Quantum technologies are becoming more and more introduced into the devices that we use in the daily life, such as quantum dot based television screens or quantum cryptographic channels for encoding financial transactions. Given the limits of the silicon computer chips, it will in the near future be necessary to integrate quantum elements into the computers in order to handle the exponentially growing amount of data that needs to be processed every day. While quantum systems can be controlled with excellent precision that comes with unexpected computational power, their monitoring and readout still seems to be a daunting task.

In my PhD, I introduced methods to tackle these problems based on experimentally relevant scenarios. In particular, I have used a newly developed theory of measurements to reinterpret the physical meaning of specific experiments in quantum optics. The same mathematical framework was then used to improve algorithms for parameter estimation of quantum systems, which is a highly important practical aspect for the characterisation of quantum technologies. Finally, I have improved conventional methods for parameter estimation protocols for quantum systems using novel machine learning techniques such as artificial neural networks.

AARHUS UNIVERSITY
DEPARTMENT OF PHYSICS AND ASTRONOMY

Eliska Greplova
Quantum Measurements: From Bayes Rule to Neural Networks
Dissertation

Quantum Measurements: From Bayes Rule to Neural Networks

Eliska Greplova

PhD Dissertation

Quantum Measurements: From Bayes Rule to Neural Networks

Eliska Greplova

PhD thesis
November 2017

Supervisor: Klaus Mølmer

Department of Physics and Astronomy
Aarhus University

English Summary

Quantum technologies are becoming more and more introduced into the devices that we use in the daily life, such as quantum dot based television screens or quantum cryptographic channels for encoding financial transactions. Given the limits of the silicon computer chips, it will in the near future be necessary to integrate quantum elements into the computers in order to handle the exponentially growing amount of data that needs to be processed every day. While quantum systems can be controlled with excellent precision that comes with unexpected computational power, their monitoring and readout still seems to be a daunting task.

In my PhD, I introduced methods to tackle these problems based on experimentally relevant scenarios. In particular, I have used a newly developed theory of measurements to reinterpret the physical meaning of specific experiments in quantum optics. The same mathematical framework was then used to improve algorithms for parameter estimation of quantum systems, which is a highly important practical aspect for the characterisation of quantum technologies. Finally, I have improved conventional methods for parameter estimation protocols for quantum systems using novel machine learning techniques such as artificial neural networks.

Dansk Resumé

Kvanteteknologier er ved at blive introduceret mere og mere i vores hverdag, for eksempel kvantedot-baseret fjernsyn eller kvantekryptografisk koding af finansielle transaktioner. Givet silicium computerchips nærtstående begrænsninger er det nødvendigt at integrere kvantelementer med vores computere for at vi kan håndtere den eksponentielt voksende mængde af data, der bliver produceret hver dag. Selvom kvantesystemer kan blive kontrolleret med excellent præcision der giver uventede computerkraft, så er målinger af kvantesystemer stadig en uoverkommenlig opgave.

I løbet af min PhD har jeg introduceret metoder til at håndtere disse problemer baseret på eksperimentielle relevante scenarier. I særdeleshed har jeg brugt en nyligt udviklet teori for kvantemålinger til at omfortolke den fysiske betydning af specifikke eksperimenter i kvanteoptik. De samme matematiske rammer blev herefter brugt til at forbedre algoritmer for parameterestimering af kvantesystemer, hvilket er et meget vigtigt praktisk aspekt for at karakterisere kvanteteknologier. Endelig har jeg forbedret traditionalle metoder for parameterestimering af kvantesystemer ved hjælp af nye maskinlæringsteknikker såsom kunstige neurale netværk.

Preface

This Thesis presents research completed in the course of my PhD at the *Deparment of Physics and Astronomy* at *Aarhus University*. The research (carried out between November 2014 and November 2017) was supervised by Klaus Mølmer and funded by Villum Fonden (within QUSCOPE, Villum Foundation Centre of Excellence).

Papers, [1–5], produced in the course of my PhD studies are listed in detail in the following section. This Thesis is mostly based on [3–5] and on Progress Report [6] written for the purpose of the PhD qualifying exam undertaken in 2016. Paper [1] fits well within the scope of this Thesis and is briefly summarized in Appendix C. Paper [2] is a work outside of the scope of my main PhD project that concerns the long standing quantum information problem of capacity evaluation of quantum channels. The illustration on the cover is an artists depiction of intersection of quantum devices with neural networks inspired by drawings of Ramon y Cahal, one of the founders of modern neuroscience.

First and foremost I would like to acknowledge Klaus Mølmer for being the best supervisor anyone could wish for in their PhD. I am deeply grateful to him for all his support, for uncountable hours of exciting physics discussions, for his astonishing physics intuition and encouragement to pursue my own ideas.

I would like to thank Andrew Briggs for welcoming me in his group at the University of Oxford, for many stimulating discussions and for being a valuable and insightful collaborator. I am grateful to Edward Laird for introducing me to the experimental realizations of semiconductor quantum devices and for all his hard work on our project. A big thank you to Natalia Ares for letting a theorist in her lab and for all the patient explanations of the experiment.

I would like to thank to all my collaborators for all their hard work on our projects, for offering fresh points of view and teaching me a lot about their expertise.

I am grateful to my colleagues from and beyond the 6th floor for creating a stimulating and fun research environment. In particular, I would like to thank to Jinglei Zhang, Jørgen Rørstad, Katérina Verteletsky, Felix Motzoi, Lukas Buchmann, Philip Blocher and Alexander Kiilerich for great discussions about physics and politics, but mainly for being great friends.

I would like to thank to Grete Flarup for her patience and administrative help.

Finally, I would like to thank to Christian Kraglund Andersen for all his support and for countless feedback when writing this Thesis.

List of Publications

[1] Qing Xu, Eliska Greplova, Brian Julsgaard, Klaus Mølmer. Correlation functions and conditioned quantum dynamics in photodetection theory. *Physica Scripta*, **90**, 12, (2015), *Invited Comment*

[2] Eliska Greplova, Geza Giedke. Degradability of Fermionic Gaussian Channels. *arXiv*:1604.01954 (2016), (*submitted to Physical Review Letters*)

[3] Eliska Greplova, Klaus Mølmer, and Christian Kraglund Andersen. Quantum teleportation with continuous measurements. *Phys. Rev. A*, **94**, 042334, (2016)

[4] Eliska Greplova, Edward A. Laird, G. Andrew D. Briggs, Klaus Mølmer. Conditioned spin and charge dynamics of a single electron quantum dot. *Phys. Rev. A*, **96**, 052104, (2017), *Editor's Suggestion*

[5] Eliska Greplova, Christian Kraglund Andersen, Klaus Mølmer. Parameter estimation of continuously monitored quantum systems using neural networks. *arXiv*:1711.05238 (2017), (*submitted to Physical Review Letters*)

Contents

Chapter 1

Introduction and outline

The main focus of this Thesis are various aspects of monitoring and parameter estimation of open quantum systems.

As a case study for monitoring of open quantum systems we use a single-electron quantum dot system, on which we illustrate an established description based on stochastic Schrödinger equation and stochastic master equation methods. We apply the so-called past quantum state analysis to unravel the electron and charge dynamics in the system. This type of analysis can in practice be applied directly on experimental currents and helps us greatly in removing uncertainty about the dynamics due to the experimental noise. Determining the precise timing of quantum jumps then proves to speed up and simplify the estimation of the parameters governing the dynamics. To that end we provide a new hybrid method of parameter estimation that allows for evaluation of coherent and incoherent parameters on the same footing. Our work presented here, carried out in collaboration with the Oxford experimental group of G. Andrew D. Briggs and Edward Laird, provides new tools for processing experimental currents obtained by measuring semiconductor devices and shifts the boundaries of precision for monitoring and parameter estimation.

When mastering the most common methods of quantum parameter estimation one may ask about the possible limitations of these approaches as parameter estimation becomes crucial in the particular context of processing of the experimentally measured currents. By construction, Bayesian parameter estimation and related methods are typically based on solving the master equation that captures the dynamics of the system, which means that whatever limitations there are on the possibility of solving (or even formulating) the master equation, these will immediately

translate into a large degree of complexity, or even an impossibility to estimate the relevant experimental parameters using Bayesian methods. One of the most important assumptions being made in order to derive the master equation is 'Markovianity' of the environment. In simple terms, this means that the signal at the time t only depends on the state of the system at that time and that $\rho(t + dt)$ only depends on $\rho(t)$ and the signal at t. Other issues that can cause a master equation not to represent the dynamics of the system faithfully are experimental considerations related to current detection such as dark counts and finite bandwidth. There has been number of theoretical proposals to expand the Markov theory to include these complications, but they typically tend to be very mathematically and computationally demanding.

In this Thesis we would like to propose one possible solution to all these types of issues that is inspired by recent developments in the field of artificial intelligence. Using the same example system we used above, a single electron quantum dot, we first show that we can reproduce the results obtained using master equations and Bayesian analysis by constructing a simple convolutional neural network and training it on sufficiently large sets of experimental currents. In addition to that, we show that an identical analysis works for experimental currents submitted to all kinds of experimental assumptions we described above, i.e. in the cases when Bayesian parameter estimation is impossible.

As a result this work should serve as a bridge between traditional open quantum system methods and applicability of artificial intelligence methods in this field.

This Thesis is organized as follows. In Chapter 2 we give a brief overview of open quantum system methods and a mathematical framework for describing the monitoring of open quantum system. We extend these standard techniques by an introduction to past quantum state method, that will allow us to get more precise idea about the underlying quantum state. We also offer a short introduction into Bayesian and Baum-Welch parameter estimation. Chapter 3 (based on [4]) reviews the direct application of these methods on a single-electron quantum dot. In particular we use a stochastic master equation approach to analyze the signal of the charge sensor used for observation of the dot. After identifying the time evolution we are in the position to estimate the parameters governing the time evolution of the system. To this end we introduce and apply a novel hybrid technique that allows for estimation of coherent parameters while a numerically efficient Baum-Welch method allows estimation of incoherent rate parameters. Given the questions about the limitations of master

equation based techniques for parameter estimation raised above, we take a small detour in a form of Chapter 4, which can be used as a short independent introduction into the language of machine learning tailored towards the needs of a physicist. With the knowledge acquired in Chapter 4, we are in the position to revisit a single-electron quantum dot from the perspective of neural networks. Chapter 5 (based on [5]) is dedicated to this new point of view, and in addition to that it offers a generalization for noise sources in the experimental current that can not be captured by Markovian theory. Finally, Chapter 6 (based on [3]) serves as an example of an impact of past quantum state based post-processing and neural network analyses in the field of quantum algorithms. Final thoughts and a discussion of possible extensions and future impact of this work are presented in Chapter 7.

Chapter 2

Monitoring of open quantum systems

The central notion of this Thesis is quantum measurements. Since the advent of quantum mechanics the concept of quantum measurement led to spirited debates about the consequences of the action of a measurement on a quantum state [7, 8]. The understanding of these consequences is largely dependent on one's chosen interpretation of what quantum state represents and is to-date an open problem [9, 10].

In the recent years, while we witnessed the rise of implementations of quantum devices, the need to start another kind of discussion related to measurements arose [11–18]. In particular, we need to be able to access the quantum information processed by these devices under the experimental restrictions and at the same time understand the impact our measurement has on the system in question. It is no longer only the question of when, how and whether at all the collapse of the wave-function happens we care about, but the pressing need to understand the issues of how a measurement modifies the state of a quantum system and how much information we can obtain through this measurement about real physical quantities governing the dynamics of the system in the realistic experimental conditions.

While the basic formalism of quantum mechanics exactly predicts what happens to a quantum state when it is measured, this formalism still needs to be adapted to cover the cases of non-ideal, noisy, and non-projective measurements. We are of course not going to depart from the predictions of quantum mechanics. We merely adapt them for purposes of realistic detection.

In this Thesis we take on this practical point of view and consider several examples of realistic quantum devices and the tools being used for their readout. Then we connect these realistic limitations to the mathematical formalism describing the underlying quantum dynamics that is only accessible through elaborate measurement methods. We discuss the methods of obtaining the maximum amount of information about a quantum system under experimental restrictions through post-processing of experimental currents, modeling the inner workings of the quantum device, and training neural networks to recognize relevant features in experimental data.

In this Chapter we give a short introduction into the standard methods of for description of monitored noisy quantum systems. This introduction overlaps with Chapter 1 of [6] and is mainly based on the mathematical tools described in [19] and [20].

In general, the state of a quantum mechanical system is described by a vector $|\Psi\rangle$ in a complex Hilbert space $\mathcal{H}$. $|\Psi\rangle$ is often referred to as a *pure state*. Should we have a statistical mixture of these, there is a straightforward generalization

$$\rho = \sum_{\alpha} p_\alpha |\Psi_\alpha\rangle\langle\Psi_\alpha|. \tag{2.1}$$

The object ρ is called a *density operator* or *density matrix* once written in a particular basis. Within this work we mostly use density matrices to describe the state of the system under consideration as in an experimental context we can hardly ever talk about pure states.

Quantum mechanics postulates unitary evolution of a closed system. In terms of ρ this is described by the *von Neumann* equation

$$\frac{d}{dt}\rho = -\frac{i}{\hbar}[H, \rho]. \tag{2.2}$$

A more realistic description of the time evolution, however, should take into account the possibility that the system in question can interact with an (possibly infinite) environment. This leads to the generalization of (2.2) called the *master equation*:

$$\frac{d\rho}{dt} = -\frac{i}{\hbar}[H, \rho] + \sum_{\mu} \mathcal{D}[\hat{c}_\mu]\rho =: \mathcal{L}\rho, \tag{2.3}$$

where $\mathcal{D}$ is a superoperator (this refers to the fact that it is an operator on the space of operators) and its action on ρ is defined as

$$\mathcal{D}[\hat{c}]\rho = \frac{1}{2}(2\hat{c}^\dagger \rho \hat{c} - \hat{c}^\dagger \hat{c}\rho - \rho\hat{c}^\dagger \hat{c}). \tag{2.4}$$

Here $\hat{c}$ is an operator that describes the degree of freedom of the quantum system interaction with the environment.

The master equation, however, does not take into account the possible observation of the system, which is, again, highly relevant for experiments. In particular, when performing an experiment we would like to monitor some aspect of it, which in general modifies the state of the system due to the back-action of the performed measurement. In other words, we would like to have the description of the state *conditioned upon the measurement results.*

If we were able to monitor each occurrence of the interaction, Eq. (2.4), with the environment, Eq. (2.3) would then be replaced by the combination of the so-called no-jump dynamics within the time increment dt

$$\frac{d\rho}{dt} = -\frac{i}{\hbar}[H,\rho] - \frac{1}{2}\sum_{\mu}(\hat{c}_{\mu}^{\dagger}\hat{c}_{\mu}\rho - \rho\hat{c}_{\mu}^{\dagger}\hat{c}_{\mu}), \tag{2.5}$$

and then followed by a quantum jump

$$\rho \to \frac{\hat{c}_{\mu}^{\dagger}\rho\hat{c}_{\mu}}{\mathrm{Tr}(\hat{c}_{\mu}^{\dagger}\rho\hat{c}_{\mu})} \tag{2.6}$$

with a (typically low) probability $dp = \mathrm{Tr}(c_{\mu}^{\dagger}\rho c_{\mu})dt$. After the jump, the density matrix, ρ, is renormalized and the evolution is continued in the next step, dt. The theory describing quantum evolution interrupted by quantum jumps is called stochastic quantum trajectory theory [21–23], and it offers a straightforward formalism of including monitoring into the master equation.

In this work we use this theory implicitly for the numerical simulations of the systems in question and on occasion we list the stochastic master equation explicitly. In general, all of these are derived following standard methods, an extensive review of which can be found in [19, 20]. All our novel contributions in this Thesis can be understood through the intuition of consecutive jumpless evolution followed by a quantum jump outlined above.

There is another, equivalent, point of view we can take on the time evolution of the monitored system, and that is through the measurement operator theory. In quantum mechanics, a measurement is described by the set of operators $\{M_i\}$, which fulfill the property $\sum_i M_i^{\dagger} M_i = \mathbb{1}$. These are called POVMs (positive operator-valued measure) and allow us to express the state of the system at the time $t + dt$ in terms of the state at the time

t as

$$\rho(t+dt) = \frac{M_i \rho(t) M_i^\dagger}{\mathrm{Tr}(M_i \rho(t) M_i^\dagger)}, \tag{2.7}$$

where M_i is the measurement performed during the time dt. In other words, we can find a POVM that reproduces Eqs. (2.6)-(2.5) (up to the first order in dt).

Let us consider the difference between discrete and continuous measurements on the example of Ref. [19]. Let us consider a measurement operator M_i, that has two outcomes $i = 0, 1$, such that

$$M_0^\dagger(dt) M_0(dt) + M_1^\dagger(dt) M_1(dt) = \mathbb{I}. \tag{2.8}$$

One example of such operators is

$$M_0(dt) = \mathbb{I} - (R/2 + iH)dt, \tag{2.9}$$

where R and H are Hermitian operators, and

$$M_1(dt) = \sqrt{dt}\hat{c}, \tag{2.10}$$

where $R = \hat{c}^\dagger \hat{c}$. In fact, we see that applying M_0 and M_1 on ρ according to Eq. 2.7 we get Eq. (2.5) and Eq. (2.6) respectively. The average dynamics of the system is then described by $\rho(t+dt) = \sum_i M_i \rho(t) M_i^\dagger$, which on average exactly reproduces Eq. (2.3).

Let us denote N(t) number of occurrences of outcome '1' until the time t. Let us make an assumption that measuring the outcome '1' will not affect measured system significantly, i.e. we have to detect many of these outcome to see the difference in the measured system. The typical experimental situation where these conditions arise naturally is shown in Fig. 2.1. $N_{+,-}$ denote the photon (or other particle) number measured in the respective arms of the interferometer. This measurement is referred to as homodyne detection and it a widely used method to detect weak signals without affecting the measured system significantly. The main idea of this approach is to mix the weak signal from the system (the direct measurement of which would most likely drastically affected the system) with the strong signal (in an optical setting this would mean mixing a few photon signal with a strong laser drive). The result of this is that most of the measured photons comes from the strong drive and not from system itself and when we measure for a sufficiently long time we can make conclusions about the system dynamics from this record. It has been

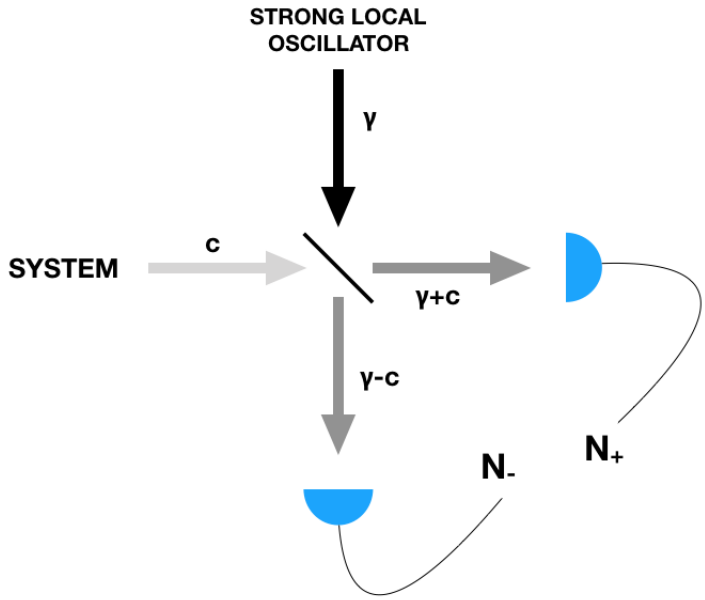

Figure 2.1: Scheme of the homodyne detection measurement: low signal from the system is mized with strong local oscillator on the 50:50 beam-splitter, the difference of the signals N_+ and N_- is detected.

shown in [24] that the operator action of such detection events on the measured system can be approximated by

$$\mathcal{O} \propto \left(\frac{\gamma}{\sqrt{2}}\right)^N (\mathcal{I} + \frac{1}{\gamma}(N_+ - N_-)\hat{c}). \tag{2.11}$$

This leads to the signal-to-noise ratio $\approx \frac{1}{\sqrt{T}}$, where T is the total measurement time. This means that we obtain a projective measurement in the long time limit (as long as $\hat{c}$ is hermitian and commutes with the Hamitonian, H). In other words, projective and continuous measurements are described by a completely equivalent manner within the realm of quantum mechanics. The weak measurement formalism turns out to be invaluable in many quantum optics experiments [25–29], because whenever we want to observe any kind of dynamical effect it is imperative not to destroy the evolution by the projective measurement.

