because only a short time interval (~ 200 ns) of data was recorded in each trace before the measurement pulse was turned on. Therefore the estimation of the DC offset has to be done over many traces.

[7] This problem is easy to identify by computing $\phi = \arg\left(\int_0^{t_{start}} dt\, f_a(a_{out}(t))\right)$, where t_{start} is the measurement start time. Each trace can then be rotated by ± 90 or 180 degrees depending on $\phi - \langle\phi\rangle$. This method works because of the imperfect analog pump cancellation.

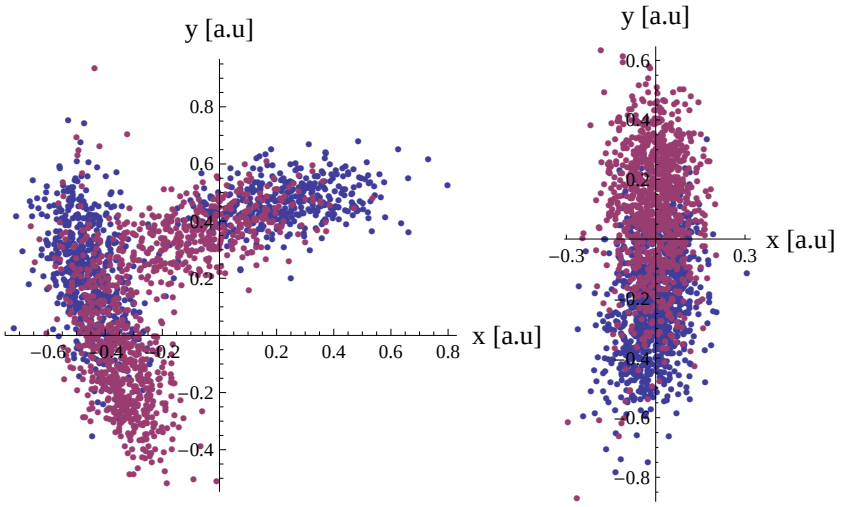

Figure 4.4: (left) The x and y quadrature values output by the ADC at a fixed time for many ground (blue) and excited (red) state traces. (right) The data after basic preprocessing that rotates all the information into Q and subtracts the remaining pump offset after the analog displacement. This step also corrects for a technical problem that occasionally rotates the entire trace by 90 or 180 degrees.

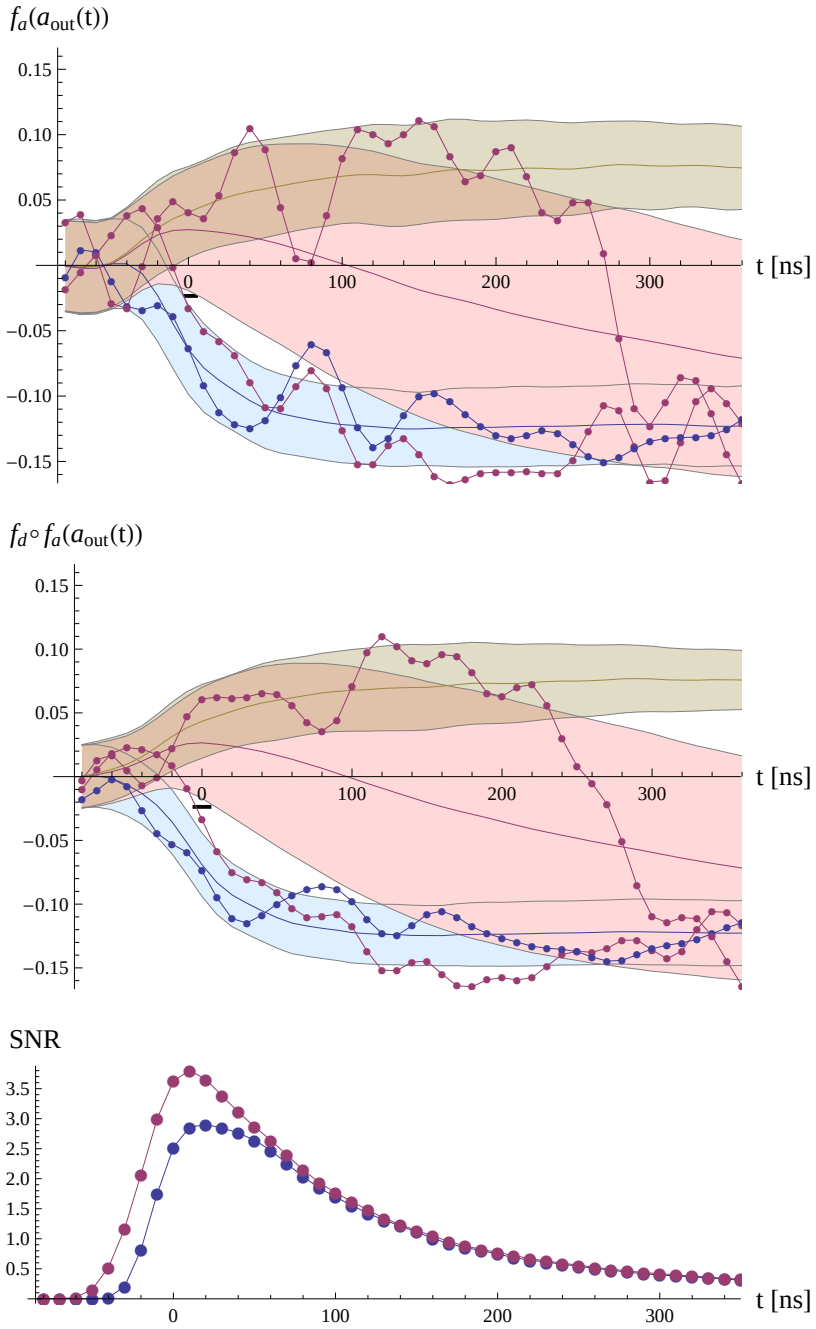

Figure 4.5: (top) Two example excited state traces (red) and one ground state trace (blue). The filled regions indicate 2σ bands around the mean values (solid lines) for the ground states (blue), excited states (red), and the excited states that were still above zero at $t = 400\,\text{ns}$. The optimal time to discriminate the traces is at $t = 0$ using the threshold indicated in black. (middle) Same but after applying a numerically optimized linear 3-point digital filter f_d. (bottom) SNR before (blue) and after (red) applying f_d. The lines connecting the points are guides to the eye.

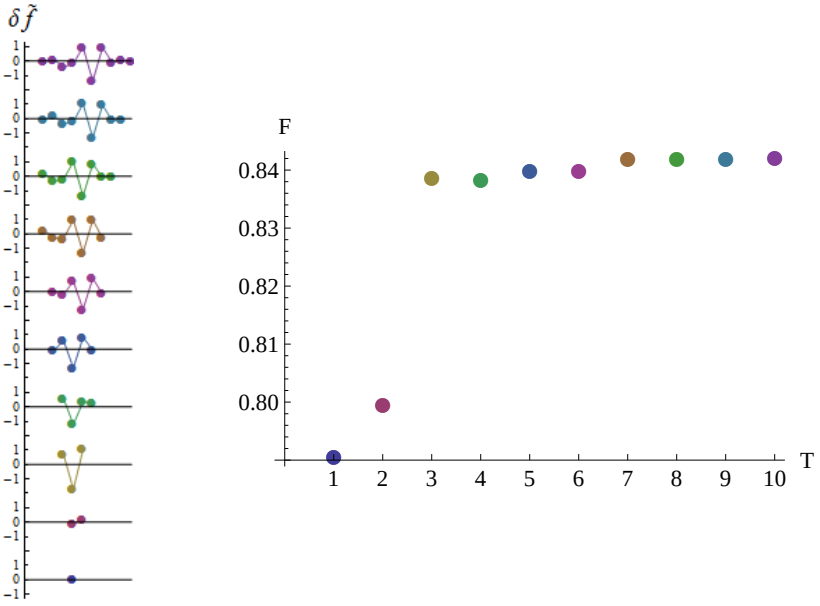

Figure 4.6: (left) Deviations of numerically optimized filters from a constant box car filter. (right) The resulting fidelities. The highest fidelity is 0.842.

offset time t and $\mathbb{R}^T$ corresponds to s_{th} and the $T-1$ independent elements of $\tilde{f}$. Such a task is a common optimization problem and has no efficient solution in the general case. In this specific case the problem is solvable because F is not expected to have many deep local minima for a fixed t and t can be restricted to a few reasonable values. On the other hand, T is restricted to small values because evaluating F is computationally expensive and it is not smooth, which makes methods relying on derivatives problematic.

The algorithm chosen here perturbs the coefficients $\tilde{f}$ randomly and systematically searches for the best threshold and offset time t with decreasing magnitude of perturbations and increasing threshold resolution. Whenever the fidelity is improved the search around the new better values is started from the beginning. The process is terminated when the fidelity has not been improved for n_{cutoff} attempts. This is grossly inefficient but relatively robust because the algorithm does not get stuck in shallow local minima and therefore the end result is not very sensitive to the initial values. Numerically optimized filters for $T \leq 10$ and the resulting fidelities are shown in Figure 4.6.

Figure 4.6 shows a sudden drop of one fifth in the infidelity $1-F$ when going from $T = 2$ to $T = 3$ but little improvement beyond that. This implies that qubit decay dominates as a source of infidelity already at an averaging time of

30 ns, corresponding to

$$SNR = \frac{\left|\langle f_d \circ f_a\left(a_{out}\left(t\right)\right)\rangle_e - \langle f_d \circ f_a\left(a_{out}\left(t\right)\right)\rangle_g\right|^2}{\sum_{i_0 \in g,e}\left\langle\left|f_d \circ f_a\left(a_{out}\left(t\right)\right) - \langle f_d \circ f_a\left(a_{out}\left(t\right)\right)\rangle\right|^2\right\rangle_{i_0}}$$

of approximately 4 as shown in Figure 4.5. The mean responses and standard deviations shown in Figure 4.5 also support this conclusion. The increasing variance in the excited state response and the constant variance in the ground state response can be naturally explained by unbiased vacuum and thermal noise around both responses and an additional systematic τ_J dependent bias for the excited state responses. This is also confirmed by inspecting individual traces that show variations in accordance with the error bars for the ground state responses but not at all so for the excited state responses. The modest decrease of the variances achieved by applying the numerically optimized 3-point filter f_d is explained by the fact that the analog filtering has already caused noise to be correlated on a time scale of a few dozen nanoseconds. This is also evident from the individual traces.

All of the filters shown in Figure 4.6 show a $\{1, -1, 1\}$ pattern near the optimal measurement time for a 1-point filter, although the exact location of the pattern moves by one point at $T = 6$ and then by another at $T = 9$. The emergence of this pattern at $T = 3$ also coincides with a significant decrease in infidelity as mentioned. This can be understood by looking at the individual traces in Figure 4.5 and noting that the bandwidth of the signals is much smaller than the bandwidth of the fluctuations, which for unbiased noise would imply that a much better signal to noise ratio would be achieved by increasing the averaging window until it starts to significantly distort the signals. However, because of the steplike biased noise due to qubit decay, increasing the averaging time is detrimental for identifying those states that decay within the averaging window. For $T > 2$ the bandwidth can, however, be reduced not just by averaging but also by giving positive (negative) weight to positive (negative) curvature. This is precisely what the pattern $\{1, -1, 1\}$ does, in addition to averaging.[8]

4.5.3 Optimal Filter

Optimal digital filtering consists of computing the probability of a particular initial state $i_0 \in \{g, e\}$ given the full measurement record $\Psi \equiv f_a\left(a_{out}\left(t\right)\right) \in \mathbb{R}^N$ and then returning $\tilde{e}$ if $P\left(e|\Psi\right) > P\left(g|\Psi\right)$ and vice versa. As also explained in [Gambetta07], these conditional probabilities are given by Bayes theorem:

$$P\left(i_0|\Psi\right) = \frac{P\left(\Psi|i_0\right)}{\sum_{i_0'} P\left(\Psi|i_0'\right)} \tag{4.6}$$

[8] In fact, such second derivative filters are commonly used for edge detection in image processing. The reason is essentially the same: they are effective in reducing the bandwidth of unbiased noise while minimally broadening jumps in the signal because of the small filter width.

where the prior probabilities $P(i_0)$ were assumed to be equal in accordance with the operational meaning of F. The issue with this approach in practice is that each $P(\Psi|i_0)$ is a map from $\mathbb{R}^N$ to $\mathbb{R}$ so even representing—let alone determining—these functions in the general case becomes exponentially harder as the number of samples N grows. In [Gambetta07] this problem is avoided by deriving analytic expressions based on the simplifying assumptions on $f_a(a_{out}(t))$ mentioned in the beginning of this section (Section 4.5). That allows efficient representation of $P(\Psi|i_0)$ even in the infinite sampling rate limit $N \to \infty$ but is not straightforward to generalize for the physically relevant case where the non-steady state dynamics of the resonator and f_a dramatically change the assumed steplike signal.

Here an alternative approach is taken where N is kept small and $P(\Psi|i_0)$ are estimated from n experimentally measured samples $\{s_i\}$ in a process known as kernel density estimation [Epanechnikov69]:

$$P(\Psi|i_0) = \frac{1}{n}\sum_{i=1}^{n} k\left(\frac{\Psi - s_i}{b}\right)$$

where k is a normalized kernel function, such as a Gaussian, and b is the smoothing bandwidth. This is similar to measuring the density by binning the measured values $\{s_i\}$ but gives smoother results for the same amount of data. On the other hand, using the above definition as such is not very convenient because applying it requires storing all the data points. However, since in the end it only matters whether $P(\Psi|g)$ is larger than $P(\Psi|e)$,[9] it is not necessary to store all the points but rather just an approximation of the $N-1$ dimensional boundary of the region where $P(\Psi|e) > P(\Psi|g)$. This was, however, not done in this thesis. Instead, each $P(\Psi|i_0)$ was evaluated on a grid once and future evaluations of the functions were performed by linear interpolation between the grid points. The loss of information caused by this procedure is not significant for an appropriately chosen grid.[10] This approach also has the advantage that it is easily parallelizable and straightforward to program, which makes it reasonable to implement it in practice in a field programmable gate array (FPGA) that constructs the probability densities in real time during calibration measurements.

Restricting N to small values is reasonable for the current experimental parameters because with high SNR and low T_1 most of the information is contained in the first few data points. Furthermore, it is also possible to apply some simple transformations to $f_a(a_{out}(t))$ before restricting N in order to include information from early as well as late times. Specifically, $N = 3$ was chosen by computing $\Psi = (\tilde{s}_1, \tilde{s}_2, \tilde{s}_3)$ by applying the numerically optimized linear 3-point filter $f_{d,t}$ to three intervals near the optimal measurement time t_0 for the three point filter. Specifically, $\tilde{s}_i = f_d \circ f_a(a_{out}(t_0 + 3(i-1)))$ was chosen. This cannot be rigorously justified as optimal but, roughly speaking, this choice ensures

[9] It is easy to see that comparing $P(i_0|\Psi)$ given by Eq. 4.6 to 1/2 is equivalent to this question.

[10] Fortunately, many tools such as *Wolfram Mathematica* provide built-in routines for automatically choosing the grid and determining the appropriate bandwidth b.