The probability for a given measurement outcome i at time t is expressed by the *Born rule* (see for example [30]), which is in terms of POVM expressed as

$$p(i,t) = \frac{\mathrm{Tr}(M_i \rho(t) M_i^\dagger)}{\sum_i \mathrm{Tr}(M_i \rho(t) M_i^\dagger)}. \tag{2.12}$$

The way to think about this equation is that it is a probability of a measurement outcome conditioned on the measurement record preceding the measurement in question.

There is a new ingredient, the so-called past quantum state method introduced in [31], that expands on the notion of conditional probability. Eq. (2.12) tells us the probability of getting an output i at the time t depends on the time evolution and measurement back-action that state ρ went through before time t. Now, in the case of continuous monitoring, we are getting constant stream of measurements through the evolution of the state, so there is no reason to base our probability prediction exclusively on the record preceding the time t, when there is an equally informative record of what happens with the state after time t.

The intuitive way one can understand this is the following (see also [6]). Say we want to predict the probability of the outcome of the measurement of the operator M_i. Omitting the normalization we simply write $p(i) \propto \mathrm{Tr}(M_i \rho M_i^\dagger)$. Now we propagate the system further, which will be equivalent to applying another POVM, say M_j, according to the relation (2.7). This results in an expression proportional to $M_j M_i \rho M_i^\dagger M_j^\dagger$. Let us re-evaluate the probability for the outcome i after getting more measurement outcomes. Now we find

$$p(i) \propto \mathrm{Tr}(M_j M_i \rho M_i^\dagger M_j^\dagger) = \mathrm{Tr}(M_i \rho M_i^\dagger M_j^\dagger M_j) = \mathrm{Tr}(M_i \rho M_i^\dagger E), \tag{2.13}$$

where, in the first equality, we use the cyclic property of the trace and, in the second one, we simply identified $M_j^\dagger M_j =: E$. In other words, we improved our probability expression by conditioning the probability also on the future measurement record.

This logic gives rise to the generalized Born rule [31]

$$p(i,t) = \frac{\mathrm{Tr}(M_i \rho(t) M_i^\dagger E(t))}{\sum_i^N \mathrm{Tr}(M_i \rho(t) M_i^\dagger E(t))}, \tag{2.14}$$

where the probability $p(i,t)$ for a measurement outcome i conditioned on the measurement record that precedes *and* follows the measurement in question. The information accumulated from the signal obtained after time

t enters the expression through a quantum operation E. The expression Eq. (2.14) follows from the quantum theory of measurements and conditional probabilities, and it has been applied to experiments with Rydberg atoms in microwave cavities [32] and with superconducting qubits [33–35], where its predictions have been confirmed and used to identify quantitative properties of the system.

As the logic of Eq. (2.13) suggests, E is a solution of a master equation propagated backwards in time

$$\frac{E(t-dt)-E(t)}{dt} = i[H,E] + \frac{1}{2}\sum_{\mu}(2c_\mu E c_\mu^\dagger - c_\mu c_\mu^\dagger E - E c_\mu c_\mu^\dagger), \quad (2.15)$$

with the initial condition $E = \mathbb{I}$, and is subjected to POVM operators as $E(t-dt) = (M_i^\dagger E(t) M_i)/\mathrm{Tr}(M_i^\dagger E(t) M_i)$, where $E(t-dt)$ is conditioned on the outcome i of the measurement at the time t.

In Chapter 3 and Chapter 6 we will show how taking into account the whole duration of the measurement signal can help us to significantly improve the readout of quantum devices and maximize the fidelity of quantum algorithms respectively. Details of the past quantum state method are discussed in [1, 6, 31].

Finally, let us comment on parameter estimation methods based on the master equation formalism introduced above. The fact, that we are able to model the dynamics of the system by finding the state, ρ, as a solution of a differential equation at every time step, means that, at least within the approximations we made to derive the master equation, we have a very good idea about how the state of the system evolves in time. In addition to that, thanks to the generalized Born rule (2.14), we maximize our ability to predict the probability of any measurement outcome. We show in Chapter 3 that we can use this with advantage to determine the precise timing of quantum jumps in the observed system. It is not surprising that we can combine these two pieces of knowledge to determine the physical parameters governing the time evolution of the system under consideration.

Parameter estimation tools used in this work are based on Bayes rule [36]

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}, \quad (2.16)$$

where $P(\theta|D)$ is the probability density of parameter, θ, conditioned on the measured data, D. Bayes rule connects $P(\theta|D)$ with $P(D|\theta)$, the probability of observing data D conditioned on the parameter θ and the

so-called prior, $P(\theta)$ about the parameter θ. Finally, the probability of the data, $P(D) = \sum_\theta P(D|\theta)P(\theta)$, normalizes the expression. This means that if we are able to recover $P(D|\theta)$ from the master equation model, we will immediately be able to infer the physical parameters governing the dynamics as well. In practice, instead of finding the maximum of $P(D|\theta)$, the so-called likelihood $L(D|\theta)$ is being evaluated ($L(D|\theta)$ is $P(D|\theta)$ renormalized by a stricly positive function $P_0(D)$, which attains the maximum value for the same θ, i.e. $L(D|\theta) = P(D|\theta)/P_0(D)$). The likelihood, $L(D|\theta)$, is related to parameter probability density via

$$P(\theta|D) = \frac{L(D|\theta)P(\theta)}{\int d\theta L(D|\theta)P(\theta)}. \tag{2.17}$$

The reason for this adjustment is that at each time, t, the likehood, $L_t(D|\theta)$, is directly related to the unnormalized solution of stochastic master equation at the time t, $\tilde{\rho}$, describing the sytem (as we can see from Eq. 2.7). In particular,

$$L_t = \mathrm{Tr}(\tilde{\rho}_t). \tag{2.18}$$

This means we can evaluate the likehood $L_t = \mathrm{Tr}(\tilde{\rho}_t)$ simply by solving the master equation describing the system for all t. Then we just need to find θ for which the likelihood is maximized. In practice, this is done by parallel simulation of the master equation for the set of candidate values for the parameter θ based on educated guess. We illustrate this method on the practical application for the case of determining parameters of the spin dynamics in a quantum dot in Chapter 3. While Bayesian estimation is the to-date most universal and powerful method for determining unknown parameters, it has a caveat. Whenever θ is not just a single (or very low-number of) parameter(s), the parameter space increases exponentially with the number of parameters and the estimation becomes numerically intractable. It is therefore important to have other methods available in order to be able to decrease the size of the parameter space for Bayesian estimation.

One answer to this problem is the so-called Baum-Welch estimation [37, 38] that, although also based on Bayes rule, uses the recurrent update approach that does not contribute to the parameter space issue. This method is purely classical, because it is based on the conditional probabilities for transitions between particular states and in its original formulation has no space for coherent superpositions. In a quantum context, this method has been applied for dynamics driven exclusively by incoherent transitions [37]. Let us consider a hidden Markov process, i.e. the Markovian time evolution of a hidden variable X_t, the properties of which we are

trying to infer from the measurement signal D_t, where subscript t corresponds to the time step of the evolution. The incoherent parameter we are trying to estimate corresponds to the transition probabilities $P(X_{t+dt}|X_t)$. The detailed derivation can be found in [37] (see also [4]). The central notion of the method is to evaluate joint probability

$$C_t(i,j) = P(X_{t+dt} = j, X_t = i|D_1, ..., D_N). \tag{2.19}$$

This quantity allows us to re-estimate the transition probabilities $P(X_{t+dt} = j|X_t = i)$, i.e. what is the probability the system is in the state j at the time $t + dt$ provided it has been in the state i at the time t. In particular, the relation of $C_t(i,j)$ to the re-estimated the transition probabilities is

$$P(X_{+dt} = j|X = i) = \frac{\sum_t C_t(i,j)}{\sum_t(\sum_j C_t(i,j))}. \tag{2.20}$$

The main idea of this method is to use the knowledge we have about the transition probability from the measured data, D, to calculate the correlation (2.19), re-estimate the transition probability for a given time, t, and then move on to the $t + dt$ and repeat the process. In Chapter 3 we show how to apply this method in practice to determine the rate at which an electron tunnel on and off the quantum dot and explain how to extend it to the quantum cases containing the coherent dynamics as well. Thus, we put the tool introduced here into practical use to analyze experimental currents from a continuously monitored quantum system evolving under combination of coherent and incoherent dynamics. There are also limitations to these tools coming primarily from experimental noise. We address these in Chapter 5.

Chapter 3

Charge and spin dynamics of a single-electron quantum dot

This Chapter is a case study of the usefulness of the theoretical framework introduced in Chapter 2 for realistic experimental readout of quantum devices. We apply stochastic master equation tools to describe the monitoring of the charge and spin dynamics of a single-electron quantum dot. This Chapter is based on [4] and the matter discussed here also overlaps with research presented in [6]. In particular, Sections 3.1-3.5 and parts of this Introduction and Section 3.7 are taken over from [4] with only minor changes. The research presented in this Chapter is a result of our collaboration with University of Oxford.

The information about a quantum system from stochastic measurement signals can be used for the purpose of state reconstruction [39–42] and precision measurements of physical parameters [36, 37, 43–45]. In this work we achieve this by a a combination of Bayesian analysis and modified Baum-Welch re-estimation that extracts the quantum state and the physical parameters governing the system dynamics from the measurement data. We apply this to the single-electron occupation of a quantum dot. Because the occupation depends on quantum tunnelling which in turn depends on the spin state [46], we have a combination of incoherent and coherent dynamics. This requires a modification to the conventional Baum-Welch estimation scheme, with repeated application until all the parameters have converged.

We consider a new scheme for repeated and continuous monitoring of a spin qubit that tunnels on and off a quantum dot [46]. A static magnetic field splits the spin-up and spin-down state energies, such that only a spin-down electron may tunnel into the dot and prevent further charging of the dot by Coulomb blockade. A resonant drive causes oscillations between the spin-up and spin-down states. There is therefore a mixture of coherent spin dynamics and incoherent tunneling events, as the Rabi oscillations are interrupted when the excited spin-up electron tunnels out of the dot. A quantum point contact (QPC) which transmits an electron current that depends on the charge but is insensitive to the spin dynamics on the quantum dot is used to continuously monitor the electron tunneling dynamics [12, 47, 48]. Our theory will apply to the analysis of real experimental data, but in this work it will be illustrated on a simulated system dynamics, where the tunneling events in and out of the quantum dot occur governed by a stochastic master equation [19]. We note that even though our experiments are not sensitive to the electron spin state, the Rabi oscillatory dynamics will reveal itself through the distribution of time intervals spent on the dot, since the electron enters and leaves in different spin states. We thus aim to recover the quantum state and the physical parameters governing the electron spin and charge from the stochastic measurement signal of the charge sensor. We show how the time dependent dot occupation can be estimated and how this estimate is improved by incorporating information from the subsequent sensor record. Finally, we show how the spin precession becomes imprinted on the charge dynamics, and we present a method to estimate efficiently the parameters of the qubit Hamiltonian from the noisy charge measurement record.

This Chapter is organized as follows: In Sec. 3.1, we introduce the master equation and quantum trajectory description of the system subject to ideal probing of the charge state. In Sec. 3.2, we present the case of continuous probing of the charge state, and illustrate the discrepancy between the true and the estimated state of the system. In Sec. 3.3, we show how the past quantum state formalism employs the full time dependent signal and provides an estimate of the time dependent charge state in better agreement with the true evolution of the system. In Sec. 3.4, we propose a combination of Bayesian analysis and modified Baum-Welch re-estimation that extracts the physical parameters governing the system dynamics from the measurement data. In Sec. 3.5, we present consideration for experimental realization of our method and its applicability in realistic conditions. In Sec. 3.6 we show the power of our technique in the circumstances even noisier than in the experiment we are considering here.

Finally, in Sec. 3.7 we conclude this Chapter and provide an outlook for further use of our technique.

3.1 Coherent and incoherent processes, conventional and stochastic master equation

We consider a gate-defined quantum dot that can contain zero or one electron (Fig. 3.1(a)). The quantum dot is coupled to an electron reservoir such that electrons can tunnel on and off. The three states available are the state $|0\rangle$ with no electron charge on the quantum dot, and the singly charged spin-up and spin-down states $|\uparrow\rangle$ and $|\downarrow\rangle$. The dot is measured via a quantum point contact (QPC) acting as a charge sensor. The current I_{QPC} through this point contact is sensitive to the charge occupation; when the dot is empty, the average current is I_0, while, when the dot is occupied with one electron, the current is I_1.

We suppose that the device is placed in a magnetic field to introduce a Zeeman splitting of the two spin states, adjusted using gate voltages so that the spin-up and spin-down levels straddle the Fermi level of the reservoir (Fig. 3.1b). Electron tunneling is now spin-sensitive; if the quantum dot is charged with an electron in state $|\uparrow\rangle$, it will tunnel off the dot with rate $\gamma_\uparrow$; if the quantum dot is empty, a spin-down electron will tunnel onto the dot with rate $\gamma_\downarrow$. A coherent drive at the qubit resonance (Larmor) frequency induces coherent precession at the Rabi frequency Ω, and when the device evolves between the three states, the charge state sensitive I_{QPC} fluctuates around different characteristic values as shown in Fig. 3.1c.

The Hamiltonian describing the resonant drive between the up and down states reads ($\hbar = 1$)

$$H = \frac{\Omega}{2}(|\uparrow\rangle\langle\downarrow| + |\downarrow\rangle\langle\uparrow|), \tag{3.1}$$

with the Rabi frequency Ω.

Let us define the operators associated with the incoherent transfer of the electron between the dot and the electron states in the Fermi sea:

$$c_\downarrow = |0\rangle\langle\downarrow|, \quad c_\downarrow^\dagger = |\downarrow\rangle\langle 0|, \tag{3.2}$$

$$c_\uparrow = |0\rangle\langle\uparrow|, \quad c_\uparrow^\dagger = |\uparrow\rangle\langle 0|. \tag{3.3}$$

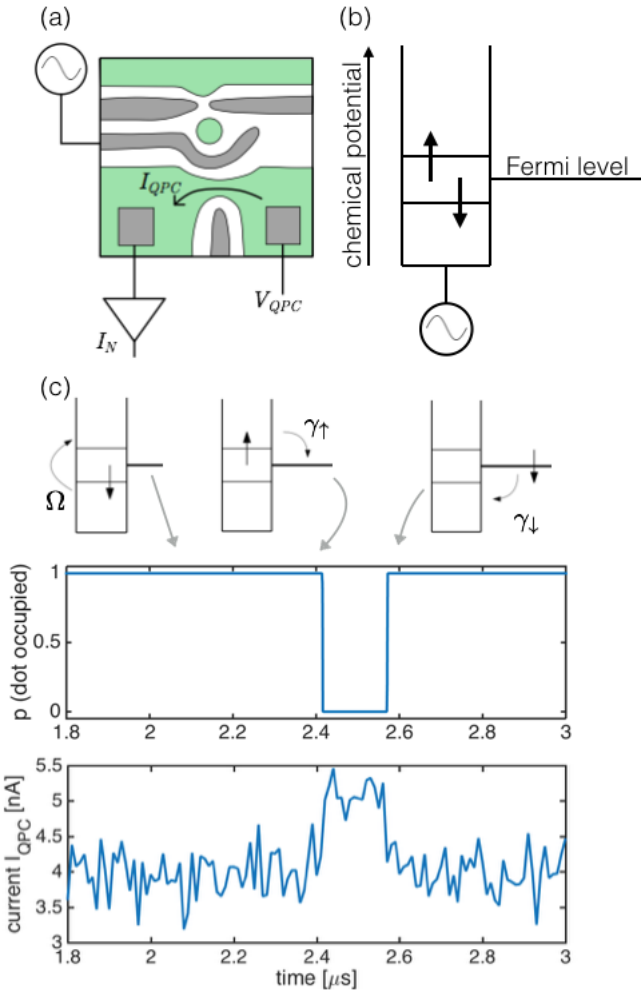

Figure 3.1: (a) Schematic of the experimental setup: GaAs quantum dot with the quantum point contact, yielding the experimentally accessible current I_{QPC} ; (b) Energy levels being populated through the exchange of electrons with the the lead electrode: a spin-down (up) electron can tunnel into (off) the dot; (c) Relation of the system dynamics to the measured current: the Rabi precession does not lead directly to a change in current, but does enable the electron to leave the dot, causing a temporary current increase until another electron tunnels onto the quantum dot.

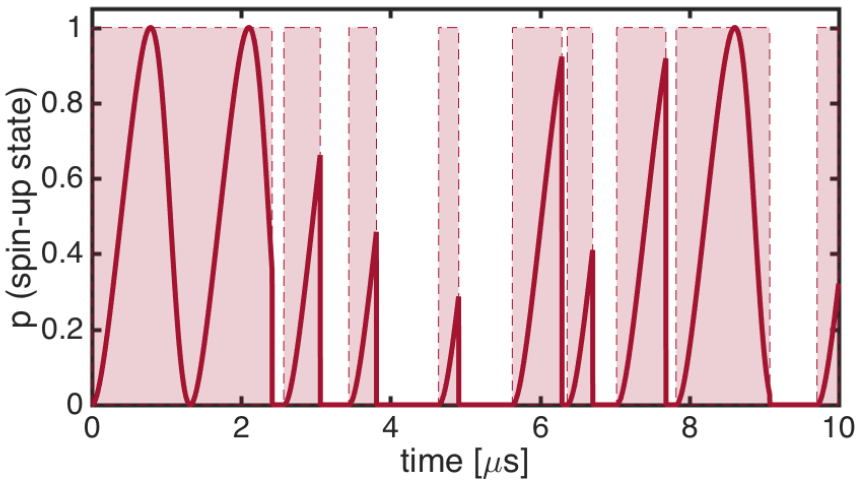

Figure 3.2: Example of a particular quantum trajectory: the occupation of the spin-up state is plotted as a function of time and shows the Rabi precession of the spin interrupted by the quantum jumps where the electron leaves the dot. The shaded red regions correspond to the time intervals when the quantum dot is occupied irrespective of the spin state. The figure is obtained with a spin Rabi frequency of $\Omega = 5$ MHz, and tunneling rates $\gamma_\uparrow = \gamma_\downarrow = 3$ MHz.

If $\gamma_\downarrow, \gamma_\uparrow$ denote the respective tunneling rates and dt is an infinitesimal time interval, the master equation [19] for the system reads

$$\begin{aligned}\frac{d\rho}{dt} &= -i[H,\rho] + \frac{\gamma_\downarrow}{2}(2c_\downarrow^\dagger \rho c_\downarrow - c_\downarrow c_\downarrow^\dagger \rho - \rho c_\downarrow c_\downarrow^\dagger) \\ &\quad + \frac{\gamma_\uparrow}{2}(2c_\uparrow \rho c_\uparrow^\dagger - c_\uparrow^\dagger c_\uparrow \rho - \rho c_\uparrow^\dagger c_\uparrow).\end{aligned} \tag{3.4}$$

This master equation describes the average dynamics of the unobserved system, subjected to both coherent driving between the spin states and incoherent tunneling onto and off the dot, described by rate equation terms. After a time of a few $\gamma_{\uparrow,\downarrow}^{-1}$ this equation causes ρ to converge to a steady state density matrix with populations of the three states and coherences between the two occupied spin states. Here, the system is modelled at zero temperature, but finite temperature can be straightforwardly added into the master equation [22]. The typical temperature in the dilution fridge is

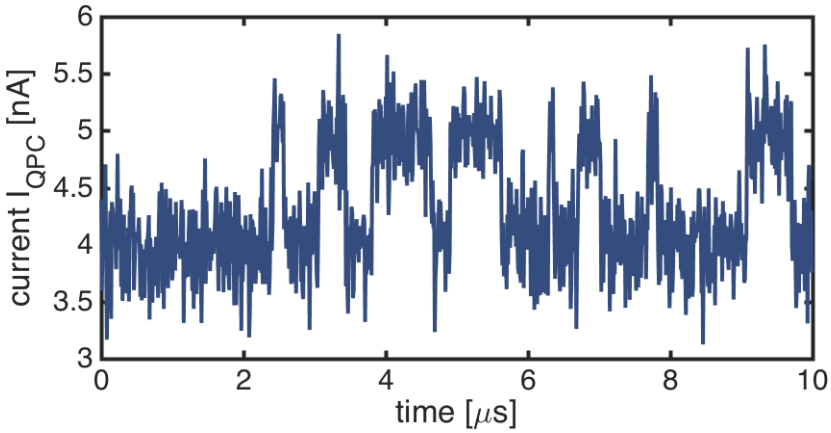

Figure 3.3: Simulation of the measured current through the QPC with the same parameters as in Fig.2, and with QPC current rates $r_0 = 31.21$ GHz and $r_1 = 24.97$ GHz. These rates lead to the charge dependent currents $I_{QPC,0} = 5$ nA and $I_{QPC,1} = 4$ nA while Poissonian count statistics (see text) leads to the standard deviation $\sigma_0 = 0.28$ nA and $\sigma_1 = 0.26$ nA for a measurement binning time of 10 ns.

below 30 mK, while the energy of Zeeman splitting of the electron spin resonance at typical magnetic fields corresponds to a temperature of ≈ 300 mK, which means that for our purposes the thermal bath of the quantum dot is effectively at zero temperature.

If we imagine that we could monitor the occurrence of each tunneling event, Eq.(3.4) would be replaced by a stochastic quantum trajectory [22, 23]. This means that in each time step, dt, the system density matrix is first evolved by the so-called no-jump dynamics

$$\begin{aligned}\frac{d\rho}{dt} &= -i[H,\rho] - \frac{\gamma_\downarrow}{2}(c_\downarrow c_\downarrow^\dagger \rho + \rho c_\downarrow c_\downarrow^\dagger) \\ &\quad - \frac{\gamma_\uparrow}{2}(c_\uparrow^\dagger c_\uparrow \rho + \rho c_\uparrow^\dagger c_\uparrow)\end{aligned} \tag{3.5}$$

and then made subject to one of the quantum jumps onto or off the quantum dot:

$$\rho \to \frac{c_\downarrow^\dagger \rho c_\downarrow}{\mathrm{Tr}(c_\downarrow^\dagger \rho c_\downarrow)} = |\downarrow\rangle\langle\downarrow|$$

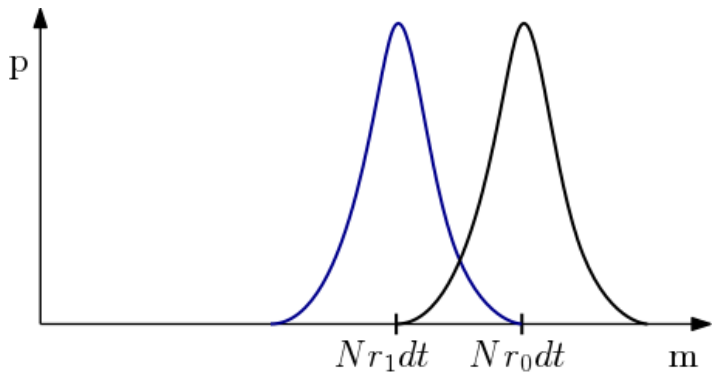

Figure 3.4: The probability that m electron counting events are registered by the QPC during a measurement time $T = Ndt$ (where N is large). For the empty and the charged quantum dot, a perfect detection yields QPC counts that are Poisson distributed with the mean values Nr_0dt and Nr_1dt and standard deviations $\sigma_0 = \sqrt{Nr_0dt}$ and $\sigma_1 = \sqrt{Nr_1dt}$ respectively.

or,

$$\rho \to \frac{c_\uparrow \rho c_\uparrow^\dagger}{\mathrm{Tr}(c_\uparrow \rho c_\uparrow^\dagger)} = |0\rangle\langle 0|.$$

These jumps occur with the probabilities $dp = \mathrm{Tr}(c_\downarrow^\dagger \rho c_\downarrow)\gamma_\downarrow dt$ and $dp = \mathrm{Tr}(c_\uparrow \rho c_\uparrow^\dagger)\gamma_\uparrow dt$, respectively, and ρ is renormalized, before the evolution is continued in the following time step. The formal relation between quantum trajectories, quantum jumps and the quantum theory of measurements is described, e.g., in [19].

We shall use trajectory dynamics to simulate the system, see Fig.3.2 which shows a sample evolution obtained in the way just described. The curve in the figure shows the spin-up population, dropping discontinuously to zero in connection with the simulated tunneling events off the quantum dot. The total occupation of the dot is unity in the intervals with the temporally modulated spin up population.

3.2 Continuous probing of the charge dynamics

We now describe the information about the quantum state available to an experimentalist who has access only to the fluctuating QPC current signal as illustrated in Fig. 3.3. We shall then compare this information with the simulated record of occupations on and off the quantum dot, depicted in Fig. 3.1, that gives rise to the QPC signal variation.

For the purpose of this analysis, we consider the current through the charge sensor as a stochastic counting signal, with two possible electron counting rates $r_{0,1} \equiv I_{0,1}/e$ depending on the charge on the dot. We assume a 1 nA difference between I_0 and I_1. While the QPC current record is measured as a continuous signal and does not resolve individual electrons, it is convenient to describe the measurement back action by the positive operator-valued measure (POVM) formalism [49] (see also Chapter 2), i.e., by operators M_c and M_{nc} yielding the probabilities and the back action associated with the count and no-count of each single electron passing the QPC in a time interval dt:

$$\begin{aligned} M_c &= \sqrt{r_0 dt}\Pi_0 + \sqrt{r_1 dt}\Pi_1, \\ M_{nc} &= \sqrt{1 - r_0 dt}\Pi_0 + \sqrt{1 - r_1 dt}\Pi_1, \end{aligned} \tag{3.6}$$

where $\Pi_0 = |0\rangle\langle 0|$ is the projection on the empty dot state and $\Pi_1 = |\uparrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow|$ is the projection operator on the charged dot state, with eigenvalues 0 and 1.

The probability for an electron to tunnel through the QPC in interval dt (a 'click event') is

$$P_{click} = \mathrm{Tr}(M_c \rho M_c^\dagger) = \rho_{00} r_0 dt + (\rho_{\uparrow\uparrow} + \rho_{\downarrow\downarrow}) r_1 dt \tag{3.7}$$

and the state conditioned on the click or no click ($x = c$ or $x = nc$), is

$$\rho|_x = (M_x \rho M_x^\dagger)/\mathrm{Tr}(M_x \rho M_x^\dagger). \tag{3.8}$$

In a real experiment one measures a macroscopic fluctuating current corresponding to the large number of electrons, m, passing through the QPC in finite time increments $\Delta t = N dt$. If the system is subject to no further evolution, Eqs. (3.6-3.8) then lead to an integrated number of electrons going through the QPC given by a sum of two Poisson distributions with mean values $N r_0 dt$ and $N r_1 dt$ and standard deviations $\sigma_0 = \sqrt{N r_0 dt}$ and $\sigma_1 = \sqrt{N r_1 dt}$ respectively (see Fig. 3.4). In the limit of large N, the current fluctuations can be modelled with Gaussian noise and various

corrections to Poissonian counting statistics [50] may be incorporated by adjusting the Gaussian widths. To simplify the presentation, we restrict our analysis to the Poissonian case given by Eqs. (3.6-3.8). As the master equation and POVM descriptions yield equivalent result, as discussed in Chapter 2, we shall refer to the conceptually simpler POVM formalism throughout this work, while stochastic differential equations may be more efficient in some numerical applications.

The QPC current and the noisy signals shown in Fig. 3.3 are determined from the "true" simulated state ρ_{true}. In contrast, we shall use the symbol ρ to represent the density matrix inferred only from the QPC measurements and from our knowledge that the spin precession and electron tunneling, while not monitored directly, cause the density matrix elements to evolve according to Eq. (3.4). We thus apply Eq.(3.4) and the POVM operators associated with the QPC detection to evolve the density matrix ρ. Note that, in general, the simulated ρ_{true} should also incorporate the back action of the QPC probing in its evolution, but as ρ_{true} populates with certainty either the charged state or the uncharged state, it is invariant under the QPC back-action Eq. (3.8). In the upper panel of Fig. 3.5 we compare the quantum dot occupation according to ρ (blue curve) and ρ_{true} (red regions). With the chosen parameters, the QPC probing is able to follow the charge dynamics of the quantum dot, but the state assignment is subject to statistical fluctuations due to the measurement noise.

3.3 Past Quantum State

For the parameters used here, the charge state changes on a similar time scale as our acquisition of measurement data, and as Fig. 3.5 shows, we cannot track the true system evolution perfectly because of the noise of the QPC current. In particular, we observe sharp features, indicating tunneling events, while subsequent rapid return to the original state, reveals that it was most probably a random signal fluctuation rather than a change of state. This is a well known problem in classical inference and has led to the introduction of data smoothing algorithms which incorporate the full signal and use data obtained both before and after t in the analysis of the state of a system at the time t. This can be done in a rigorous analysis known as the "forward-backward" analysis for classical hidden Markov models [37] and as the past quantum state (PQS) [1, 3, 31] and the quantum smoothing [51] formalism for quantum systems.

As the quantum formalism deals with the assignment of probabilities

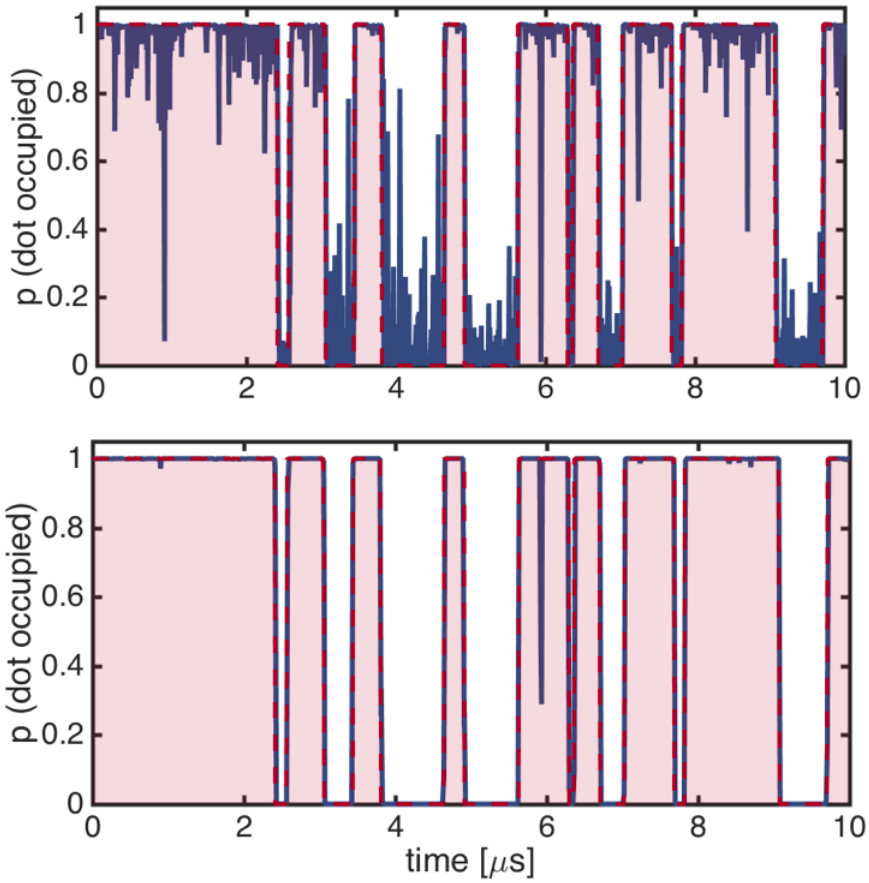

Figure 3.5: Comparison of true and inferred dynamics: the blue solid curve shows the occupation of the quantum dot as inferred from the QPC current. The red regions show the time intervals where the dot is actually occupied according to the simulated tunneling events on and off the quantum dot. The upper panel shows this comparison when the probability from the QPC current is inferred using the conditioned master equation. In the lower panel the probability is calculated using the past quantum state method.

to measurement outcomes, we can present our knowledge about a quantum system at time t by a general expression for such probabilities. The most general measurements are described by the formalism of POVMs [49]. We already saw examples of this formalism in our treatment of the QPC counting signal, and it quite generally assigns to any measurement a set of operators $\{M_i\}$, with $\sum_i M_i^\dagger M_i = I$, each representing an outcome (i) of the measurement.

The probability for outcome i is conventionally given by the density matrix expression,

$$P(i) = \frac{\mathrm{Tr}(M_i \rho M_i^\dagger)}{\sum_i \mathrm{Tr}(M_i \rho M_i^\dagger)}, \tag{3.9}$$

where the density matrix ρ may be the solution to our stochastic quantum dynamics described above, i.e., $\rho(t)$ is conditioned on the QPC measurement outcomes that occurred and were read out until time t.

The past quantum state formalism [31] offers an improved expression that incorporates the subsequent signal record via an "effect matrix" $E(t)$. In this formalism, the probability that a measurements of the observable corresponding to the POVM $\{M_i\}$ yielded the outcome i is

$$P_{PQS}(i) = \frac{\mathrm{Tr}(M_i \rho(t) M_i^\dagger E(t))}{\sum_i \mathrm{Tr}(M_i \rho(t) M_i^\dagger E(t))}. \tag{3.10}$$

Here, $\rho(t)$ depends only on the QPC measurement outcomes obtained *before* t and the matrix $E(t)$ depends only on the QPC measurement outcomes obtained *after* time t. In particular, E follows from an adjoint master equation that is solved backwards in time. To evaluate the matrix E introduced in [31], we assume the final value $E = I$ in the future of all probing measurements. Then we propagate the matrix elements backwards in time by incorporating the known Hamiltonian evolution and master equation damping terms of Eq. (3.4), as well as the QPC measurement outcomes, in a way that is adjoint to the evolution of ρ [31]. Between QPC counting events, the matrix E pertaining to the experimental observer, who does not observe the spin and charge dynamics, obeys the equation

$$\begin{aligned} \frac{E(t-dt) - E(t)}{dt} &= i[H, E] + \frac{\gamma_\downarrow}{2}(2c_\downarrow E c_\downarrow^\dagger - c_\downarrow c_\downarrow^\dagger E - E c_\downarrow c_\downarrow^\dagger) \\ &+ \frac{\gamma_\uparrow}{2}(2c_\uparrow^\dagger E c_\uparrow - c_\uparrow^\dagger c_\uparrow E - E c_\uparrow^\dagger c_\uparrow). \end{aligned} \tag{3.11}$$

In addition to the dynamics, decribed by Eq. (3.11), E undergoes changes associated with the QPC probing of the system. This is described by and

is analogous to the POVM operations acting on ρ in Eqs. (3.6)-(3.8), $E|_x = (M_x^\dagger E M_x)/\mathrm{Tr}(M_x^\dagger E M_x))$. We note that unlike the master equation (3.4) the evolution by Eq. (3.11) does not preserve the trace of E, but the explicit normalization of the outcome probabilities ensures the physical applicability of Eq. (3.10).

The results of the PQS analysis are depicted in the lower panel Fig. 3.5, again, in comparison with the simulated 'true' state. We observe that the assignment of charge states to the dot is significantly improved compared to the upper panel in Fig. 3.5.

3.4 Parameter Estimation

In order to infer the time dependent occupation of the dot from the measurement signal, we solved the master equation for $\rho(t)$ and the adjoint equation for $E(t)$, taking into account the measurement data, which could come from a real experiment but was synthesized by simulating the dynamics of a "true state". While the master equation (3.4) contains terms governed by the numerical values of the Rabi frequency Ω and the tunneling rates $\gamma_\downarrow$, $\gamma_\uparrow$, the abrupt changes of the occupation are mainly governed by the random measurement back-action, cf. the variations in the current signal shown in Fig. 3.3.

Larger or smaller values of the Rabi frequency and the rates would not be in conflict with short segments of the signal and the inferred occupation dynamics, but we expect a statistical agreement with the physical parameters when the system is observed over longer times. Thus, the periods with no occupation of the dot should follow an exponential distribution with the rate constant $\gamma_\downarrow$, and since the electron always enters the quantum dot in the spin down state and leaves in the spin up state, we expect to see the intervals of occupation of the dot cluster around odd multiples of $\sim \pi/\Omega$, and there should be very few intervals close to even multiples of $\sim \pi/\Omega$.

This intuition is confirmed by making a histogram of the electron dwell times on the dot, extracted from the PQS analysis of the dot occupation (Fig. 3.6). Even more so, it permits a fit to the analytical dwell time distribution for different values of Ω,

$$w(t) = -\frac{2\Omega^2\gamma_\uparrow}{\kappa^2}\exp\left(-\frac{t\gamma_\uparrow}{2}\right)\left(\cos\left(\frac{t\kappa}{2}\right)-1\right), \tag{3.12}$$

where $\kappa = \sqrt{4\Omega^2 - \gamma_\uparrow^2}$ with $2\Omega > \gamma_\uparrow$.

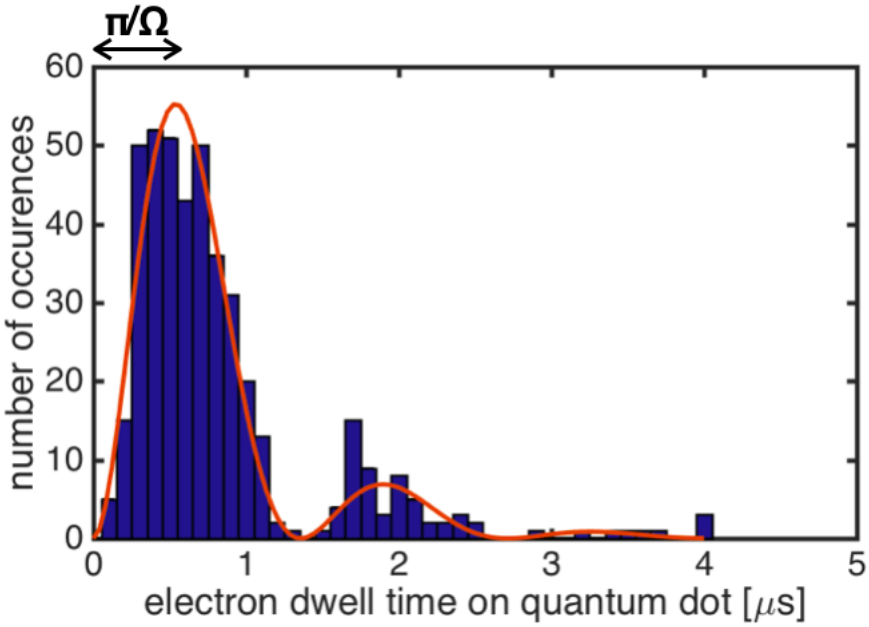

Figure 3.6: Histogram of the time intervals in which the quantum dot is occupied by an electron according to the past quantum state analysis of the QPC measurement data. The modulation reflects the Rabi oscillation between the spin states and the red curve represents the fit of the histogram by the expression (3.12) which leads to an estimated Rabi frequency of ~ 5.2 MHz in reasonable agreement with the value of 5MHz used in our simulation.