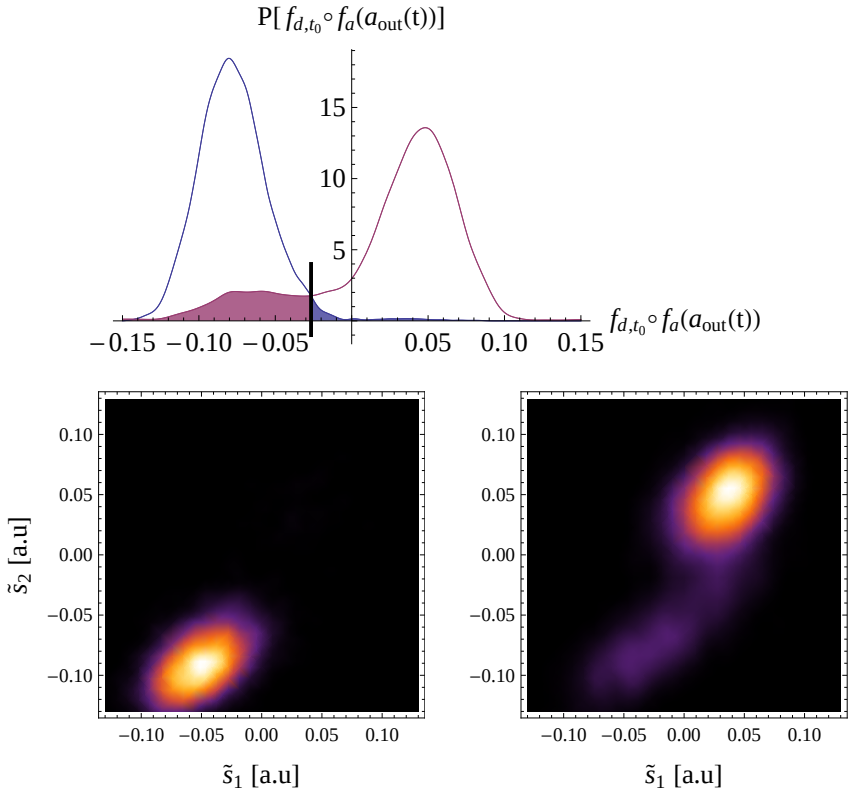

Figure 4.7: (top) $P(\Psi|g)$ and $P(\Psi|e)$ for $N = 1$, i.e. linear filtering. The areas of the blue and red regions correspond to $P(\tilde{e}|g) \approx 0.15$ and $P(\tilde{g}|e) \approx 0.02$, respectively. (bottom) The same for $N = 2$. Unfortunately it is hard to plot the densities in higher dimensions.

that the traces that stay excited past $t \approx t_0 + 5$ are always easily identified from the value of the last point while those excited state traces that decay very early have some chance of being distinguished from a ground state response if they show a particularly strong response in the first point. For comparison, the fidelities for the cases $N = 2$ with $\tilde{s}_i = f_d \circ f_a(a_{out}(t_0 - 2 + 3i))$ and $N = 1$ with $\tilde{s}_1 = f_d \circ f_a(a_{out}(t_0))$ were also computed.

The resulting fidelities for these Bayesian filters and the set of measured data are 0.830, 0.843, and 0.853 for N equal to 1, 2, and 3, respectively. The value for $N = 3$ represents a modest 7 % decrease in infidelity compared to the best numerically optimized linear filters shown in Figure 4.6. Pictorially, this indicates that the $N-1$ dimensional surface that divides the regions $P(\Psi|e) > P(\Psi|g)$ and $P(\Psi|e) < P(\Psi|g)$ is almost flat in the region where the densities are non-negligible. Therefore the densities $\{P(\Psi|i_0)\}$ can be projected onto a

line perpendicular to this surface without reducing the fidelity much.[11] This can be seen for $N = 2$ in Figure 4.7 where it is clear that projecting the distributions onto a roughly diagonal line has little effect on the overlap of the distributions.

4.6 QND Nature of the Measurement

Deviations from the ideal QND measurement can be characterized roughly as incomplete projection of the state onto the measurement basis states, complete projection but incorrect reporting of the result, and correct projection and reporting but subsequently changing the qubit state. The first two kinds of errors manifest themselves in $F < 1$ but the third type of errors could exist even if $F = 1$ in the single short measurements analyzed so far. Therefore it is necessary to look at the statistics of long continuous or short repeated measurements in order to determine the claimed QND nature of the measurement scheme.

The simplest analysis one can perform is to plot the decay of the ensemble averaged excited state population over a long time to see whether the relaxation rate T_1' or the steady state excited state population is affected by continuous measurement as predicted in [Boissonneault09].[12] Figure 4.8 shows examples of long measurements where so called quantum jumps are clearly visible in the data. The figure also shows the mean populations as computed from the traces filtered by the full measurement procedure, including the binary decision $\theta(s - s_{th})$. The mean excited state population is seen to reach a steady state value of approximately 1.0 % regardless of the initial state of the qubit. For qubits prepared in the excited state the time scale can be fit with high confidence and is found to be $T_1' = 303 \pm 2\,\mathrm{ns}$, which is within the error bars of $T_1 = 300 \pm 10\,\mathrm{ns}$ in the absence of measurement. For the traces where a ground state is prepared the steady state is reached on a faster time scale of $T_{g.s.} = 84 \pm 30\,\mathrm{ns}$. All of these results are in qualitative agreement with [Boissonneault09] that predicts a small deviation of T_1' from T_1 and a small excited state population for the case where the pure dephasing rate is negligible compared to the relaxation rate of the qubit.

4.6.1 Repeated Measurements

In future experiments the measurement pulse cannot be as long as shown in Figure 4.8 if any qubit in the resonator is to be used in further operations after the measurement. It is therefore useful to make the measurement pulse as short as possible without significantly compromising the readout fidelity. By applying such short pulses repeatedly at intervals that allow the resonator to relax to its ground state, one can also be sure that correlations between consecutive

[11] Recall that it only matters whether $P(\Psi|e) > P(\Psi|g)$, i.e. the fidelity is equivalent to the trace distance between the two distributions.

[12] T_1 in the absence of measurement can be determined independently by varying the delay between qubit preparation and measurement.

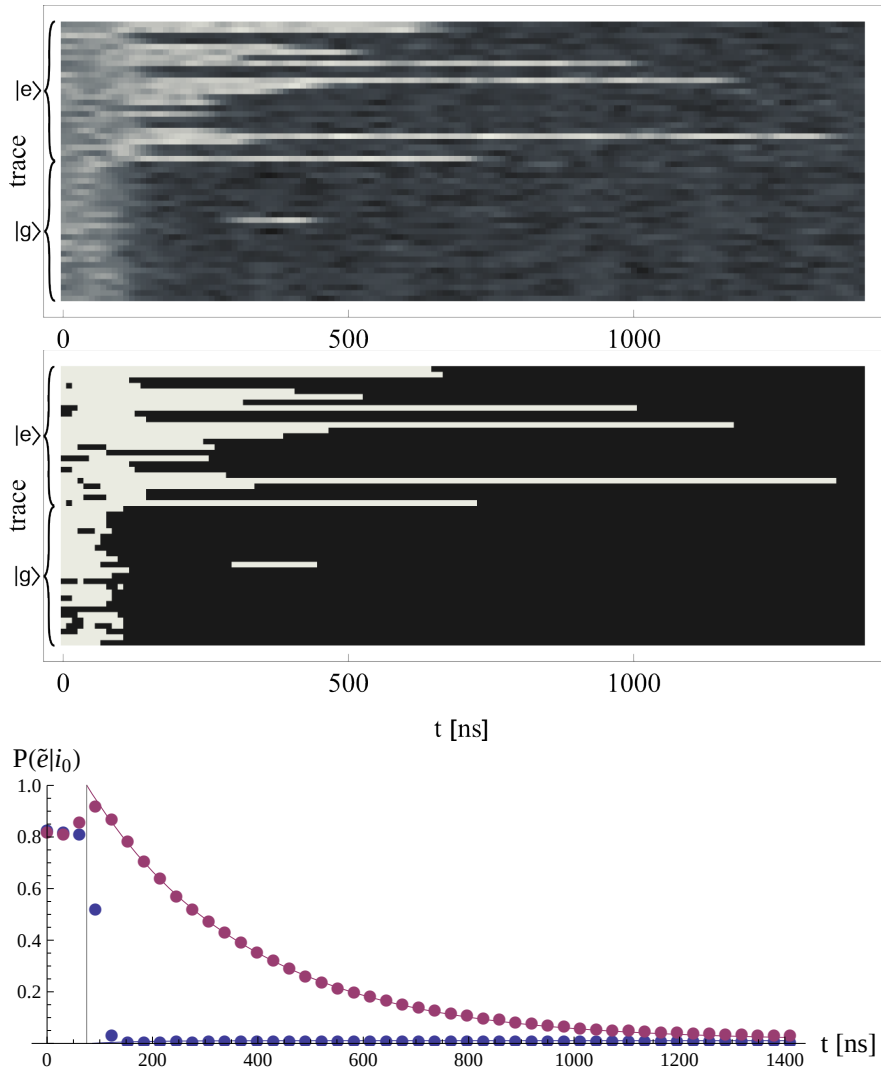

Figure 4.8: (top) Typical signal traces from measurements using a step pulse and the numerically optimized 3-point digital filter f_d. In the lower (upper) half of the traces the qubit was initially prepared in $|g\rangle$ ($|e\rangle$). (middle) The same traces with white indicating that the qubit exceeds the discrimination threshold between ground and excited state. (bottom) Mean populations calculated from the above and 10,000 other similar binary traces. The solid lines are exponential fits to the data points after $t = 200\,\text{ns}$ with time constants of $T_1' = 303 \pm 2\,\text{ns}$ and $T_{g.s.} = 84 \pm 30\,\text{ns}$ and a steady state offset of $(1.05 \pm 0.02) \times 10^{-2}$. The gray vertical line indicates the time at which the extrapolated qubit population reaches one.

measurement results are truly due to the qubit state and not due to dynamics of the resonator or the paramp.

Figure 4.9 shows some examples and mean populations for a set of traces where measurement pulses of 50 ns were applied every 250 ns. The result for $T_1 = 298 \pm 4$ ns agrees well with the result from the continuous measurements but the steady state ground state populations of $P(\tilde{e}_{t\to\infty}) \approx 0.04$ is four times higher than that observed in the continuous case. Based only on the ensemble averaged first order moments $P(\tilde{e}_t|i_0)$,[13] it is difficult to say whether the increase in $P(\tilde{e}_{t\to\infty})$ comes from truly increased population of the $|e\rangle$ state, denoted $P(e_t)$, or from misidentified $|g\rangle$ states, denoted $P(\tilde{e}_t|g_t)$. One can, however, inspect higher order moments involving more than one measurement result in order to separate these contributions. These moments describe the ensemble averaged correlations between measurements at different times, rather than just the independently averaged results.

Specifically, one can compute the second order moment

$$P(\tilde{e}_{t+\Delta t}|\tilde{e}_t) \equiv \frac{\langle \sigma_z(t)\sigma_z(t+\Delta t)\rangle}{\langle \sigma_z(t)\rangle}$$

where $\sigma_z(t) \in \{0, 1\}$ is the measurement result from measurement number t. This correlator is normalized in such a way that it describes the probability of measuring $\tilde{e}$ in measurement number $t + \Delta t$ given that $\tilde{e}$ was measured in measurement number t.[14] For $\Delta t \to \infty$, $\tilde{e}_t$ and $\tilde{e}_{t+\Delta t}$ are expected to become uncorrelated and therefore the unconditioned steady state value of $P(\tilde{e}_{t+\Delta t}|\tilde{e}_t)$ can be assumed to approach $P(\tilde{e}_{t\to\infty}) \equiv \langle \sigma_z(t \to \infty)\rangle \approx 0.04$.

The measured values of $P(\tilde{e}_{t+\Delta t}|\tilde{e}_t)$ shown in Figure 4.10 immediately show that a significant proportion of the excited state detections $\tilde{e}_t$ in steady state, i.e. when $P(e_t) \ll 1$, come from misidentified ground states. The lack of correlation between these false positives is particularly clearly manifested in the significant reduction in $P(\tilde{e}_2|\tilde{e}_1)$ for $i_0 = g_0$ compared to that for $i_0 = e_0$. This means that detecting $\tilde{e}$ in a single measurement in a situation where the unconditioned probability $P(g_t)$ is biased towards the ground state does not necessarily imply a high probability of the qubit actually being in the excited state. This illustrates the briefly mentioned fact that optimizing the measurement for F may not be optimal if $P(g_t) \neq 0.5$.

Quantitatively the data can be analyzed using a model where the measurements are point-like in time and the qubits are either truly excited with probability $P(e_t)$ and decay with probability $1 - e^{-T_{meas}/T_1}$ between measurements, or are truly in the ground state and cause uncorrelated false positives with probability $P(\tilde{e}_{t+\Delta t}|g_{t+\Delta t})$. Then for $\Delta t \geq 1$,

[13] Here the convention is to label the initially prepared qubit state as $i_0 \in (g_0, e_0)$ and refer to the first measurement as $t = 1$.

[14] This definition treats $\tilde{e}$ and $\tilde{g}$ very differently. This is intentional because the correlation between $\tilde{g}$ measurements is expected to be very strong and hence makes the $\tilde{e}$ correlations negligible in a symmetrically defined correlator.

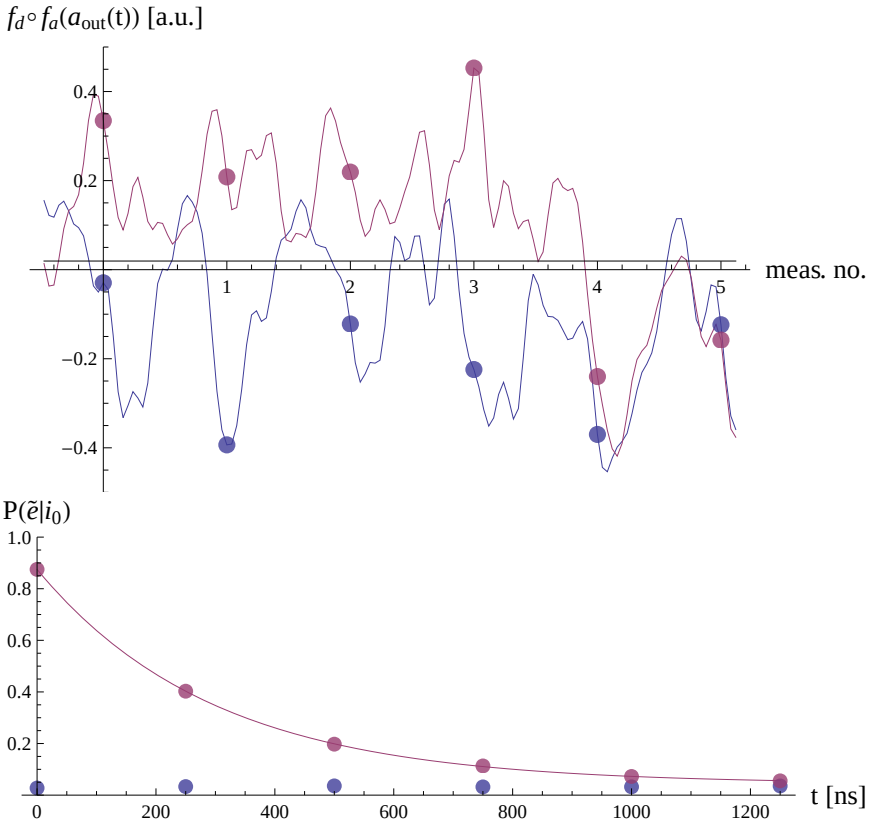

Figure 4.9: (top) Example signal traces from repeated pulsed measurements after preparing $|g\rangle$ (blue) or $|e\rangle$ (red). The solid black line is the threshold used to deduce whether the qubit is excited at the optimal measurement time for each measurement, i.e. at integer multiples of 250 ns. (bottom) The average qubit population after preparing the qubit in the ground (blue) or excited state (red). The solid red line is an exponential fit with $T_1 = 298 \pm 4$ ns and steady state offset of $(4.4 \pm 0.4) \times 10^{-2}$. 12,800 traces were used for the averages.