So, even though the experiment is not directly sensitive to the spin state, it is able to reveal the spin dynamics and determine its characteristic frequency. Supplemented with a histogram for the periods with no electron on the dot, we may also estimate the tunneling parameters.

In our particular simulations, the past quantum state analysis provides decisive information about the occupation of the dot, and the occupation dynamics is well resolved. A more general procedure allows extraction of the physical parameters, even when the inferred dynamics does not assign the state with certainty.

An optimal estimation of the Rabi frequency is thus made by Bayes rule, which assumes a prior, e.g., uniform, probability distribution for Ω,

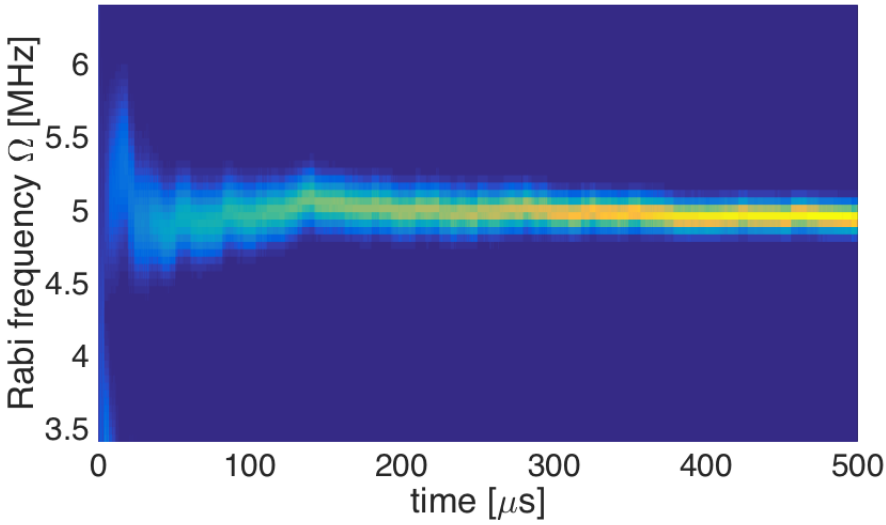

Figure 3.7: Bayesian estimation of Rabi frequency Ω. The color scale depicts the likelihood function $L(\Omega)$. We used a grid of 60 candidate values between 3.5 MHz and 6.5 MHz. We see clear convergence towards the correct value of $\Omega = 5$ MHz.

and evolves a separate density matrix $\rho_\Omega(t)$ for each of a set of candidate values. The QPC current outcomes occur with higher probabilities for some than for other conditional states ρ_Ω, and the likelihood $L(\Omega)$ for the different Ω are merely multiplied by these probabilities at each time step. During the parallel time evolution of the different candidate density matrices, the value of Ω with the highest likelihood represents the best estimate for the Rabi frequency. An example of the Bayes rule evolution of $L(\Omega)$ with time is shown in Fig. 3.8. We see that the method quickly identifies the plausible range of values for Ω, and that the estimate becomes more peaked as data accumulates. A more detailed analysis shows that the estimation error scales as $1/\sqrt{T}$ with long probing times T [36].

It is in principle possible to identify multiple parameters by Bayes method, but it requires propagation of conditional density matrices, with each parameter exploring a grid of candidate values. One therefore has recourse to other methods if the number of parameters is too high. For incoherent processes, the so-called parameter re-estimation method consists in determining the dynamics of the system subject to a choice of parameters,

and then extracting the apparent rates from the dynamics. Using an initial guess for the rate parameters, one determines the time dependent occupation of the different states as well as the joint distribution for occupying state i at time t, and state j at the next time step according to the full measurement record. From the same probabilities and joint probabilities, one subsequently extracts new candidate values for the transition probabilities. The whole data record is then analyzed again, but with these inferred rates, and the procedure is repeated until it has converged and the rates applied in the dynamics are in agreement with the rates inferred from the population dynamics.

This so-called Baum-Welch estimation scheme has been developed for classical parameter estimation, which means that it also applies to incoherent quantum processes as shown in case of atomic dynamics in a cavity QED experiment [37]. It cannot be applied directly here, since we have a combination of coherent (Rabi frequency Ω) and incoherent (tunneling rates $\gamma_{\uparrow,\downarrow}$) dynamics. In particular, the correlations between the population of the states at different times are not only due to the incoherent rates but also due to the coherent processes coupling the populations of states i and j via the coherence ρ_{ij}, $i \neq j$. Results of the modified Baum-Welch method designed to include the coherent part of the dynamics are shown in Fig. 3.8. The details of the method are discussed in Appendix A.

Understanding how to estimate coherent parameters assuming known values of the incoherent ones and vice versa, we propose the hybrid scheme illustrated in Fig. 3.9. The method assumes (guessed) values for the rates, and applies the Bayes rule method to find the most likely value of Ω. That value, together with the rates is subsequently used in the propagation of the master equation for $\rho(t)$ and the adjoint equation for $E(t)$. With the values of $\rho(t)$ and $E(t + dt)$, we then calculate the joint probability for finding the system in state i at t and in j at $t + dt$, which should agree on average with the rates assumed in the calculation of the dynamics. As illustrated in Fig. 3.9, this effective Baum-Welch protocol can be iterated until convergence (n times) for the given Ω, and the resulting rates are used in a new Bayes estimation of Ω. The step is repeated (N times) until all parameters are converged and consistent with the data.

When applying this method to our simulated data $n = 5$ modified Baum-Welch iterations and $N = 5$ iterations of the full protocol were sufficient to estimate all three parameters. We observe in Fig. 3.8 and Fig. 3.10 that our estimation does not identify the 'true' parameters used for the simulation exactly. This is due to the use of finite data sampling and better agreement is expected for longer probing times. Even though

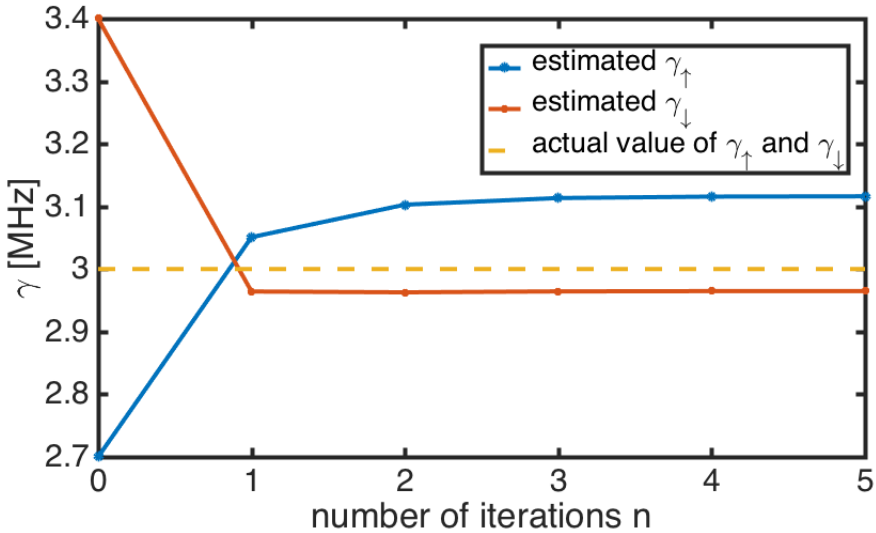

Figure 3.8: Estimation of $\gamma_\downarrow$, $\gamma_\uparrow$ using the modified Baum-Welch procedure. Assuming the correct Rabi frequency, Ω, $\gamma_{\uparrow,\downarrow}$ is shown as a function of number of iterations n.

the rates are identical $\gamma_\uparrow = \gamma_\downarrow$, the fact that they govern oscillatory and exponentially damped waiting time distribution, respectively, explains that they may be subject to different estimation errors.

3.5 Experimental Considerations

In our model, the tunneling dynamics is correlated with the spin precession, and while the spin dynamics is not directly observable in the experiment, it is possible to infer its properties from the charge measurements. In particular, the Rabi frequency is evident in the distribution of time intervals where the quantum dot is charged. A Bayesian analysis of the noisy measurement current record permits a reliable estimate of Ω, while a generalization of the classical Baum-Welch algorithm allows estimation of the tunneling rates.

By adjusting the bias across the charge sensor and therefore the strength of the measurement, our method offers a versatile approach to weak measurements of spin qubits [52]. All of the ingredients of this approach have now been demonstrated experimentally [46]. The key technology of

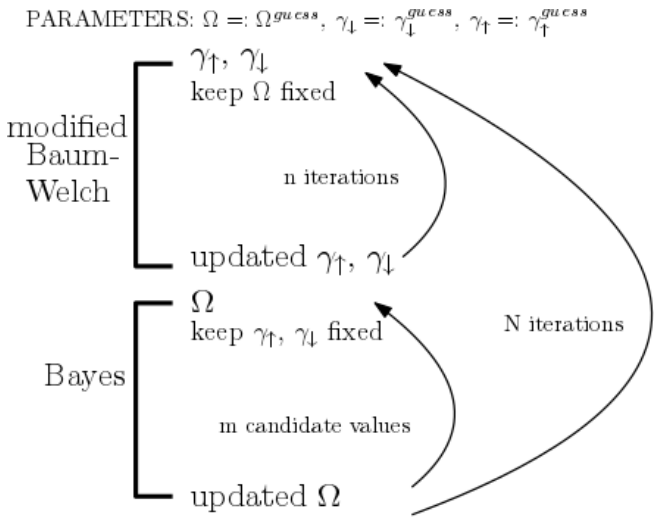

Figure 3.9: The scheme of our numerical approach to estimate both coherent and incoherent physical parameters: We fix the Rabi frequency and use n iterations of the Baum-Welch reestimation procedure to obtain consistent values of the rates. With these rates fixed, we use Bayesian estimation to pick the lost likely Rabi frequency among m candidate values. The whole protocol is repeated N times until all parameters have converged.

spin readout via a charge sensor has been achieved both using dc current as in Fig. 1 [12, 47, 48] as well as via the faster but conceptually similar technique of radio-frequency reflectrometry [53–59], to which our analysis can also be applied. Single-spin rotation can be achieved in several ways; by applying a local magnetic field [60, 61], by electrically driven spin resonance [62, 63], or by coupling to an on-chip field gradient [64].

Our simulations used realistic physical parameters. Thus, by operating at a qubit frequency of 40 GHz (corresponding to a magnetic field ~ 7 T in GaAs or ~ 1.4 T in Si), thermal fluctuations at 30 mK may be neglected. The variation of our simulated QPC signal is compatible with the reported ability to perform charge readout and distinguish the occupation in a quantum dot in 100 ns [65]. Finally, Rabi oscillations as slow as 1.5 MHz [62] have been measured and may thus form comparison with the achievement of the method.

Let us now turn the discussion to the use of the method beyond the

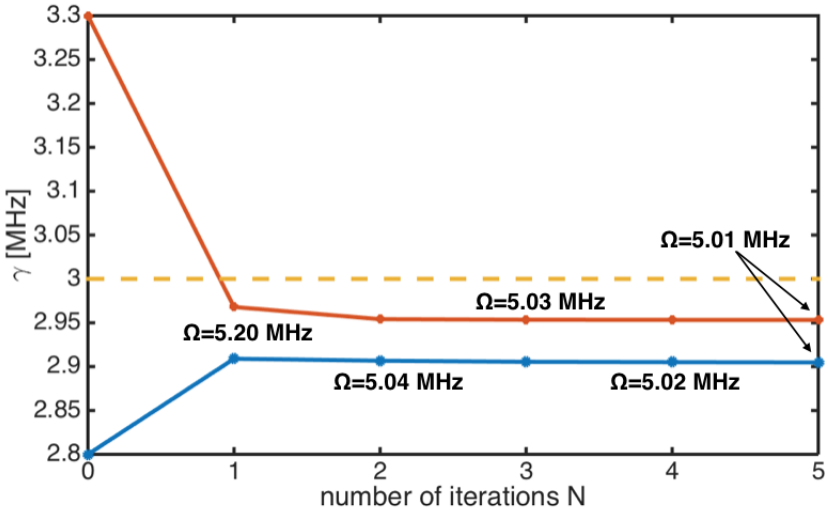

Figure 3.10: Estimation of $\gamma_\downarrow$, $\gamma_\uparrow$ using the modified Baum-Welch procedure. Parameters $\gamma_{\uparrow,\downarrow}$ are shown as a function of number of iterations N. Rabi frequency Ω associated with N-th iteration is shown at each point.

parameter regime relevant for this particular experimental realization.

3.6 Adding noise

Let us consider the case in which the noise would prevent us even more aggressively to recognize any structure in the experimental current. In the experiment described above, all the technical advances in the field of semiconductor quantum information allow for routinely detecting QPC currents similar to the simulated one shown in Fig. 3.3, where the difference between occupied and un-occupied quantum dot translates into approximately 1 nA difference in the measured current. This means that, albeit noisy, the bimodal structure of current is directly visible. That is also a reason, why we are able to construct the histogram in Fig. 3.6. To illustrate, how powerful past quantum state can be, in this section we consider the case when the difference between empty and occupied states of the dot would manifest itself with the measured current difference that is 10 times smaller than one typically measured in the experiment we considered above, i.e. 0.1 nA.

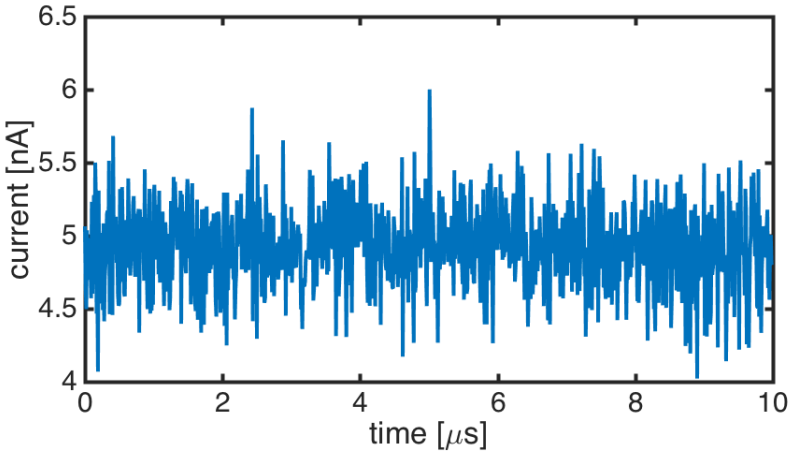

Figure 3.11: Estimation of $\gamma_\downarrow$, $\gamma_\uparrow$ using the modified Baum-Welch procedure. Assuming the correct Rabi frequency, Ω, $\gamma_{\uparrow,\downarrow}$ is shown as a function of number of iterations n.

This assumption leads to the current shown in Fig. 3.11 and we can see that in this instance we can not recognize by the naked eye whether the quantum is occupied or not. Repeating the calculations outlined in Sections 3.1-3.2 we obtained results shown in Fig. 3.12. As above, we show a trajectory of the spin and the 'true' occupation intervals in red and the QPC-inferred dynamics in blue. While we are not able to distinguish the electron dwell times on the dot as sharply as before, we can see that our methods recognizes the intervals quite reliably. Using this information about the probability of the dot occupation we can in principle go ahead and repeat the parameter estimation shown above even for this extreme case. Our simulation returned $\gamma_\uparrow = 3.1$ MHz, $\gamma_\downarrow = 2.9$ MHz, and $\Omega = 5.4$ MHz. So while these values are less close then the ones we have estimated for the experimentally relevant case, they are close to the correct values $\gamma_{\uparrow,\downarrow} = 3$ MHz and $\Omega = 5$ MHz especially in the light of the fact there is no visible structure in the experimental current.

3.7 Conclusion

In the preceding sections we have investigated the prospects of monitoring the spin and charge dynamics of a single electron quantum dot with a QPC. We have shown that the conditional master equation allows inference of

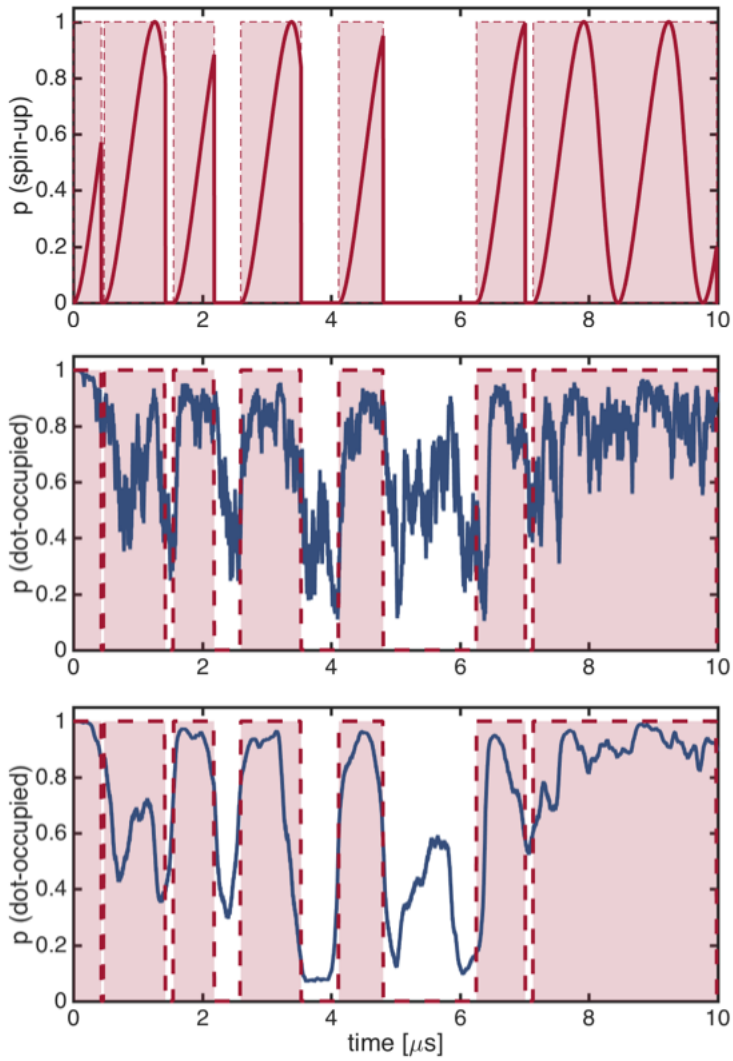

Figure 3.12: Recovering the spin dynamics using QPC probing in the extremely noisy case: the upper panel depics the quantum trajectory and the shaded regions show when the dot is occupied, the middle panel shows the probability of the dot being occupied using forward master equation analysis (blue line), bottom panel shows the same probability calculated using past quantum state.

the quantum state of the system, and that analysis of the full temporal signal improves the statistical certainty about the tunneling events in the system.

For a detailed description of the numerical simulations of the system in question we would like to direct reader's attention to Appendix D. In addition to that, when conditioning on both future and past events many interesting question arise from a quantum information point of view. The connection of the generalized Born rule to the so-called 3-box paradox [66–68] is discussed in Appendix B.

In spin-based quantum computing, one method for mitigating the fluctuating hyperfine coupling that limits operation fidelity is to track the instantaneous hyperfine field dynamically from its effect on the electron spin precession frequency [69–71]. Also, a record of the spin qubit precession frequency can be used to estimate a local electric field [72]. We believe that our results can be applied in a number of such situations where parameters that govern spin evolution must be estimated efficiently.