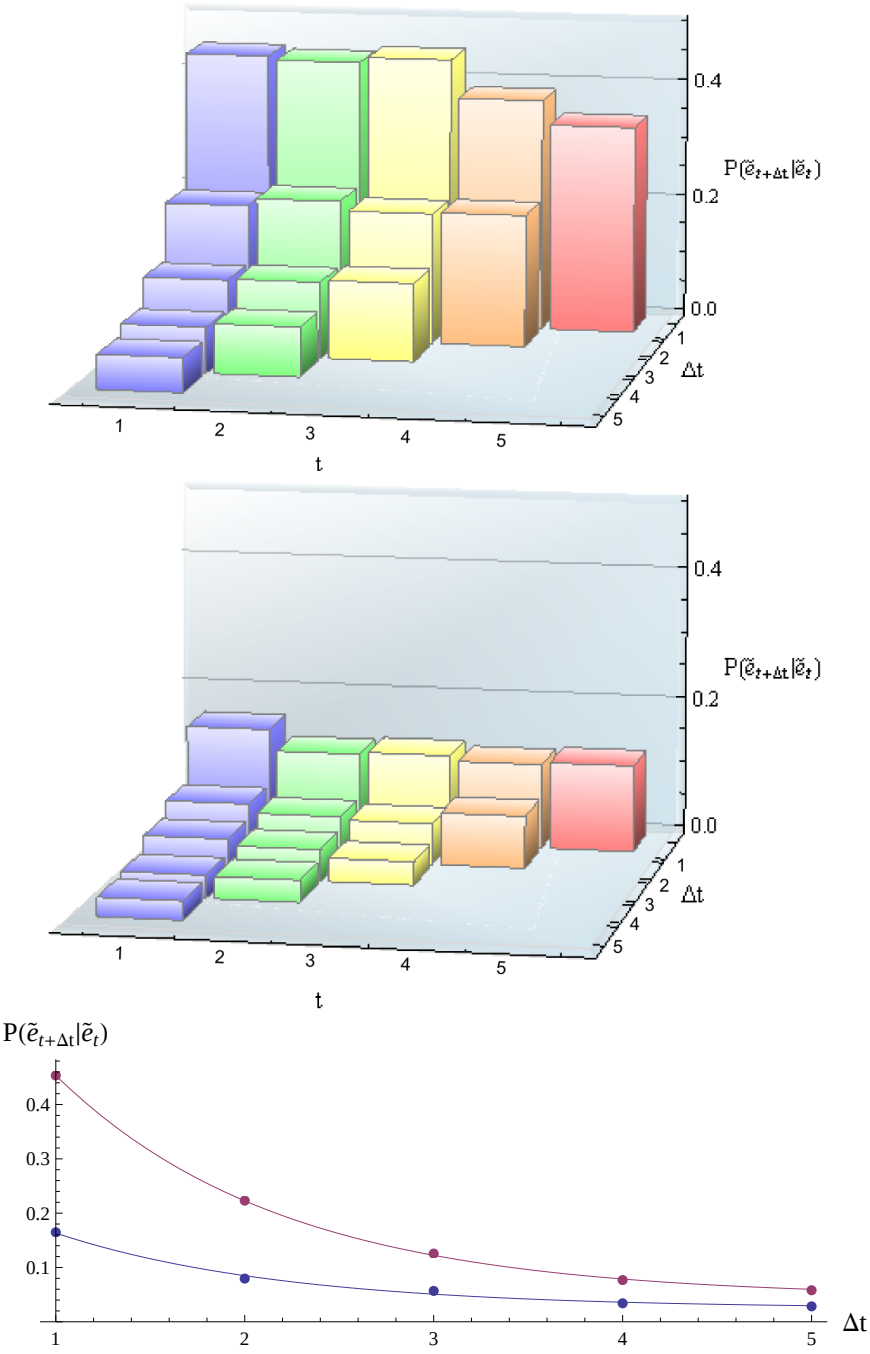

Figure 4.10: $P(\tilde{e}_{t+\Delta t}|\tilde{e}_t)$ after preparing $|e\rangle$ (top) or $|g\rangle$ (middle). The solid lines on the bottom are fits of measured $P(\tilde{e}_{1+\Delta t}|\tilde{e}_1)$ (blue bars) to Eq. 4.7 for the qubit initially in the ground (blue) or excited (red) state. The fit parameters $P(e_1|\tilde{e}_1)\,F_1$ and $P(\tilde{e}_1|g_1)$ were determined to be, respectively, 0.32 ± 0.05 (0.94 ± 0.03) and 0.025 ± 0.010 (0.045 ± 0.005) for the blue (red) curve, while $T_{meas} = 250\,\text{ns}$ and $T_1 = 300\,\text{ns}$ were fixed.

$$\begin{aligned}
P\left(\tilde{e}_{t+\Delta t}|\tilde{e}_t\right) =& P\left(e_{t+\Delta t}|\tilde{e}_t\right) P\left(\tilde{e}_{t+\Delta t}|e_{t+\Delta t}\right) \\
&+ P\left(g_{t+\Delta t}|\tilde{e}_t\right) P\left(\tilde{e}_{t+\Delta t}|g_{t+\Delta t}\right) \\
\approx& P\left(e_t|\tilde{e}_t\right) e^{-\frac{\Delta t}{T_1/T_{meas}}} P\left(\tilde{e}_{t+\Delta t}|e_{t+\Delta t}\right) \\
&+ \left[P\left(g_t|\tilde{e}_t\right) + P\left(e_t|\tilde{e}_t\right)\left(1 - e^{-\frac{\Delta t}{T_1/T_{meas}}}\right)\right] P\left(\tilde{e}_{t+\Delta t}|g_{t+\Delta t}\right)
\end{aligned}$$

where $T_{meas} = 250\,\text{ns}$ is the repetition interval of the measurements and second order processes corresponding to spontaneous excitation after the first measurement have been ignored. This is reasonable when Δt is small compared to the lifetime $T_{1,g}$ of the ground state but note that it leads to the wrong $\Delta t \times T_{meas}/T_{1,g} \to \infty$ limit since $P\left(e_{t+\Delta t}\right) \to 0$ when re-excitation is ignored.[15] Nevertheless, for $1 \leq \Delta t \ll T_{1,g}/T_{meas}$ the above terms can be rearranged without any further approximations as:

$$P\left(\tilde{e}_{t+\Delta t}|\tilde{e}_t\right) = P\left(e_t|\tilde{e}_t\right) F_{t+\Delta t} e^{-\frac{\Delta t}{T_1/T_{meas}}} + P\left(\tilde{e}_{t+\Delta t}|g_{t+\Delta t}\right)$$

where $F_{t+\Delta t} \equiv 1 - P\left(\tilde{e}_{t+\Delta t}|g_{t+\Delta t}\right) - P\left(\tilde{g}_{t+\Delta t}|e_{t+\Delta t}\right)$ is the fidelity of the measurement at time $t + \Delta t$.[16] Furthermore, it is reasonable to assume that F_t and $P\left(\tilde{e}_t|g_t\right)$ are nearly time independent since they do not depend on $P\left(g_t\right)$ and otherwise the measurements should be identical. Hence, with the additional assumptions that F_t and $P\left(\tilde{e}_t|g_t\right)$ are constant on a time scale of Δt, the correlator finally becomes:

$$P\left(\tilde{e}_{t+\Delta t}|\tilde{e}_t\right) = P\left(e_t|\tilde{e}_t\right) F_t e^{-\frac{\Delta t}{T_1/T_{meas}}} + P\left(\tilde{e}_t|g_t\right). \tag{4.7}$$

The form of Eq. 4.7 is convenient because T_1/T_{meas} is a known parameter and hence the two remaining free parameters $P\left(\tilde{e}_t|g_t\right)$ and $P\left(e_t|\tilde{e}_t\right) F_t$ can be extracted reliably by fitting $P\left(\tilde{e}_{t+\Delta t}|\tilde{e}_t\right)$ to the experimentally measured points.

As seen from the fit results in Figure 4.10, $P\left(e_t|\tilde{e}_t\right) F_t$ varies from 0.94 to 0.32 between the non-steady state and steady state situations where $P\left(g_t\right) \ll 1$ and $P\left(e_t\right) \ll 1$, respectively. This trend is expected since $1 - P\left(e_t|\tilde{e}_t\right) F_t$ corresponds approximately to the proportion of $\tilde{e}_t$ detections coming from false positives. The high proportion of false positives in steady state implies that the true steady state excited state population $P\left(e_t \to \infty\right) \sim 0.32 \times P\left(\tilde{e}_t \to \infty\right) \approx 0.014$ does not dramatically differ from the continuous measurement case where $P\left(\tilde{e}_{t\to\infty}\right)$ was 0.01 and could be assumed to be close to $P\left(e_{t\to\infty}\right)$ due to high SNR.

On the other hand, the values of the second fit parameter $P\left(\tilde{e}_t|g_t\right)$ show an unexpected change from 0.045 to 0.025 between the steady and non-steady state situations. The values themselves are of the expected order of magnitude but,

[15] $P\left(e_{t\to\infty}\right) \lesssim P\left(\tilde{e}_{t\to\infty}\right) = 0.04$ implies that $T_1/T_{1,g} \lesssim 0.04 \to T_{1,g} \gtrsim 7.5\,\mu\text{s}$.

[16] Note that $F_1 = 1 - P\left(\tilde{e}_1|g_1\right) - P\left(\tilde{g}_1|e_1\right)$ is not the same as the fidelity F defined for single measurements, which in this notation would be $F = 1 - P\left(\tilde{e}_1|g_0\right) - P\left(\tilde{g}_1|e_0\right)$ and takes into account qubit decay before the first measurement.

since $P(\tilde{e}_t|g_t)$ is the false positive rate conditioned on the qubit being in the ground state, there is no reason why it should depend on the qubit preparation. A more likely explanation is that the approximation $T_{1,g} \gg \Delta t \times T_{meas}$ is not completely accurate and the $P(g_t)$ dependent effects of spontaneous excitation show up at large values of Δt. Finally, note that small changes in this offset parameter do not dramatically affect the best fit value of the much larger other fit parameter $P(e_t|\tilde{e}_t) F_t$. Therefore the conclusion of the previous paragraph remains valid regardless of uncertainty in the precise value of $P(\tilde{e}_t|g_t)$.

5 Outlook

In this thesis fast dispersive single-shot readout with a fidelity of 0.85 was demonstrated. The time required to reliably determine the qubit state was only about 50 ns and, at least for the measured qubit with a short T_1, the 0.15 infidelity is dominated by spontaneous decay of the qubit rather than measurement induced transitions. In other words, no unexpected lower bound for the infidelity using the dispersive readout was found in these experiments. Therefore, based on these experiments there is no reason to expect that the infidelity should not keep decreasing linearly with $1/T_1$ in the future [Gambetta07]. It should also be possible to further reduce the measurement time and hence improve the fidelity significantly by taking advantage of the two-step pulse sequence more aggressively. As explained in Section 4.4, this improves the crucial initial scaling of the signal from quadratic to linear in time.

Besides improving fidelity, it would be interesting to study the QND nature of the pulsed measurement in a more systematic way. A generalization of the scheme of repeated measurements described in Section 4.6.1 would involve applying qubit operations between the measurements. Using such measurements one could determine not just how the qubit state in the measurement basis is correlated between measurements but also whether some correlations in other basis survive. An ideal projective measurement would completely remove these correlations regardless of imperfections in the detection chain after the resonator output but this may not be the case for very short pulses. In fact, one can argue that, if the vacuum noise is a significant source of infidelity, the measurement will not be perfectly projective since it is not possible to distinguish the measurement outcomes perfectly, even in principle.

Another possible way to completely describe the readout is to find the so called Kraus operators $\{\xi_i\}$ that describe the measurement as an operation on a larger space of density matrices that includes the qubit state as well as a classical variable $|z\rangle \in \{|\tilde{g}\rangle, |\tilde{e}\rangle\}$ that stores the measurement result. Since $|z\rangle$ is classical by definition, the Kraus operators are constrained to always project

the density matrix to the subspace $|\tilde{g}\rangle\langle\tilde{g}|\otimes\mathcal{S}(\mathcal{H}_q)\cup|\tilde{e}\rangle\langle\tilde{e}|\otimes\mathcal{S}(\mathcal{H}_q)$, where $\mathcal{S}(\mathcal{H}_q)$ is the space of reduced qubit density matrices. Furthermore, the measurement result should be independent of the initial value of the classical variable $|z\rangle$ and $|z\rangle$ should not be initially entangled with the qubit state. Therefore the measurement operation $\lambda(\rho_q)$ can be written as

$$\begin{aligned}\lambda(\rho_q) &= \sum_i \xi_i \left(|z\rangle\langle z| \otimes \rho_q\right) \xi_i^\dagger \\ &= |\tilde{g}\rangle\langle\tilde{g}| \otimes \sum_i G_i \rho_q G_i^\dagger \\ &\quad + |\tilde{e}\rangle\langle\tilde{e}| \otimes \sum_i E_i \rho_q E_i^\dagger\end{aligned}$$

where $\{G_i\}$ or $\{E_i\}$ are applied to the qubit state depending on the measurement outcome. A perfect QND measurement corresponds to $\{G_i\} = \{|g\rangle\langle g|\}$ and $\{E_i\} = \{|e\rangle\langle e|\}$. These can be rewritten in the Pauli basis as $(\sigma_0 \pm \sigma_z)/2$. In notation resembling usual process matrices:

$$\begin{aligned}\lambda(\rho_q) &= |\tilde{g}\rangle\langle\tilde{g}| \otimes \sum_{i,j} \chi_{G,ij} \sigma_i \rho_q \sigma_j^\dagger \\ &\quad + |\tilde{e}\rangle\langle\tilde{e}| \otimes \sum_{i,j} \chi_{E,ij} \sigma_i \rho_q \sigma_j^\dagger\end{aligned}$$

$$\text{where ideally } \chi_{G/E} = \frac{1}{2}\begin{pmatrix} 1 & & & \pm 1 \\ & 0 & & \\ & & 0 & \\ \pm 1 & & & 0 \end{pmatrix}.$$

However, χ_G and χ_E differ from normal process matrices in that only the sum of their traces is one.

Since both $\{\{G_i\}, \{E_i\}\}$ and $\chi_{G/E}$ describe the complete evolution of the reduced qubit density matrix conditioned on the measurement outcome and since no assumptions about the implementation of the binary measurement were made, knowing $\chi_{G/E}$ would completely characterize the measurement. In particular they could be directly applied to any input state of interest to calculate how much coherence from ρ_q remains in $\lambda(\rho_q)$ in different basis.