In general, the approach introduced in this Chapter can be readily applied to any other quantum system, where the parameter of interest governing the dynamics is not accessible by direct experimental measurement. One immediate example significant in the development of future quantum technologies would be precession of electronic spin inside Nitrogen Vacancy (NV) center [73] used in magnetometry [74].

Chapter 4

Brief introduction to machine learning for physicists

Machine learning generally stands for a wide number of techniques and sub-fields. In this Chapter we give an introduction to the to-date most successful machine learning technique, artificial neural networks. This Chapter overlaps with the Supplemental Material of [5]. It is based on the lectures by Florian Marquardt [75] and the books [76, 77] and it should provide a concise introduction into artificial neural networks in order to ease reader's way into Chapter 5, where machine learning (namely convolutional neural network) is applied in the context of open quantum systems and parameter estimation.

Artificial neural network model is based on the idea of reproducing neurons connection in human brain. As schematically shown in Fig. 4.1, the neural network is supposed to be trained in a fashion similar to the human learning with the ultimate result of being able to process complex inputs and conclude correct outputs just like a brain does.

A neural network is defined as a non-linear function of number of variables that depends on multiple parameters. It is composed of neurons that are organized within layers, which are interconnected with each other in such a way that the strengths of these connections represent parameters that can be optimized. This optimization is done by so-called training of the neural network.

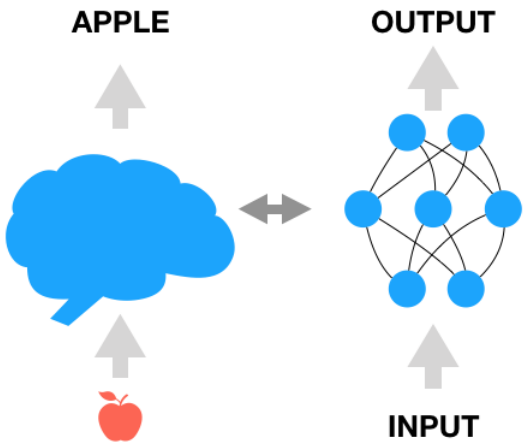

Figure 4.1: Illustration of a single neuron within neural network.

To understand these connections, let us a consider a single neuron in a given layer as shown in Fig. 4.2. This neuron is connected to n neurons in the preceding layer $\{y_1, \ldots, y_n\}$ with the weights $\{w_1, \ldots, w_n\}$. The output of a neuron is given by nonlinear function, f, applied on the weighted sum of inputs y_i. More specifically, this weighted sum is given by

$$z = \sum_j w_j y_j + b, \tag{4.1}$$

where w_j is above mentioned weight representing the strength of a given connection and b is a constant offset called bias. The superposition, z, then serves as an input for a nonlinear operation, f, designed to heighten differences between different values. Some simple examples of the function f (often called activation function) are for instance the sigmoid function or the so-called rectified linear unit (ReLU) function shown schematically in Fig. 4.3. The sigmoid function asymptotically approaches the Fermi function and the ReLU function is defined to return 0 for negative inputs and input itself for the positive ones.

When scaling up and creating neural network, each neuron has it's own bias and set of weights (these are the parameters that will be adjusted by training), while the nonlinear function is typically fixed for the

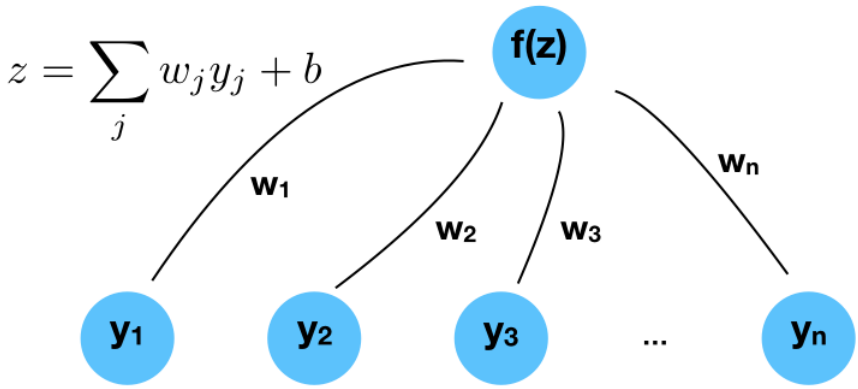

Figure 4.2: Illustration of a single neuron within neural network.

whole network (although the output layer often has a different activation function).

In order to train the network we need to have a way to check whether we are adjusting the weights and biases correctly for reaching the required goal. The success of the optimization is measured by the so-called cost function. Let us consider the input vector y^{in} and the output vector y^{output} such that the neural network acts as

$$y^{output} = F_w(y^{in}), \tag{4.2}$$

where index w denotes the vector of all flexible weights and biases that are to be adjusted by training. Now, if the network would function perfectly, this transformation would be given by

$$y^{output} = F(y^{in}), \tag{4.3}$$

where F is the desired target function (this can both be well defined analytical function or a function that is only defined on the inputs of the training data). One typical example of cost a function definition is then the following

$$C(w) = \frac{1}{2}\langle ||F_w(y^{in}) - F(y^{in})||^2 \rangle, \tag{4.4}$$

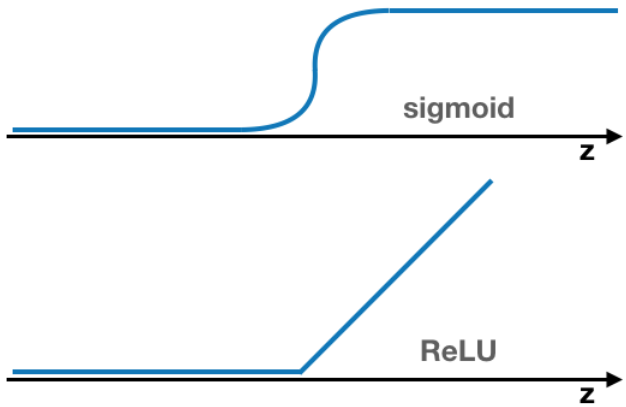

Figure 4.3: Examples of the non-linear functions, f, applied to the result of linear operation, z.

where the norm is standard vector norm and the average is taken over all the input vectors y^{in} used for training (the so-called training set). Of course, to analytically evaluate this function precisely we would need to average over the whole training set, which, in most practical situations, is if not intractable, very impractical. In practice, to minimize the the cost function we find an approximation of the cost function and its derivative, and in discrete steps we adjust the weights w_j according to

$$w_j \mapsto w_j - \eta \frac{\partial \tilde{C}(w)}{\partial w_j}, \tag{4.5}$$

where $\tilde{C}(w)$ is an approximation of the cost function from the previous step and η specifies the step size. This is the main idea of the so-called gradient descent method [78], which is one common way to adjust the cost function in practice.

To make the calculation of the gradient of the cost function (4.5) computationally tractable, the so-called back-propagation algorithm [79] is being used. Roughly speaking, the back-propagation algorithm is based on the fact, that since the whole structure of the neural network is known, one may just use the output of the network and trace-back the calculations through the network while adjusting all the weights and biases.

Figure 4.4: Examples of the hand-written digits from the MNIST database.

More precisely, the back-propagation allows us to compute $\frac{\partial \tilde{C}(w)}{\partial w_j}$ first for a single input vector (training example) and then estimate $\frac{\partial C(w)}{\partial w_j}$ by averaging over many elements of the training set. The details of the algorithm are well explained for example in [76].

Let us now apply these basic concepts of machine learning onto an example that is often being called 'hello-world' of neural networks, and that is the recognition of hand-written digits from MNIST database [80].

In this problem we want to train the neural network on the set of 28x28 pixels (like those shown in Fig. 4.4), such that it learns to recognize digits in any hand-writing. This is an example of category classification problem. We have a huge set of input data that can we wish to sort into 10 categories (digits).

The neural network for this problem is constructed as follows: the input layer contains $28 \cdot 28 = 784$ neurons, each of which contains information about one pixel of the input figure. The output layer contains 10 neurons, each of which we wish to contain information about probability that the input digit was in a given category.

In terms introduced above, the 784 digit input is transformed according to Eq. (4.1). For the activation function of the final output layer, f, we want something that would transform the output into the probabilities, thus the candidate activation functions presented above will not do. There is, however, an easy generalization of the sigmoid function that normalizes output just right

$$f_j(y_1, \ldots, y_n) = \frac{\exp(y_j)}{\sum_{k=1}^{n} \exp(y_k)}. \tag{4.6}$$

This is the so-called softmax activation function and it is often the activation function of choice for category problems.

As for the cost function, since we are now working with probabilities we are able to make a slightly more sophisticated choice than above that

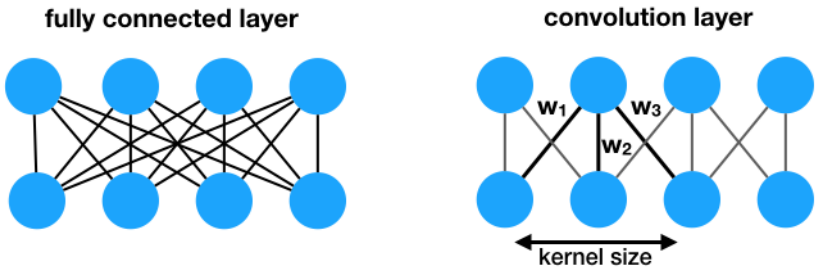

Figure 4.5: Illustration of the difference betweem fully connected dense layer and convolution layer.

is also physically meaningful. For any probability distribution $\{p_j\}$ we can define the entropy function as $S = -\sum_j p_j \ln p_j$. To quantify how 'far' we are from the desired output we introduce the cross-entropy as

$$H(y^{target}, y^{output}) = -\sum_j y_j^{target} \ln y_j^{out}. \tag{4.7}$$

Here, y^{output} is a vector outputted by the neural network, and y^{target} is a vector encoding the correct classification (the correct outputs are stored as one-hot vector, where there is 1 at the position corresponding to the correct digit and 0 elsewhere), i.e. $y_j^{target} = F_j(y^{in})$. This function, Eq. (4.7) is always non-negative and becomes zero only for $y^{target} = y^{output}$.

Now the neural network can be trained on a part of MNIST database by minimizing the cross-entropy (4.7) for a set of training samples. A very important point when doing this is to divide the training set into two non-overlapping parts, where one is solely dedicated to training and the other to validation. When the training is finished, the validation set serves to establish the fidelity with which the network can recognize the digits (using the training examples for this could lead to over-estimation of this fidelity for the cases when the network is sufficiently large to 'remember' the training samples by overfitting). A simple Python implementation using Tensorflow library [81] shown in [82].

So far we discussed the activation and cost functions as the main features for characterization of our neural network and its learning. There is, however, one more significant factor that requires further discussion and that is the structure of how the neurons are interconnected within the network. Up until now we implicitly assumed that all neurons in a given layer are connected to all neurons of the following layer and so

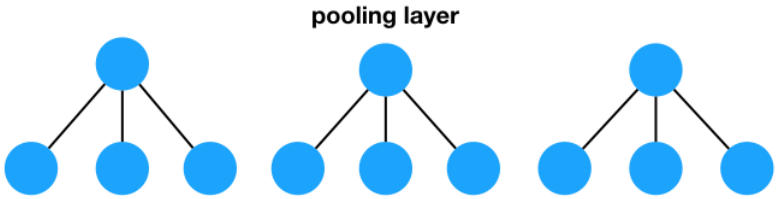

Figure 4.6: Illustration of the functionality of the pooling layer.

on (as schematically illustrated in Fig. 4.1). These types of connections are an integral part of ANNs and are almost always needed to construct well-functioning neural network. Not all layers however have to have this type of complex connections and for many cases we can get away with much less.

When we are in the situation described above, and all the neurons of the given layer have a connection to each neuron of the following layer, the layer is called *fully connected dense layer*. Another type of layer that found huge importance especially in image recognition is the so-called *convolution layer*. This type of layers explores the translational invariance of many problems. The difference between a fully connected layer and a convolution layer is shown in Fig. 4.5. A significant simplification of the convolution layer comes from the fact that instead performing optimization for all the weights of all the connections shown in the left panel, we only choose a finite size neighborhood of a given neuron that will influence the corresponding neuron in the following layer. In our case this amounts to nearest neighbors, i.e. 3 connections in total, that will be translationally invariant along the layer.

Technically, we are performing a discrete convolution between the layers and the neighborhood we choose to connect to the upper layer correspond to the kernel size of the Green's function we use for this convolution [1]. The translational invariance is then a direct consequence of the same invariance of the kernel of the convolution integral. In the language of image recognition this means that all the features learned in one part of the image will be authomatically recovered in the other parts. This approach also comes with a clear reduction in the computation complexity: while the

[1] A convolution for functions f and g is defined as $f * g = \int_{-\infty}^{+\infty} f(\tau)g(t-\tau)d\tau = \int_{-\infty}^{+\infty} f(t-\tau)g(\tau)d\tau$, where $g(t-\tau)$ $(f(t-\tau))$ is the kernel.

fully connected layer requires us to store (number of neurons)^2 of weights, the convolutional layer gets by with the number of weights corresponding to the size of the kernel of the convolution. An excellent resource on the technical aspect of convolutional neural networks are lectures by Fei-Fei Li [83].

The kernel is called a filter, and the size of the kernel corresponds to the number of weights in the filter. The filter size is an important parameter for convolutional layers. In practice multiple filters are typically being used to address different features of input data.

The last type of layer highly relevant for further discussion is the so-called *pooling layer*. This layer is used for a down-sampling operation that reduces amount of parameters and computation in the network. In practice this means that we divide given layer into areas and then pick one representative value for each, so we can represent that whole area with a single neuron in the following layer. The choice of this representative value depends on the particular problem, but typical choices are to pick either maximum value of average value. The pictorial representation of the pooling layer functionality is straightforward and is shown in Fig. 4.6.

An example of an evolved artificial neural network is shown in Fig. 4.7. This particular structure of neural network is used in image recognition. The convolution layer is very well suited for classification of input pixels and is followed by pooling which reduces the amount of data. After applying several convolution filters the densely connected is layer is used before producing the output. This is the network structure we use in context of recognizing parameters of quantum systems in Chapter 5.

Most of the explanations in this Chapter have been oriented towards the image recognition, because that is going to be most relevant in the upcoming Chapters. Before concluding, let us point out other applications for ANNs.

An important example is autoencoding [84, 85]. Autoencoders are used as a form of compression adapted to the specific features of a particular type of input. An example of auto-encoding would be reconstruction of the image for which we have only limited amount of storage space. While in general terms autoencoding can not compete with universal machine-learning unrelated methods, it is still very good at representation of particular input-dependent features of the data. Autoencoders are also rising in importance in the context of quantum machine learning [86, 87], because of the limited size of the near-term quantum processors extra input-dependent compression will be extremely important to down-size the input size into one suitable for initialization of small quantum register.

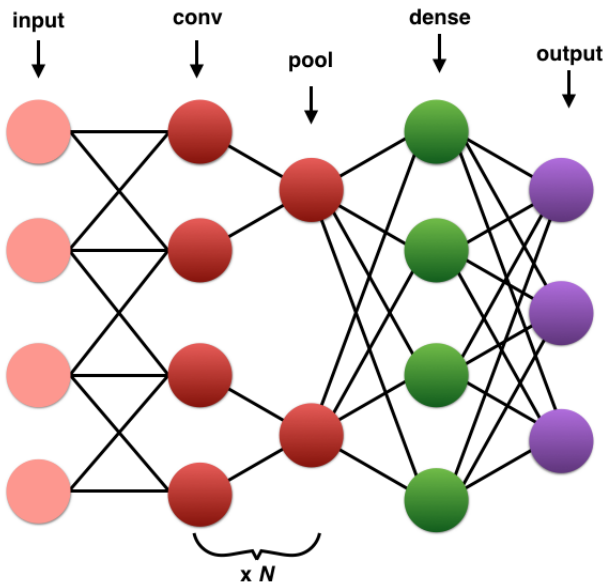

Figure 4.7: Example of evolved neural network for image recognition: convolutional layer followed by pooling layer N times connected with with dense layer.

Finally, autoencoding is an example of the so-called unsupervised learning. With the image-recognition example above, we were considering supervised learning in a sense that we know the correct answer and we use the comparison with this correct answer when training the network. Now, when we are in the situation that we simply want to reproduce an input or enhance certain of its features, we do not need the correct answer to compare to, autoencoding is learning just by analyzing the input data (and hence without supervision).

There is also somehow intermediate stage between supervised and unsupervised learning and that is reinforcement learning [88], which is learning based on the rare reward. In particular, probabilistic action is used by the network and whenever the answer turns out to be correct all the actions in the same class get assigned higher probability. This is the learning strategy typically used to develop AI capable to perform well in various computer games, most prominent example being Google's neural network mastering Go [89, 90].

There is a number of ways aritificial neural networks can be emplyed for applications in quantum physics [91–95]. In Chapter 5 (see also [5]) we show, that the convolutional neural network introduced in this Chapter can be directly used for the purposes of parameter estimation in quantum systems. In particular, we show that machine learning methods work in the presence of detector imperfections that rule out Bayesian analyses.

Another direction, taken by [91], is to use a neural network as a new ansatz for the description of many body quantum states. This has led to new ways to understand entanglement in many-body systems [93].

Finally, there is an option to make use of entanglement in quantum systems and design a quantum neural network that offers learning speed-up in comparison with classical neural networks [86].

Chapter 5

Quantum parameter estimation with neural networks

5.1 Introduction

In this Chapter we would like to use the tools briefly introduced in Chapter 4 and apply them to a problem of inference of relevant physical parameters from experimental currents. We propose to use neural networks to estimate the rates of coherent and incoherent processes in quantum systems from continuous measurement records. For this purpose, we treat experimental current signals as one-dimensional images and we succeed in training the network to extract the relevant features of the signals and correctly distinguish between different candidates for the underlying physical parameters. We demonstrate the method by simulating measurements on a quantum dot system subject to both coherent spin dynamics and incoherent tunneling dynamics. Under ideal, infinite bandwidth and noise-less detection, a maximum likelihood analysis of the unknown parameters can be obtained from a Markovian, stochastic master equation, but such equations do not apply if, e.g., finite bandwidth amplifier noise is added or a finite detector bandwidth introduces non-Markovian correlations in the measured signal. We simulate such distortion of ideal signals, and we show that the neural network is perfectly capable of distinguishing signals pertaining to different values of the coherent and incoherent physical parameters of the quantum dot.

Sections 5.2-5.5 and parts of this introduction overlap with [5]. The technical details entering the simulation of the model introduced in this Chapter are summarized in Appendix E, which is part of Supplemental Material of [5].

5.2 Analyzing experimental currents with neural network

The development of quantum technologies with improved performance for a wide number of tasks rely on our ability to manipulate and control individual quantum systems which, in turn, assumes precise knowledge of system parameters such as transition frequencies, coupling strengths and dissipation rates. The operation of quantum communication and computation devices [49] may require frequent verification and calibration measurements [96], while the whole purpose of quantum sensing [97] and metrology [98] is to extract such parameters from measurement records.

For repeated measurements on a quantum system, the probability, or likelihood, for each measurement outcome is governed by the system density matrix or state vector which in turn depends on the parameters governing its dynamics, and sampling of N independent measurement results leads to an estimation error that decreases as $1/\sqrt{N}$. For continuous measurements on a single quantum system, each random signal output is accompanied by measurement back action, which has consequences for future measurement outcomes. For many dissipative quantum systems, measurements are only correlated over finite time, and the estimation error decreases as $1/\sqrt{T}$ as function of the total duration, T, of the measurement [99] (see also Chapter 2). The optimal estimation of unknown parameters $\{\theta\}$ from a measurement signal is governed by the Cramér-Rao bound [100] and the Fisher information [101] which, in turn, involve the likelihood $L(D|\{\theta\})$ of the signal D conditioned on the values of the parameters in question [36] (see also discussion in Chapter 2). If this likelihood function is known, Bayes' rule, $L(\{\theta\}|D) \propto L(D|\{\theta\}) \cdot P_{prior}(\{\theta\})$ saturates the Cramér-Rao bound and provides the optimal maximum likelihood estimate for $\{\theta\}$. But, it is in general very complicated to evaluate $L(D|\{\theta\})$, and hence to obtain an optimal parameter estimate when the signal is in the form of a continuous time dependent record $D = I(t)$.