In practice, χ_G and χ_E could be measured the same way as in normal process tomography where qubit rotations are performed before and after the measurement to prepare and measure the qubit in different basis. The only difference would be that as an additional preprocessing step each trace would need to be sorted according to the measurement result $|z\rangle$ and later used only for computing the corresponding χ matrix. In addition the two matrices would at the end need to be normalized according to the ratio of $\tilde{e}/\tilde{g}$ counts so that their traces sum to one. Since the matrices $\chi_{G/E}$ are small, it should be feasible to simply store complete traces from each measurement and do all processing offline on a regular computer.

Finally, with modest improvements of the qubit coherence time it would become reasonable to attempt proof-of-principle quantum feedback experiments such as teleportation, where operations are conditioned on the outcome of a measurement. This could be especially fruitful at the Quantum Devices Lab at ETH Zürich since related work on teleportation has been already performed [Baur12]. Furthermore, the existing knowledge on FPGA programming should be helpful in incorporating real-time digital signal processing in the feedback loop.

Bibliography

[Baur12] M. Baur, A. Fedorov, L. Steffen, S. Filipp, M. P. da Silva, and A. Wallraff. "Benchmarking a quantum teleportation protocol in superconducting circuits using tomography and an entanglement witness." *Phys. Rev. Lett.*, **108**, 040502 (2012).

[Bianchetti09] R. Bianchetti, S. Filipp, M. Baur, J. M. Fink, M. Göppl, P. J. Leek, L. Steffen, A. Blais, and A. Wallraff. "Dynamics of dispersive single-qubit readout in circuit quantum electrodynamics." *Phys. Rev. A*, **80**, 043840 (2009).

[Blais04] 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**, 062320 (2004).

[Boissonneault09] M. Boissonneault, J. M. Gambetta, and A. Blais. "Dispersive regime of circuit qed: Photon-dependent qubit dephasing and relaxation rates." *Phys. Rev. A*, **79**, 013819 (2009).

[Boissonneault10] M. Boissonneault, J. M. Gambetta, and A. Blais. "Improved superconducting qubit readout by qubit-induced nonlinearities." *Phys. Rev. Lett.*, **105**, 100504 (2010).

[Chiorescu03] I. Chiorescu, Y. Nakamura, C. J. P. M. Harmans, and J. E. Mooij. "Coherent quantum dynamics of a superconducting flux qubit." *Science*, **299**, 1869–1871 (2003).

[Clerk10] A. A. Clerk, M. H. Devoret, S. M. Girvin, F. Marquardt, and R. J. Schoelkopf. "Introduction to quantum noise,

measurement, and amplification." *Rev. Mod. Phys.*, **82**, 1155–1208 (2010).

[Cohen-Tannoudji98] C. Cohen-Tannoudji, J. Dupont-Roc, and G. Grynberg. *Atom-Photon Interactions: Basic Processes and Applications.* Wiley Science Paperback Series (1998).

[Eichler11] C. Eichler, D. Bozyigit, C. Lang, M. Baur, L. Steffen, J. M. Fink, S. Filipp, and A. Wallraff. "Observation of two-mode squeezing in the microwave frequency domain." *Phys. Rev. Lett.*, **107**, 113601 (2011).

[Epanechnikov69] V. A. Epanechnikov. "Non-parametric estimation of a multivariate probability density." *Theory of Probability and its Applications*, **14**, 153–158 (1969).

[Fowler09] A. G. Fowler, A. M. Stephens, and P. Groszkowski. "High-threshold universal quantum computation on the surface code." *Phys. Rev. A*, **80**, 052312 (2009).

[Gambetta07] 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**, 012325 (2007).

[Gardiner85] C. W. Gardiner and M. J. Collett. "Input and output in damped quantum systems: Quantum stochastic differential equations and the master equation." *Phys. Rev. A*, **31**, 3761–3774 (1985).

[Girvin09] S. M. Girvin, M. H. Devoret, and R. J. Schoelkopf. "Circuit qed and engineering charge-based superconducting qubits." *Phys. Scr.*, **2009**, 014012 (2009).

[Houck09] A. Houck, J. Koch, M. Devoret, S. Girvin, and R. Schoelkopf. "Life after charge noise: recent results with transmon qubits." *Quantum Inf. Process.*, **8**, 105–115 (2009).

[Koch07] J. Koch, T. M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. 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 (2007).

[Mallet09] F. Mallet, F. R. Ong, A. Palacios-Laloy, F. Nguyen, P. Bertet, D. Vion, and D. Esteve. "Single-shot qubit readout in circuit quantum electrodynamics." *Nat. Phys.*, **5**, 791–795 (2009).

[Martinis02] J. M. Martinis, S. Nam, J. Aumentado, and C. Urbina. "Rabi oscillations in a large Josephson-junction qubit." *Phys. Rev. Lett.*, **89**, 117901 (2002).

[Nielsen00] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press (2000).

[Paik11] H. Paik, D. I. Schuster, L. S. Bishop, G. Kirchmair, G. Catelani, A. P. Sears, B. R. Johnson, M. J. Reagor, L. Frunzio, L. I. Glazman, S. M. Girvin, M. H. Devoret, and R. J. Schoelkopf. "Observation of high coherence in josephson junction qubits measured in a three-dimensional circuit qed architecture." *Phys. Rev. Lett.*, **107**, 240501 (2011).

[Schuster07] D. I. Schuster. *Circuit Quantum Electrodynamics.* Ph.D. thesis, Yale University (2007).

[Scully97] M. O. Scully and M. S. Zubairy. *Quantum Optics.* Cambridge University Press (1997).

[Siddiqi04] I. Siddiqi, R. Vijay, F. Pierre, C. M. Wilson, M. Metcalfe, C. Rigetti, L. Frunzio, and M. H. Devoret. "Rf-driven josephson bifurcation amplifier for quantum measurement." *Phys. Rev. Lett.*, **93**, 207002–4 (2004).

[Vijay11] R. Vijay, D. H. Slichter, and I. Siddiqi. "Observation of quantum jumps in a superconducting artificial atom." *Phys. Rev. Lett.*, **106**, 110502 (2011).

[Wallraff04] A. Wallraff, D. I. Schuster, A. Blais, L. Frunzio, R. S. Huang, J. Majer, S. Kumar, S. M. Girvin, and R. J. Schoelkopf. "Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics." *Nature*, **431**, 162–167 (2004).

[Wallraff05] A. Wallraff, D. I. Schuster, A. Blais, L. Frunzio, J. Majer, S. M. Girvin, and R. J. Schoelkopf. "Approaching unit visibility for control of a superconducting qubit with dispersive readout." *Phys. Rev. Lett.*, **95**, 060501 (2005).

[Walls94] D. Walls and G. Milburn. *Quantum Optics.* Springer Verlag, Berlin (1994).

[Walther06] H. Walther, B. T. H. Varcoe, B.-G. Englert, and T. Becker. "Cavity quantum electrodynamics." *Reports on Progress in Physics*, **69**, 1325–1382 (2006).

[Yurke89] B. Yurke, L. R. Corruccini, P. G. Kaminsky, L. W. Rupp, A. D. Smith, A. H. Silver, R. W. Simon, and E. A. Whittaker. "Observation of parametric amplification and deamplification in a josephson parametric amplifier." *Phys. Rev. A*, **39**, 2519–2533 (1989).

[Yurke96] B. Yurke, M. L. Roukes, R. Movshovich, and A. N. Pargellis. "A low-noise series-array josephson junction parametric amplifier." *Appl. Phys. Lett.*, **69**, 3078–3080 (1996).

[Yurke06] B. Yurke and E. Buks. "Performance of cavity-parametric amplifiers, employing kerr nonlinearites, in the presence of two-photon loss." *J. Lightwave Technol.*, **24**, 5054–5066 (2006).

A

Paramp Calibration

Calibrating the paramp pump power is in practice non-trivial because of the highly non-linear response of the paramp near the bifurcation point. Furthermore, in order to prevent the strong pump signals from entering the input and output ports of the paramp it is necessary to apply a second signal, called the displacement signal, that cancels the pump tone reflected from the paramp as indicated in Figure 3.1. Unfortunately, this cancellation and the bias point of the paramp are not diagonal in the pump and displacement tone basis because of the combined effect of non-zero cross-coupling in the directional coupler, large magnitudes of the pump and displacement signals, and the non-linearity of the paramp near the bifurcation point. In other words, it is not possible to find the optimal pump strength first and then calibrate the displacement signal so that the steady state power reaching the HEMT is close to zero. Instead a more complicated search algorithm that adjust both parameters is required for efficiently reaching the calibration that maximizes cancellation but still biases the paramp as desired.

Another practical problem is simply that for large cancellation, say 40 dB in power, it is necessary to control the amplitude of the pump and displacement signals down to 1 % on a linear scale. On the exponential decibel scale this implies control of the attenuation down to a level of 0.09 dB. Similarly, for 40 dB cancellation the resolution in the relative phase of the pump and displacement tones must be 0.01 rad. Furthermore, it would be highly desirable that the calibration is stable over at least several hours so that the calibration algorithm does not have to be constantly re-executed. For this reason a single high-quality microwave source was used from which the pump and displacement signals were split with a 3-dB splitter. The relative phase and amplitude of the displacement signal were then adjusted using a "displacer" that consists simply of a mechanical delay line and a mechanical attenuator controlled by two stepper motors as shown in Figure A.1. This mechanical approach also has the advantage that it remains calibrated regardless of crashes or other problems in measurement

software. On the other hand, this approach has the practical disadvantage that there is always some degree of hysteresis in the displacement as a function of the stepper motor positions due to play in the mechanical components.

The complete process of calibrating the paramp can be divided into three parts after fixing the desired signal frequency ω_s:

1. Record gain at ω_s and $\omega_s + 2\pi\Delta_b$, where Δ_b is of the order of the desired bandwidth, while sweeping pump power P_p and the B-field that controls the bare resonance frequency ω_0 of the paramp.

2. Choose a bias point (B, P_p) with desirable gain and bandwidth.

3. Minimize some objective function $h(P_p, A_d, \theta_d)$ that is maximal at the desired paramp bias point and when the cancellation is perfect. Here A_d and θ_d are the attenuation and phase shift applied to the displacement tone, respectively.

Currently all of these steps can be performed fully automatically once the sweep range for step number one has been set. However, in practice step number two is often performed manually because the desirable gain and bandwidth depend on the application.

The most challenging part of the process is choosing an appropriate objective function h and a robust but reasonably fast algorithm for finding the global minimum of h even in the presence of hysteresis, noise, and slow drifts. Currently the objective function is defined as

$$h(P_p, P_d, \theta_d) = -\tilde{G}\log(|G - G_0|) + \tilde{C}\log(P_{p,out}/P_s)$$

where G_0 is the desired gain, $P_{p,out}$ is the measured output power at the pump frequency, P_s is an arbitrary small power used to measure the gain G, and $\tilde{G}, \tilde{C} \in \mathbb{C}$ are user-defined weights. The bandwidth does not appear here since it is assumed to be related to the gain by the gain-bandwidth product.

The algorithm used to minimize h is inspired by how a human would do it, which is not extremely fast but very robust. It consists of first choosing an initial value for the acceptable objective function value $h_{th,outer}$ and then performing the steps described in the following pseudo-code:

```
function optimize(thOuter) {
  h={measureH()};
  i=0;

  while(h[-1] > thOuter) {
    i++;

    param=chooseParamToAdjust();
    delta=chooseDelta(param,i);

    j=0;
    thInner=h[-1];

    while(j<3 || h[-1] > thInner || (h[-1]<h[-2]<h[-3])) {
      j++;
      adjustParam(adjust,delta);
      appendTo(h,measureH());
      if(h[-1] > h[-2]) delta=-delta;
      thInner=increaseThreshold(thInner,h[-j;;-1]);
      thOuter=increaseThreshold(thOuter,h);
    }
  }
}

function measureH() {
  /* Returns the current value of the objective function.   */
}

function chooseParamToAdjust() {
  /* Returns "power" if either "phase" or "attenuation"     */
  /* were last adjusted. Otherwise, return the parameter    */
  /* that has been least recently adjusted.                 */
}

function adjustParam(param, delta) {
  /* Adjusts the pump power, attenuation or phase by delta. */
}

chooseDelta(param, i) {
  /* Returns the magnitude of delta to use depending on how */
  /* long the algorithm has been running.                   */
}

increaseThreshold(current,h) {
  return current+(max(h)-current)/T;
  /* where T is some userdefined time-constant, possibly    */
```

```
  /* different for the inner and outer thresholds.          */
}
```

The approach described by the pseudo-code is robust in the sense that it gets closer to the optimum point even in the presence of noise and hysteresis because it compares the current value to the value at the beginning of the inner loop. On the other hand, it is also robust in the sense that it does not get stuck in an inner loop if the initial value was a fluke because it increases the acceptable threshold slowly depending on how far off the measured values are from the threshold. The additional condition in the inner while loop ensures that the algorithm keeps adjusting the same parameter in the same direction as long as there is a consistent decrease in the objective function. Also note that in practice the initial value for $h_{th,outer}$ can be chosen by initially running the above algorithm without terminating the outer loop for some minimum number of iterations. Alternatively, a reasonable value may be known from previous rounds of optimization.

Finally, in the future it might be useful to also perform some measurements of the dynamic range and the noise number in step number one in order to choose the (B, P_p) point more wisely in step number two. It would also be convenient to automate the measurement of gain curves, dynamic range, and noise number once the optimization algorithm has completed.

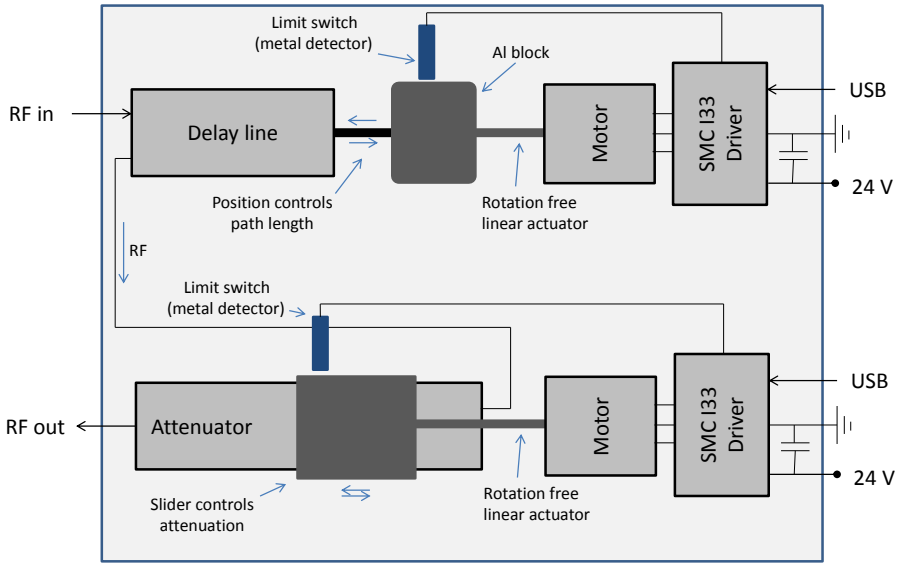

Figure A.1: Mechanical displacer. The delay line provides a phase shift of up to $\pi/3\,^{\mathrm{rad}}/_{\mathrm{GHz}}$ with a resolution of approximately $1.5\times10^{-4}\,^{\mathrm{rad}}/_{\mathrm{GHz}}$, which is much smaller than fluctuations in phase due to thermal expansion of a 1 meter cable under a 1 degree temperature change. The attenuator has a range of 10 dB and resolution of 0.01 dB.