In this Chapter, we propose to use an artificial neural network to analyze detector signals and identify the most likely set of parameters based on machine learning from independently generated training signals.

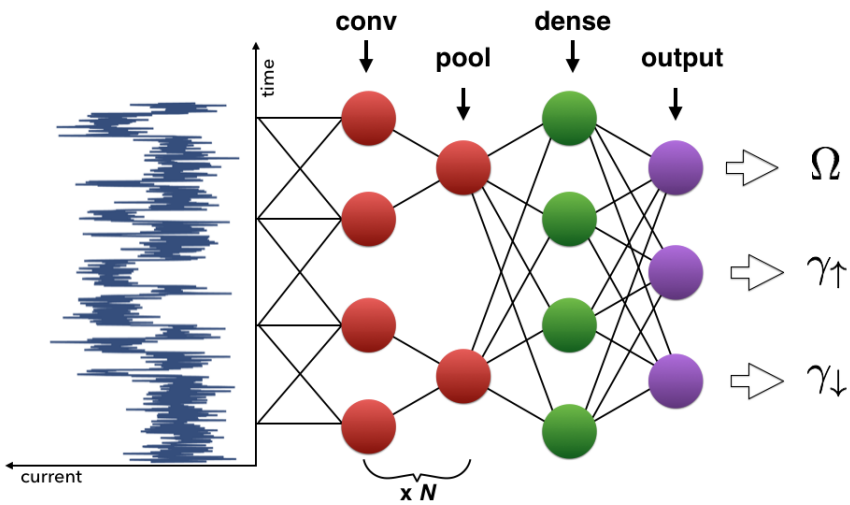

Figure 5.1: The neural network used for parameter estimation based on analyses of experimental signals. From left to right, the figure shows how the experimental data are stored in an input layer of nodes that is connected to a convolutional layer which applies one or several series of filters followed by a pooling layer and a densely connected layer that identifies finer features in the data, and is then connected to the output. The output layer yields the probabilities and most likely candidate values of the parameters governing the production of each signal. The function of the neurons, shown as circles in the figure, is to forward output values that are simple functions of the inputs, and the purpose of the machine learning is to properly adjust these functions for optimal recognition on a large training set.

Machine learning has recently found numerous applications in quantum physics [91–94] and it has been used for parameter estimation by projective measurements on quantum systems [95]. Our work differs from these previous efforts by applying machine learning to the analysis of the weaker and more noisy measurements associated with continuous monitoring of an experimental system. Our measurement signal thus consist of a time series of data which we regard as an one-dimensional image in order to directly adapt machine learning algorithms developed for image recognition. Specifically, we employ convolutional neural networks, which are known to be highly effective for classification of images (see general introduction to machine learning and neural networks in Chapter 4, [75–77, 83]). The neural network is exposed to signals obtained for a finite number of parameter sets and the network is not provided with any information about how these parameters are associated with the dynamics of the system and the ensuing signals.

In Chapter 3 and [4] we have simulated sensing of the charge on a quantum dot by the noisy current probed by a quantum point contact (QPC). Under a Markovian assumption for the tunneling dynamics and the QPC current signal, the system was simulated and analyzed by a stochastic master equation [19, 21]. By propagating quantum trajectories associated with each candidate set of parameters, we are able to continuously update the likelihood for each set in a Bayesian manner [36, 43, 44, 102, 103] as discussed in Chapter 2. Such Bayesian analysis becomes cumbersome for large parameter spaces and, as shown in Chapter 3, it is also possible to combine Bayesian estimation of some parameters with a more straightforward frequency analysis of the others. Note, however, that these analyses make very specific assumptions about the dynamics and formation of the signal, i.e, they assume the validity of a stochastic master equation and perfect knowledge of all other parameters than the ones that we wish to estimate. Inefficient detection and more complicated effects such as dead time and finite band width effects can be incorporated in the stochastic master equation and Bayesian theory, but only with considerable effort, and only if their statistical properties are fully characterized [104–106]. There is, therefore, a need for methods that can effectively distinguish different physical cases without relying on assumptions on physical parameters that cannot be certified. Based on the same principles as the neural network recognition of hand written characters, our machine learning approach assumes no particular underlying theory or parameter values; it only assumes that the training sets are representative of the signals that will later be presented to the algorithm.

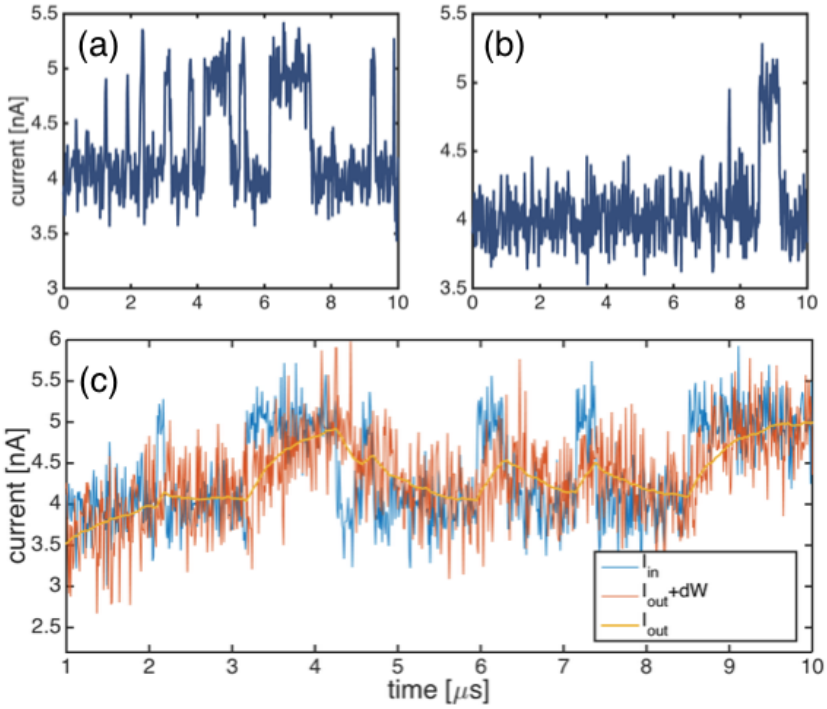

Figure 5.2: Panels (a) and (b) show examples of Markovian QPC current with Rabi frequencies and tunneling rates, $\Omega = \gamma_\uparrow = 5$ MHz and $\gamma_\downarrow = 4$ MHz for panel (a) and $\Omega = 4$ MHz, $\gamma_{\uparrow,\downarrow} = 1$ MHz for panel (b). The panel (c) shows the application of the non-Markovian filters on the Markovian current of the type shown in (a) and (b).

To illustrate the flexibility of machine learning techniques for parameter estimation of quantum systems, we consider the continuous monitoring of a single electron that may tunnel on and off a quantum dot as considered in Chapter 3 [4, 46]. The energy splitting between the spin-up and spin-down states can be tuned by a magnetic field, such that only a spin-down electron may tunnel into the dot. When an electron occupies the dot, further charging of the dot is prevented by the Coulomb blockade. Similarly, the electron may only tunnel off the quantum dot when it occupies the spin-up state. In addition to this incoherent tunneling on and off the dot, governed by the tunneling rates $\gamma_\downarrow$ and $\gamma_\uparrow$, respectively, the spin may

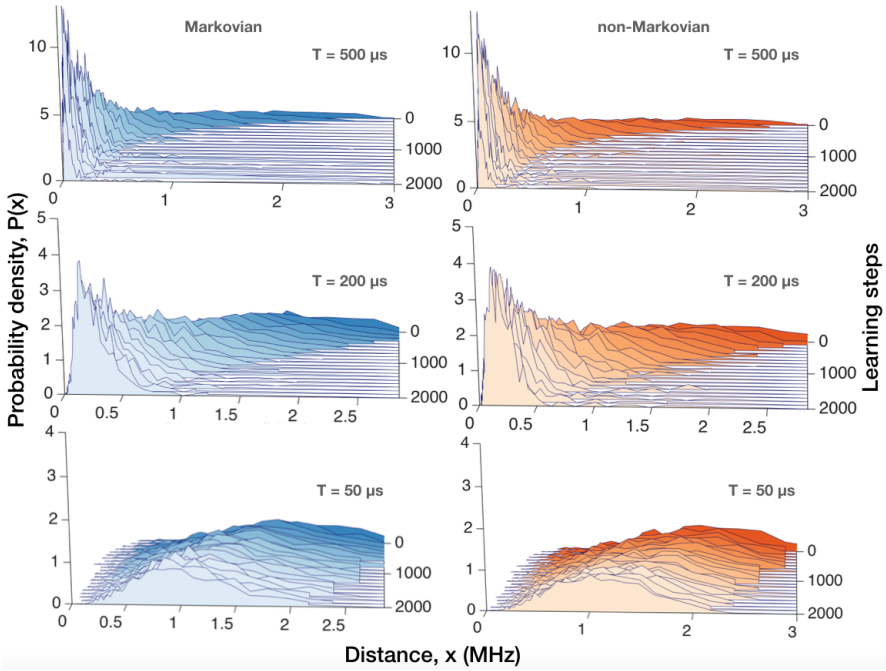

Figure 5.3: Probability density of the distance from the correct value of all parameters given for different number of learning steps. The upper (lower) panel shows the first 2000 learning step of learning based on simulated experimental currents from a Markovian (non-Markovian) data.

precess between the spin states with a Rabi frequency Ω when an external resonant drive is applied. The electron is, thus, subject to both incoherent and coherent dynamics and is simultaneously monitored by a nearby QPC, which permits a noisy transmission current with a mean rate that depends on the charge on the quantum dot. The QPC current is insensitive to the spin state of the electron [12, 47]. The left hand side of Fig. 5.1 and the panels (a), (b) in Fig. 5.2 show examples of simulated QPC current data and clearly witness periods of low and high transmission, pertaining to the occupied and unoccupied states of the quantum dot. From the data shown, it becomes evident that one may apply a statistical analysis of the duration of low and high transmission intervals and, thus, obtain an estimate of the tunneling rates and the Rabi frequency. Assuming the validity of the stochastic master equation and the resulting Bayesian filter, an effective

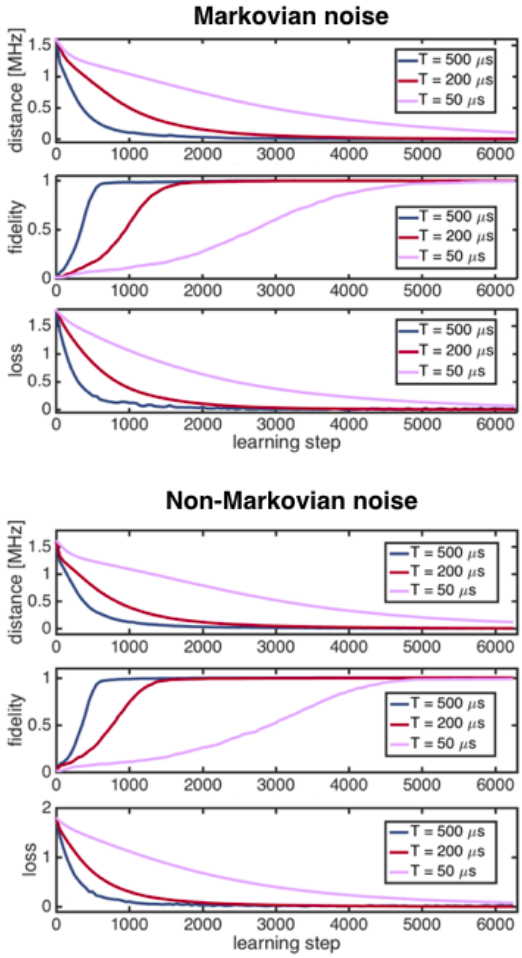

Figure 5.4: Quantities characterizing learning success for three different lenghts of the time trace. We show average distance from the correct parameter, fidelity, and value of the loss function as a function of learning steps for Markovian (non-Markovian) dynamics in the upper (lower) panel.

and precise estimation of the Rabi frequency and the tunneling rates from the QPC current was demonstrated in Chapter 3, and is possible even when one cannot clearly discern low and high transmission intervals due to the signal noise, as discussed in section 3.6.

To study the achievements of machine learning for the estimation of

parameters leading to QPC signals as the ones shown in the figures, we synthesize many such signals, which we then treat as 1D image inputs for the convolutional neural network [77, 83] (CNN) as schematically shown in Fig. 5.1. To understand why image recognition is a suitable approach to parameter estimation consider a standard case for image recognition problems namely the recognition of hand-written images of digits between 0 and 9 [80–82]. A CNN, when trained, can assign any hand-written input to one of the 10 basic classes. In our case, the classes are chosen as different combinations of parameters governing the dynamics of an experiment. Hand-writing from thousands of people include inherent randomness which the CNN learns to ignore and, similarly, the CNN learns to disregard irrelevant experimental noise fluctuations and extract only the features that distinguish the experimental parameters.

The neural network used in this work is shown in Fig. 5.1 and consists of convolution layers that perform a discrete convolution with a filter function, pooling layers that down-sample the data and dense fully connected layers that recognizes finer structure in the input data. The functionality and construction of these layers are described in detail in Appendix E. The weights of the convolution filters and the fully connected layers define the connections between the neurons in the consecutive layers and are the parameters that are adjusted during the training of the CNN so as to minimize the so-called loss function (also called the cost function) [78, 107]. For classification problems that assign a probability to all candidate values, the cost function, as also discussed in Chapter 4 is of the form $H(y^{target}, y^{output}) = -\sum_j y_j^{target} \ln y_j^{output}$, where y^{output} are the output probabilities of the network, and y^{target} are the expected (correct) outputs. During the training, we re-evaluate the weights of the connection between the neurons in the network in order to minimize the distance between the output and the correct values [81, 107].

When the training is completed, we test the network to validate how well it performs on new signals. The learning and the validation may be done with purely experimental data, but in this work, both are numerically simulated. As an example, we consider 6 candidate values for each of the parameters: the Rabi frequency Ω (equidistantly distributed between 4 MHz and 10 MHz) and the tunneling rates on and off the quantum dot $\gamma_{\uparrow,\downarrow}$ (equidistantly distributed between 1 MHz and 6 MHz). This gives us $6^3 = 216$ classes for which an experimental current can be assigned to. In the upper panel of Fig. 5.3, the training of the network is visualized by the probability density of the average distance between the correct underlying parameters and the output parameters of the machine learning algorithm

as a function of the number of learning steps. The numerical effort of the learning process depends on the total length, T, of the measurement record, and we considered signals of different duration, 50μs, 200μs, 500μs, noting that the shorter intervals provide less information as witnessed by the broader range of estimation errors. Fig. 5.3 shows how the length of the trace influences the number of training sets needed for convergence towards the correct value of all parameters as a function of the number of learning steps (for the first 2000 learning steps). When the network is trained, we perform validation on new signal currents (non-overlapping with the training data) and evaluate quantities that shows us the performance of the neural network as a function of the number of learning steps. In the three upper panels of Fig. 4 we show the average distance between the predicted parameters and the correct values, the fidelity (or accuracy) defined as the probability of predicting the right parameter and the loss function each as a function of the number of the learning steps (for details see Appendix E). Even though the convergence rate differs depending on the length of the time trace in all cases the CNN recognizes all parameters with more than 99% fidelity and the average distance from the correct value is on the order of 10^{-3} MHz for the longer time traces and 10^{-2} MHz for 50 μs (see Appendix E for precise values). While for the first 2000 training steps the distribution of distances is quite broad (as shown in Fig. 5.3), we see in Fig. 5.4, that (albeit more slowly) the algorithm correctly learns to distinguish the values perfectly for all lengths of the time trace. Further technical and quantitative details of our CNN simulation are described in detail in the Appendix E.

5.3 Realistic experimental noise and distortion

So far, we considered a system amenable to a stochastic master equation analysis, but it is easy to extend our approach to realistic situations where a stochastic master equation does not apply. Assume for example that the QPC signal is modified by a convolution with a filter function, $I_{out}(t) = (I_{in} \otimes u)(t) = \int_{-\infty}^{\infty} u(t')I_{in}(t-t')dt'$ where $I_{in\,(out)}$ is the ideal (actually detected) current. Due to the temporal convolution, we shall refer to the signal as non-Markovian. Moreover, we model the inclusion of further amplifier noise by adding an additional white noise dW to $I_{out}(t)$ as shown in Fig. 5.2 (c). It is apparent that the step structure of the ideal current is now masked by the finite bandwidth of the filter together with the added random noise. Still, we may expect that the neural network can

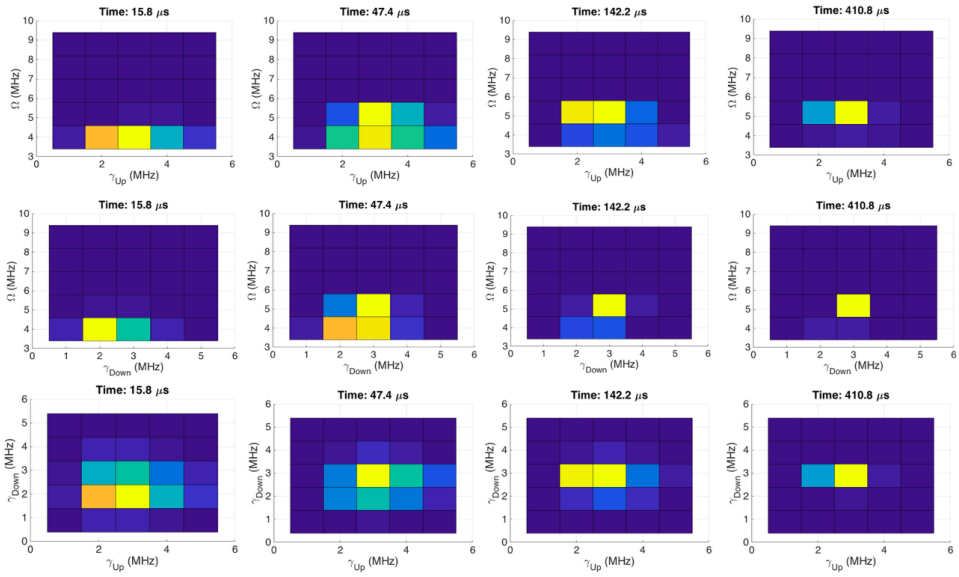

Figure 5.5: Likelihood function projected onto two-dimensional subspaces shown for different lengths of the time trace. The correct parameters are $\{\Omega, \gamma_\uparrow, \gamma_\downarrow\} = \{5.2, 3, 3\}$ MHz for this trajectory.

identify characteristic, albeit less intuitive, differences between the signals pertaining to the different parameter sets.

The CNN in Fig. 5.1 is now trained on the non-Markovian currents in the same way as described above, and it is interesting to note the close similarity of the performance of the system in the Markovian and the non-Markovian case, cf. the upper and lower panels in Fig. 5.3. Also, the average distance, the fidelity and the loss functions, depicted in the lower panel of Fig. 5.4 for the non-Markovian currents follow the same evolution as for the Markovian case. Despite a slight decrease of the speed of convergence in the non-Markovian cases, it is clear that qualitatively, the neural network treats the two problems with similar success. Thus, while no practical analytical theory exist to describe parameter estimation from the non-Markovian current signals, the machine learning approach allows us nevertheless to estimate the parameters with high accuracy.

5.4 Comparison with Bayesian approach

For the case where the Markovian approximation holds, the underlying parameters governing the quantum dynamics can be estimated using Bayesian methods. For the situation considered in the main text, this leads to the task of solving a stochastic master equation for 216 sets of parameters for each trajectory in order to estimate the most likely parameter, as discussed in Chapter 2 and Chapter 3.

The result of the master equation simulations is a likelihood function $L_t(i, j, k)$ indicating the likelihood at time t for the candidate values given by the indices's i, j and k corresponding to Ω, $\gamma_\uparrow$ and $\gamma_\downarrow$ respectively. Normalizing the likelihood function at each time t gives the probability distribution of the three variables, $P_t(\Omega, \gamma_\uparrow, \gamma_\downarrow)$. In Fig. 5.5 we show the marginal distributions of P_t at 4 different times for a single trajectory. The most likely set of parameters converges as expected to the correct underlying set of parameters.

5.5 Outlook and conclusion

In summary, we proposed to use training of neural networks for the purpose of parameter estimation. We showed that this method is capable of dealing with uncertainties in, or complete absence of, models describing the experimental noise during measurements. In particular, our neural network approach translates a quantum parameter estimation problem into an image classification problem and therefore does not suffer from issues related to the characterization of quantum and classical noise.

We imagine that our machine learning approach may have further applications, e.g., in connection with the use of the quantum systems for quantum communication and computing, where certain operations conditioned on measurement outcomes, feedback and error correction may be decided and executed based on the neutral network. Another daunting perspective is the application of a quantum processor for the machine learning process itself which, apart from offering a potential speed-up, [86, 87, 108, 109], may implement an almost autonomous quantum control and feedback systems with entangled system and observer.

Chapter 6

Applying smoothing techniques to quantum algorithms: Quantum teleportation

6.1 Introduction

In this Chapter we look into applications of the techniques explained in the previous Chapters in the context of quantum algorithms. We discussed how smoothing and neural network methods can be of help in the context of monitoring of experimental quantum systems. Let us now bring this logic into an algorithmic setting. Quantum teleporation is a notorious example of an entanglement-assisted quantum protocol, and it is also an example of a case when classical information is used for feedback or feed-forward as an inherent part of the protocol. We can therefore implement post-processing techniques (both master equation and machine learning based) to decrease the likelihood of en error in the feed-forward part of the protocol.

Sections 6.2-6.6, Section 6.8 and parts this introduction are taken over from [3] and [6]. The application of neural network for fidelity improvement of the protocol presented in Section 6.7 presents our latest currently unpublished results.

This Chapter is organized as follows: In Sec. 6.3 we describe our physical model and teleportation scheme. In Sec. 6.4 we simulate the dynamics during continuous probing of the system and we present a simple

signal analysis for the teleportation protocol. In Sec. 6.5 we apply the past quantum state analysis to (simulated) measurement data, and we show that we can obtain better teleportation fidelities from this approach. In Sec. 6.7 we compare past quantum state-based approach with machine learning method explained in Chapters 4 and 5. We present our conclusions and an outlook in Sec. 6.8.

6.2 Quantum teleportation

Quantum teleportation [110] is a protocol allowing the application of non-local quantum superposition states in quantum information [49], computation [111], and cryptography [112]. Quantum teleportation has been experimentally demonstrated in a number of systems [113–119], and it allows two parties, A and B, who share a maximally entangled state, to 'teleport' an unknown quantum state, $|\psi\rangle$, from the location of A to the location of B using only local operations and classical communication. Quantum teleportation is also refereed to as disembodied transfer, emphasizing that no properties of the teleported quantum state are at any time detectable in the spatial region between the locations A and B.

We propose a scheme for teleportation between qubits, where we use a cavity field as the communication channel, see Fig. 1. After preparation of an entangled state of the cavity field and qubit B, the qubit is detuned away from the cavity resonance and no longer interacts with the field. A measurement of the joint state of the field and the unknown state $|\psi\rangle$ of qubit A, now leads to a random outcome and an accompanying measurement back action on qubit B. Neither the measurement outcome nor the projected state of qubit B reveal any property of $|\psi\rangle$. But, the measurement outcome can be communicated classically to select and implement a local unitary on qubit B which finally prepares it in the state $|\psi\rangle$.

So far, the description of our proposal follows the protocols applied in teleportation experiments between two ions, coupled sequentially to a third ion [115, 116] and between superconducting qubits, coupled sequentially to a third superconducting qubit [118], with the main differences being that the communication channel is a different physical system, extending over both locations A and B (but decoupled by frequency detuning of the components).

While a cavity field offers continuous variable teleportation [112], to teleport the discrete states of qubits we employ the so-called hybrid quantum

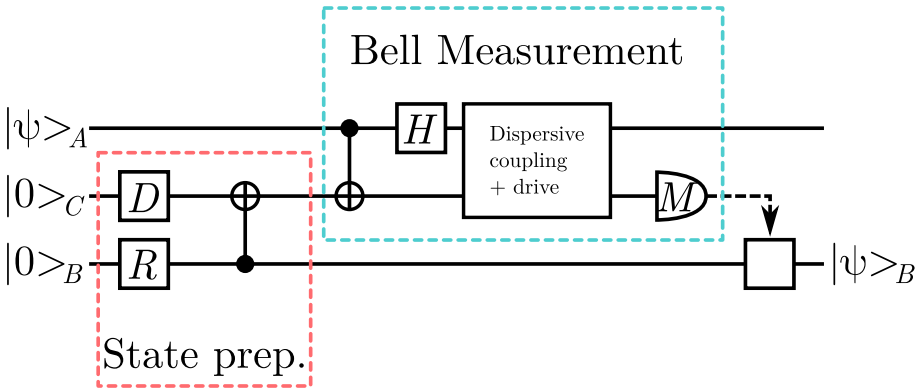

Figure 6.1: Quantum circuit implementing teleportation of an unknown state of a qubit A to another physical qubit B. System C is a harmonic oscillator, initially entangled with qubit B: a B qubit superposition state, prepared by the $\pi/2$-rotation R, interacts dispersively with a coherent state $|\beta\rangle_C$, prepared by the displacement operator D. The symbol in the figure indicates that resulting, conditional change of phase of the coherent state amplitude to $|-\beta\rangle_C$ is equivalent to C-NOT gate on the effective oscillator qubit states. A Bell measurement in the discrete basis of A qubit states and coherent states $|\pm\beta\rangle_C$ is accomplished by a dispersive interaction between the systems followed by probing the field leaking from the cavity and then dispersively probing the state of the qubit A in a rotated (by H) basis. Finally qubit B is subject to a unitary operation depending on the outcome of the Bell measurement. The protocol is explained in detail in the text.

teleportation schemes [120] where a qubit degree of freedom is associated with a pair of coherent states of the field. For the initial entanglement generation we use that a moderately detuned qubit causes a qubit-state dependent frequency shift of the cavity, equivalent to a frequency shift arising from a Kerr effect [121, 122]. While a Bell state measurement conventionally involves projective measurements on both subsystems, our procedure consist of sequential steps where we only probe the signal leaking from the cavity by homodyne detection [123]. The restriction of our readout mechanism to continuous homodyne probing rather than instantaneous projective measurements requires a signal analysis to infer the optimal local operation on the isolated target qubit B. In this article we compare the direct application of the accumulated signal as an approximate measurement of the state of the system to a Bayesian analysis which attempts to retrodict from the whole signal record what would have been the most likely outcome of a projective measurement at the beginning of the probing sequence. Such inference problems, accounting for our knowledge about the state of a quantum system at a (past) time t are of the type addressed by the past quantum state (PQS) formalism [31] as explained in Chapter 2.

We note that our use of retrodiction in connection with teleportation is related to delayed-choice experiments and to implementation of entanglement swapping protocols, where the heralding of a certain entangled state was only done after the state had already been detected and consumed [124, 125]. In this, article however, we do not aim to herald past states, but we shall rather use the past quantum state formalism, to infer which local operation may best accomplish the teleportation protocol. We demonstrate how our protocol is implemented in circuit QED, where cavities and qubits are constructed by superconducting waveguides and Josephson junctions respectively [126, 127], but the analysis is general and can in principle be implemented in any system where the readout out of one or some of the components occurs sequentially and simultaneously with unitary or dissipative dynamics. The improvement over a more straightforward measurement and feed-forward protocol is the main result of this work.

6.3 The protocol

The protocol that we are proposing is depicted in Fig. 6.1. We consider two qubits and one cavity. At the beginning, one of the qubits, A, is in the unknown state, $|\psi\rangle_A$, that we intend to teleport. The second qubit, B, and the cavity are in the qubit ground state $|0\rangle_B$ and the oscillator ground

state $|0\rangle_C$. We use the subscripts A, B, C for qubit A, qubit B and cavity respectively.

The teleportation protocol consists of three steps [110]. First we create an entangled state between the qubit B and the cavity (a qubit-cavity Bell state) [128–130]. Then, we perform a Bell state measurement of A and C, involving an entangling gate operation and probing of the field amplitude, followed by a field measurement that is sensitive to the qubit state. The third and final step is the application of a unitary operation on the qubit B chosen according to the outcome of the field measurements. Ideally, the combined protocol returns qubit B in the desired state $|\psi\rangle_B$.

We encode a qubit degree of freedom in a pair of coherent states $|\pm\beta\rangle_C$ of the cavity oscillator, and we prepare a Bell state between B and C by first exciting the oscillator coherent state $|\beta\rangle_C$ and a qubit B superposition state: $(\sqrt{2})^{-1}(|0\rangle_B + |1\rangle_B)$, followed by a dispersive interaction between the two systems which implements a qubit-controlled phase shift of the coherent state amplitude,

$$\begin{aligned}|\phi\rangle_{BC} &= U_{BC}\frac{1}{\sqrt{2}}(|0\rangle_B \ + |1\rangle_B)|\beta\rangle_C \\ &= \frac{1}{\sqrt{2}}(|0\rangle_B \otimes |\beta\rangle_C + |1\rangle_B \otimes |-\beta\rangle_C).\end{aligned}$$

The operation U_{BC} is equivalent to a C-NOT gate on the effective qubit subspace of the oscillator spanned by $|\pm\beta\rangle_C$ and the resulting state is equivalent to a two-qubit Bell state in the limit of a vanishing overlap of the two coherent states, $\langle -\beta|\beta\rangle = \exp(-2|\beta|^2) \approx 0$. We shall apply $\beta \simeq 2$ in our simulations resulting in a non-zero overlap and a correspondingly reduced fidelity of our protocol. We emphasize that after the state, $|\phi\rangle_{BC}$, is generated, qubit B is not interacting with the cavity any further and the protocol is formally equivalent to true teleportation among spatially separated quantum systems.

The full system is now in a product state of the unknown qubit state $|\psi\rangle_A = \alpha_0|0\rangle_A + \alpha_1|1\rangle_A$, and the entangled Bell-like state of the qubit B and the cavity C, $|\phi\rangle_{BC}$, and following the teleportation procedure a Bell state measurement is performed on qubit A and the cavity C. The Bell measurements can be decomposed into the application of a C-NOT gate and a qubit A rotation, followed by qubit projective measurements on the

separate systems A and C. By expanding the state:

$$\begin{aligned} H_A U_{AC}|\psi\rangle_A|\phi\rangle_{BC} = \frac{1}{2}[&|0\rangle_A(\alpha_0|0\rangle_B + \alpha_1|1\rangle_B)|\beta\rangle_C \\ &+ |0\rangle_A(\alpha_0|1\rangle_B + \alpha_1|0\rangle_B)|-\beta\rangle_C \\ &+ |1\rangle_A(\alpha_0|0\rangle_B - \alpha_1|1\rangle_B)|\beta\rangle_C \\ &+ |1\rangle_A(\alpha_0|1\rangle_B - \alpha_1|0\rangle_B)|-\beta\rangle_C]. \end{aligned} \tag{6.1}$$

We see, that depending on the measurement on qubit A and the cavity, the conditional state of qubit B acquires the unknown state amplitudes of the input state, and we can immediately read out the conditional mapping between the qubit B output state and the input state:

$$\begin{aligned} 0_A, \beta_C &\mapsto \alpha_0|0\rangle_B + \alpha_1|1\rangle_B = |\psi\rangle_B \\ 0_A, -\beta_C &\mapsto \alpha_0|1\rangle_B + \alpha_1|0\rangle_B = \sigma_B^x|\psi\rangle_B \\ 1_A, \beta_C &\mapsto \alpha_0|0\rangle_B - \alpha_1|1\rangle_B = \sigma_B^z|\psi\rangle_B \\ 1_A, -\beta_C &\mapsto \alpha_0|1\rangle_B - \alpha_1|0\rangle_B = -i\sigma_B^y|\psi\rangle_B, \end{aligned} \tag{6.2}$$

expressed in terms of the single-qubit Pauli operators $\sigma^x, \sigma^y, \sigma^z$.

The four Bell state measurement outcomes (6.2) occur with equal probabilities, independently of the input state, and they hence reveal no information about the teleported state. In the experiment, one merely has to perform the measurements, and the combination of the measurement back action on qubit B due to the entanglement with the cavity and the conditional unitary operation, cf., (6.2) should ensure the correct state transfer.

Now, in the architecture, presented in the next section, we restrict ourselves to continuous homodyne detection of the output signal from the resonator. Such measurement can be used to first distinguish the coherent states, $|\pm\beta\rangle$, as the field leaks from the cavity and, subsequently, we can drive the cavity with a resonant input field, which undergoes a phase shift due to the interaction with qubit A and which hence permits detection of the qubit state [123]. Both measurements are noisy and take time, but since qubit B is not addressed during the probing, we assume that the optimal unitary operation on B after the measurement is complete, is the one pertaining to our best estimate of which Bell state was occupied at the initial time of the measurement ($t = 0$ in Fig. 6.2). To assess the fidelity of our protocol, we choose random input states, simulate the protocol and, in particular, the probing, and we check how well our different strategies succeed on average in teleporting the state. Note, however, that since

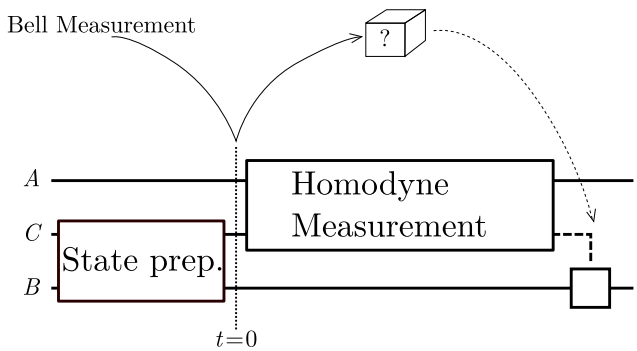

Figure 6.2: Schematic representation of the time dependent measurement procedure. Our implementation uses continuous homodyne observation of the field emitted by the cavity, and after preparation of the entangled and input states and the controlled unitary operations on the input qubit and the cavity, we read out first the leaking cavity field amplitude and, subsequently, we drive the cavity and use the dispersive interaction to probe the state of qubit A. Both measurements are subject to noise and they take a finite duration, and in the text we compare two analyses of the signals leading to different choices for the optimal operation on the undisturbed target qubit B: a direct approach based merely on the integrated signals, and an approach using the past quantum state formalism to retrodict what would have been the most likely projective Bell measurement outcome at the earlier time $t = 0$.

our strategies should not invoke any knowledge of the input state, we are not allowed to use any properties of the conditioned quantum state in the choice of operations on the system. We shall thus design the procedure to depend only on the simulated signal (equivalent to an actual measured signal in an experiment).

There has been a number of successful proposals for tests and analyses of the teleportation protocol based on the study of the density matrix describing the system as a whole, i.e. including the entangled pair and the state to be teleported [131–134]. The evaluation of a forward evolution of a quantum master equation, however, requires prior knowledge of the initial state, which is not known. This can be compensated by averaging

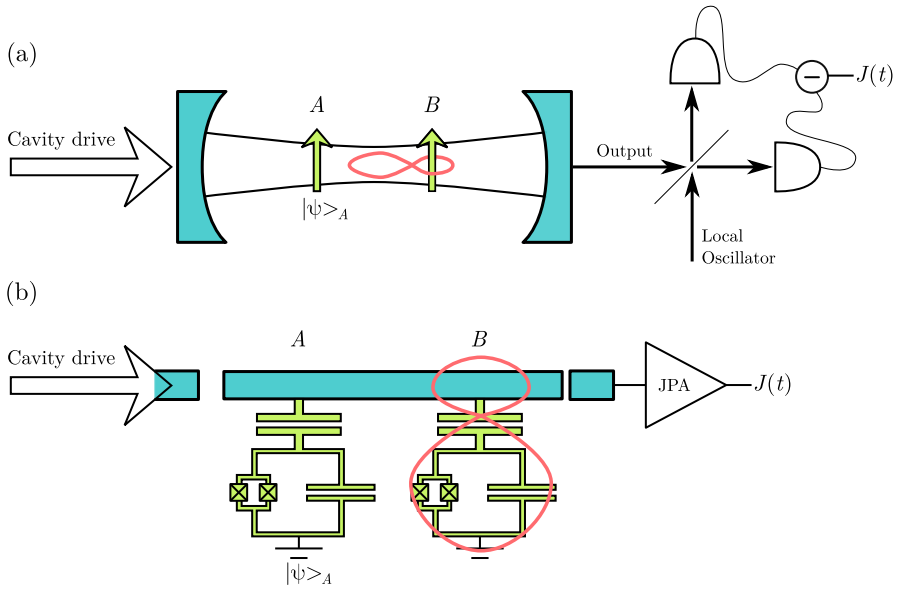

Figure 6.3: Schematic presentation of the experimental setup. Panel (a) depicts a quantum optics rendering of the experiment with optical homodyne detection, while panel (b) shows the scheme for superconducting circuits with two transmon qubits, one resonator cavity and a Josephson parametric amplifier (JPA) for amplification and measuerement of a field qudrature variable.

over all possible initial states or by adding an ancillary degree of freedom [135]. Our approach, on the other hand, does not presuppose anything about the state to be teleported and is experimentally reproducible for any state $|\psi\rangle_A$.

6.4 Implementation and simulation

We shall implement the protocol with two qubits and a cavity field, illustrated in a quantum optics and in a circuit QED schematic in Fig. 6.3. A two-level system coupled to a cavity field is described by the

Jaynes-Cummings Hamiltonian,

$$H = \omega_r a^\dagger a + \frac{\Omega_q}{2}\sigma_z + g(a^\dagger \sigma_- + a\sigma_+), \tag{6.3}$$

where $a(a^\dagger)$ is the annihilation (creation) operator for the field in the cavity, $\sigma_{-(+)}$ is the lowering (raising) operator for the two-level system, ω_r is cavity frequency, Ω_q is the qubit frequency and g is the coupling strength. This Hamiltonian has been demonstrated with numerous effective two-level systems coupling to cavity fields, e.g. atoms [136], quantum dots [137] and in particular superconducting qubits in circuit QED [126, 127]. In the regime where the detuning between the cavity and qubit frequencies, $\Delta = \Omega_q - \omega_r$, is much larger than the coupling strength, g, we arrive at the dispersive Hamiltonian

$$H = \omega_r a^\dagger a + \frac{\Omega_q}{2}\sigma_z + \chi\, \sigma_z a^\dagger a, \tag{6.4}$$

with the dispersive coupling $\chi = g^2/\Delta$. In circuit QED, the two-level systems are superconducting qubits and the cavity field is the microwave field of a coplanar waveguide or the field within a 3D microwave cavity. Circuit QED also makes it possible to tune the frequency of qubits and thereby the dispersive coupling χ and similarly a transmon with tunable coupling, g, can be constructed [138, 139]. In the relevant regime of transmon qubits [140], χ, attains a form similar to the result for general two-level systems (see [3]).

Considering the scheme presented in Sec. 6.3 we have two qubits dispersively coupled to the cavity field. We initialize the system such that [3] $\chi_A = \chi_B \approx 0$. A coherent displacement of the cavity field is achieved by applying a coherent drive, $H_d = \epsilon_d(t)a^\dagger + \epsilon_d^*(t)a$, to the cavity. Similarly, a single qubit rotation can prepare the qubit in an arbitrary superposition state, as required for the input state and for the protocol. Tuning qubit B close to resonance with the cavity yields a non-zero dispersive coupling, $\chi_B > 0$. This results in the application of a phase gate, U_{BC}, when the dispersive interaction is active for a time $\tau = \pi/\chi_B$ [130] (details presented in [3]), and we obtain the Bell-state required for the teleportation protocol. The same gate is equivalently applied for qubit A later in the protocol, where qubit B is far-detuned from the cavity and, thus, well-separated, albeit in frequency space rather than in physical space.

For the readout we leave both qubits far-detuned from the cavity, and we monitor the signal leaking from the cavity using a quantum-limited parametric amplifier [141–144] which performs a highly efficient homodyne

detection. As we are not directly measuring the intra cavity field quadrature $X_c = a + a^\dagger$ but the radiation emitted over time, we describe the dynamical evolution of the homodyne measurement by the stochastic master equation as discussed in Chapter 2 [19]

$$\begin{aligned} d\rho = &-i[H,\rho]\,dt + \kappa\Big(a\rho a^\dagger - \frac{1}{2}(a^\dagger a\rho + \rho a^\dagger a)\Big)dt \\ &+ \sqrt{\kappa\eta}\Big(a\rho + \rho a^\dagger - \mathrm{Tr}(X_c\rho)\,\rho\Big)\,dW(t), \end{aligned} \tag{6.5}$$

with κ being the linewidth of the cavity and $dW(t)$ being a stochastic Gaussian process with $\langle dW^2\rangle = dt$ and $\langle dW\rangle = 0$ which represents the stochastic part of the homodyne measurement signal

$$J(t) = \sqrt{\kappa\eta}\,\mathrm{Tr}(X_c\rho) + \frac{dW(t)}{dt}, \tag{6.6}$$

which directly corresponds to the experimentally available current. The stochastic part is zero in the mean and averaging $J(t)$ over time yields the quantity

$$S_\beta = \mathrm{sign}\Big(\int_0^{T_\beta} J(t)\,dt\Big). \tag{6.7}$$

which can be used to distinguish the two states $|\beta\rangle$ and $|-\beta\rangle$ as they lead to mean amplitude signals with opposite sign. The integral (6.7) includes a noise component which is dominant for short integration times, and the duration T_β, devoted to distinguish $\pm\beta$ should be chosen on the order of the cavity lifetime κ^{-1}, after which the signal has decayed.

As a next step, the frequency of qubit A is tuned close to the cavity resonance and a non-zero dispersive shift is obtained, that we can probe via the phase of a transmitted coherent drive with constant amplitude ϵ_r. The mean-field coherent state amplitude of the cavity field now follows the equation

$$\frac{d}{dt}\beta = -i\chi_A\langle\sigma_z\rangle\beta - \frac{\kappa}{2}\beta - i\epsilon_r, \tag{6.8}$$

where χ_A is the dispersive shift. The steady state solution ($d\beta_{ss}/dt = 0$) yields

$$\beta_{ss} = \frac{-i\,\epsilon_r\kappa - 2\epsilon_r\chi_A\langle\sigma_z\rangle}{\frac{\kappa^2}{2} + 2\chi_A^2\langle\sigma_z\rangle^2} \tag{6.9}$$

with the real part, probed by the JPA,

$$\mathrm{Re}(\beta_{ss}) = \frac{-2\epsilon_r \chi_A \langle \sigma_z \rangle}{\frac{\kappa^2}{2} + 2\chi_A^2 \langle \sigma_z \rangle^2}. \tag{6.10}$$

The sign of the average integrated signal is in steady state governed by the sign of $\langle \sigma_z \rangle$, i.e., the qubit state. To avoid transient contributions without a definite sign to the signal, we accumulate the probe signal after a finite waiting time T_w, that we shall identify by numerical optimization,

$$S_A = \mathrm{sign}\Big(\int_{T_\beta + T_w}^{T_\beta + T_m} J(t)\, dt \Big). \tag{6.11}$$

A negative sign of this integrated current implies the excited state qubit $|1\rangle$, while a positive sign corresponds to the ground state $|0\rangle$. The measurement may be in error due to the Gaussian noise contribution to the integrated signal, but the readout is a quantum non-demolition measurement, and increasing T_m will increase the measurement fidelity until we approach the qubit lifetime.

Knowing the time dependent mean field envelopes during the two probing periods allows accumulation of the signal with a time dependent weight factor. We have implemented such weighing, but for our parameters we did not see a significant change of the resulting teleportation fidelity, beyond what we achieve by optimizing the parameters T_β, T_m and T_w.

6.5 Past quantum state simulation

The past quantum state (PQS) formalism is a generalization of classical smoothing algorithms and of the so-called forward-backward analysis of hidden Markov models [37, 145] to the case of quantum states, as discussed in Chapter 2. The aim of the formalism is to infer from measurements on quantum system before and after a given time t what would have been the outcome of an arbitrary measurement if carried out at time t. It may also be viewed as a generalization of the Aharonov, Bergmann and Leibowitz (ABL) rule, for projective measurements performed between the preparation and post-selection of a quantum system in arbitrary pure states [146], and of the theory of weak value measurements for the post selected average of weak, non-projective measurements [147]. See also Appendix B for how the concepts such as 3-box paradox arise within past quantum state formalism. It exactly reproduces the results of these theories, but for our case of dynamically evolving and continuously monitored quantum

systems we shall need the more general theory of [31] that we presented in Chapter 2.

This density matrix, $\rho(t)$, yields the probability for the outcome of any measurement on the system as represented most generally by positive operator valued measures [49] (POVMs), $\{M_i\}$, with $\sum_i M_i^\dagger M_i = \mathbb{1}$, for which outcome i occurs with the probability, $p_\rho(i) = \mathrm{Tr}(M_i \rho M_i^\dagger)$.

The further knowledge due to later measurements is represented by another matrix $E(t)$, which together with $\rho(t)$ yields the POVM outcome probabilities, conditioned on all, earlier and later measurements on the system [1, 31],

$$p(i) = \frac{\mathrm{Tr}(M_i \rho M_i^\dagger E)}{\sum_i \mathrm{Tr}(M_i \rho M_i^\dagger E)}. \tag{6.12}$$

In practical calculations we find $E(t)$ by solving a stochastic master equation backwards in time from the last moment, T, of measurements, where $E(T) = \mathbb{1}$. The master equation for E is the adjoint to the master equation for $\rho(t)$, and it involves the same measurement current signal $J(t)$ as applied for the calculation of $\rho(t)$. While $\rho(t)$ only depends on $J(t')$ for $t' \leq t$, $E(t)$ depends on $J(t')$ for $t' \geq t$. From Eq. (6.5), we thus obtain

$$\begin{aligned} dE = {} & i[H, E]\, dt + \kappa\Big(a^\dagger E a - \frac{1}{2}(a^\dagger a E + E a^\dagger a)\Big) dt \\ & + \sqrt{\eta\kappa} J(t) \Big(a^\dagger E + E a\Big)\, dt. \end{aligned} \tag{6.13}$$

Note that a c-number term is missing compared to Eq.(6.5). This will only affect a common factor on E, but since the adjoint master equation is not trace preserving anyway, and since Eq. (6.12) explicitly renormalizes the probabilities, this does not affect retrodictions made by the theory.

Our aim is to infer which action to apply on the target qubit. Under ideal circumstances this action is inferred from a projective Bell state measurement at time $t = 0$, but we are not able to perform that measurement. As an alternative candidate procedure, we shall employ the past quantum state formalism to infer from our continuous measurements until the later time $T = T_\beta + T_m$ what would have been the most likely outcome of a projective Bell measurement, if it had taken place at $t = 0$.

Without knowing the input state $|\psi\rangle_A$, our prior knowledge about the outcome is equivalent to a fully mixed density matrix $\rho(t = 0) \propto \mathbb{1}/4$ on the tensor product state space of $|0\rangle_A$, $|1\rangle_A$ and $|\beta\rangle_C$, $|-\beta\rangle_C$. We calculate $E(0)$ by solving the backward evolution Eq. (6.13) from $T = T_\beta + T_m$ to 0 (note that we need only to solve this equation for the cavity and

qubit A, as qubit B is a passive spectator under the homodyne detection sequences). Finally, to retrodict the most likely Bell state outcome, the gates are performed according to Fig. 6.1 and then we need the outcome of a qubit A projective measurement, and a measurement of the sign of the intra cavity field quadrature (see Eq. (6.2)).

This measurement is described by the POVM

$$M_i = |i\rangle_A |X\rangle_{CC}\langle X|_A\langle i|, \tag{6.14}$$

where $X = \pm\beta$ and $i = 0, 1$. This, together with the backward evolution (6.13), gives us the tools for evaluating the generalized Born rule (6.12) at the time zero and therefore calculate the outcome probabilities for the 4 different outcomes of the Bell measurement (6.14).

Having used the PQS formalism to retrodict what would have been the most likely Bell state outcome, in every run of the experiment (simulation), we choose to apply the corresponding unitary corrective operation on qubit B as in Eq. (6.2). That state, in turn, can be directly compared with the randomly chosen input state, and we can determine the average fidelity of the protocol as function of the physical parameters.

6.6 The results

We carry out numerical simulations for a cavity with a damping rate κ, initially prepared in a coherent state with $\beta = 2$. We assume a dispersive interaction with $\chi = 13.5\kappa$, and a coherent drive of $\epsilon_r = 2\chi$ during the qubit readout. We measure for a time $T = T_\beta + T_m$ and we fix $T_\beta/T = 0.4$. To represent real experiments, we also perform simulations with different values of the detector efficiency. With these parameters, the signal-to-noise ratio is expected to only exceed unity for T larger than $1/\kappa$ [148]. The forward evolution of the conditioned master equation, Eq. (6.5), yields the state of the entire system, and after application of the unitary operation σ_B, chosen according to Eq. (6.2), we obtain the reduced density matrix of the target qubit, $\rho_B = \mathrm{Tr}_{AC}[\sigma_B^i \rho(T) \sigma_B^i]$. The overlap, ${}_A\langle\psi|\rho_B|\psi\rangle_A$, yields the fidelity of the final state and we estimate the protocol fidelity as the average over the state fidelities [131, 132] by choosing 500 random input states from a uniform distribution over the Haar measure [149]. The results of such simulations are shown as function of the detector efficiency in Fig. 6.4 and as function of the probing time in Fig. 6.5, where the red dashed lines show the fidelity of the procedure when σ_B is chosen from the values of the integrated homodyne signals and the solid blue lines show the result

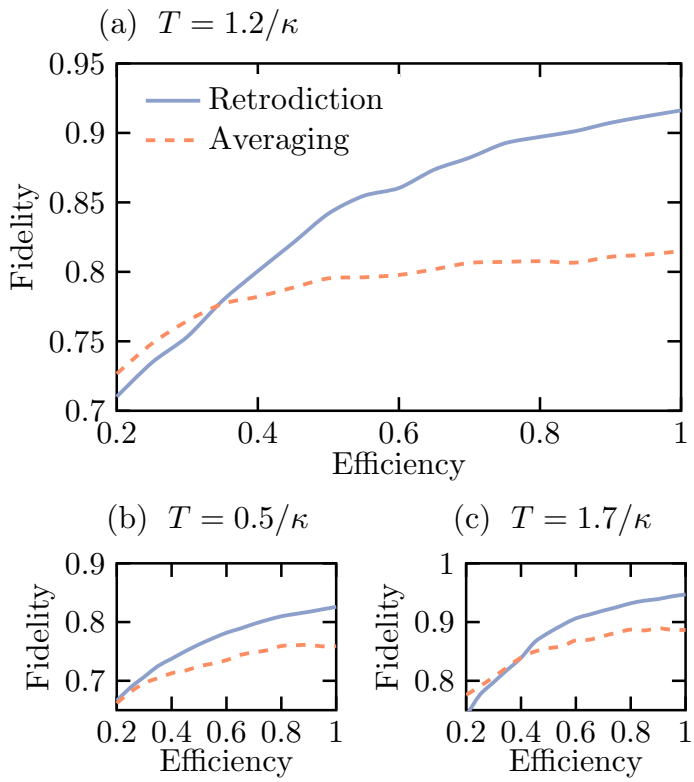

Figure 6.4: The fidelity of the protocol as a function of efficiency pre- and post the past quantum state retrodiction is applied. Here the dashed blue red curve is the results obtained used the avereging in Eqs. (6.7) and (6.11), while the solid blue curve is the fidelity obtained using the past quantum state analysis to estimate the most like Bell-measurement outcome. The parameters used in the simulations are explained in the text.

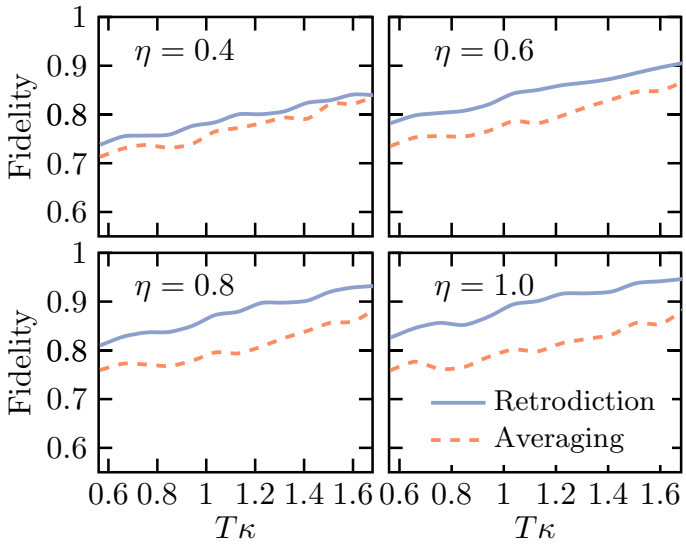

Figure 6.5: The fidelity of the protocol as a function of the total measuring time after the gates were applied. Each panel displays the fidelity of the protocol for the efficiency shown in the panel with the parameters explained in the main text.

when the past quantum state retrodiction is applied. For low detector efficiency and for short probing times, the methods have comparable and rather low fidelities, while for higher efficiencies and longer probing times, the past quantum state protocol generally performs better. In Fig. 6.4 (b) we see that the short readout time does provide enough data for more than a marginal improvement while the long readout time in Fig. 6.4 (c), accumulates enough data that also the integrated signal allows a good estimation of the optimal σ_B, and hence there is less room for improvement. Our interest in the intermediate probing time in Fig. 6.4 (a), where the past quantum state offers the largest improvement, is motivated by the existence of incoherent processes in the qubit B, which is not included in the present analysis but which will become important for long probing times.

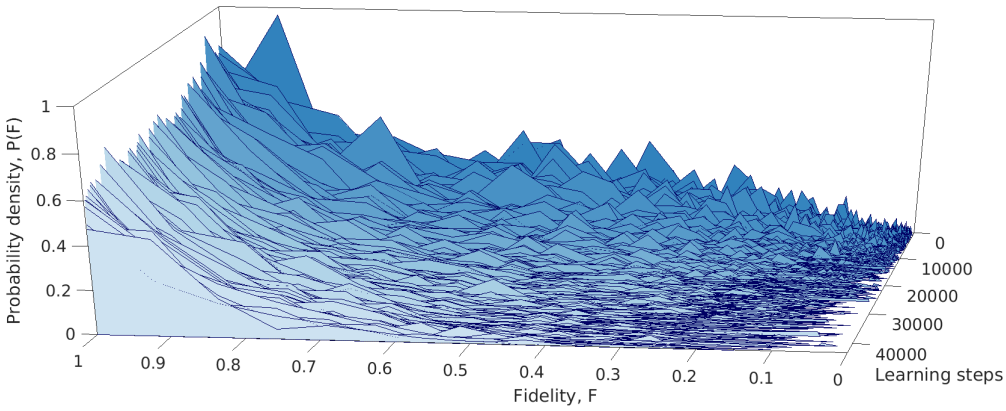

Figure 6.6: Histogram visualization of the training of neural network for visual recognition of the experomental currents for realization of the teleporatation protocol. Probability desity of the fidelity of protocol is shown as a function of the learning steps.

6.7 Neural network based post-processing

In the previous Sections, we offered an analytical microscopic model, that allowed us to simulate dynamics of the superconducting system. Since we had a good idea about inner dynamics of the system, we were able to employ the past quantum state method to improve our prediction about the most likely outcome of the Bell-measurement and correct upon the most error-prone part of the protocol. The appropriate post-processing of experimental currents led to a significant fidelity increase of the protocol. Let us now apply the methods used in Chapter 5 to reach the same goal.

The logic of this approach is precisely the same like the one used for post-processing. We have a quantum protocol, the crucial part of which is feed-forward of the classical information. This classical information is based on the result of a quantum measurement, and because of the readout limits and errors, this is a part of the protocol where we are most likely to make mistake (and obtain lower average fidelity for the protocol as a result).

Now, instead of post-processing the current and improving our prediction by the means of the generalized Born rule, we can apply the image recognition tools instead. We used the measured current leaking out of

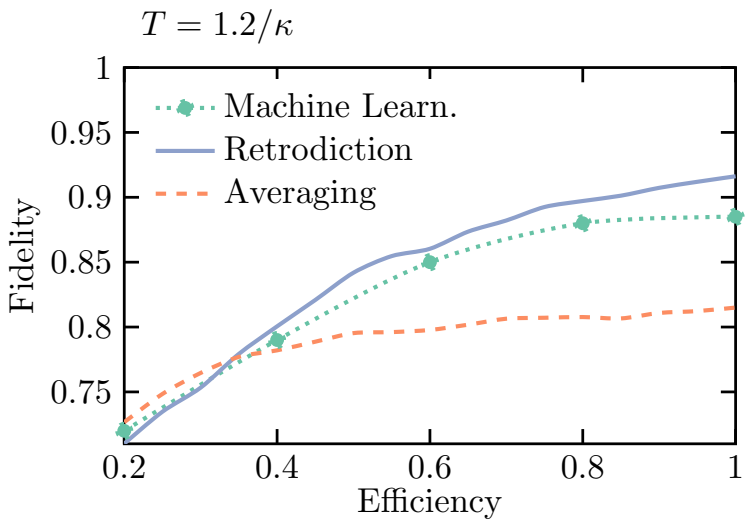

Figure 6.7: The fidelity of the protocol as a function of efficiency pre- and post the past quantum state retrodiction is applied in comparison with the machine learning approach. The dashed red curve is, again, obtained used the averaging, while the solid blue curve is the fidelity obtained using the past quantum state analysis to estimate the most like Bell-measurement outcome. The dotted green curve is a result of a neural network assisted analysis of the experimental currents.

the cavity directly as an input of the neural network depicted in Fig. 5.1. When we train the network on a sufficiently large training set it is able to recognize, which output of the Bell measurement (that we do not have direct access to) the current represents. In other words, we proposed to abandon any kind of averaging or current post-processing for plain image recognition of the features of the current as we did in Chapter 5 and [5].

In fact, the present situation is even simpler than the parameter estimation problem, because now we only have to be able to distinguish 4 options reliably (in difference from all the possible combinations of relevant parameters). Of course, depending on the experimental parameters the currents representing different outcomes of Bell measurement might be

visually very similar, but, ideally, a sufficiently large training set should allow us to distinguish them.

We were able to use exactly same neural network as used in Chapter 5[1], which shows the universality of our approach for the analysis of continuous measurements. The training of the neural network is visualized by the histogram in Fig. 6.6, where the probability density of the protocol fidelity is shown as a function of the learning step. In Figure 6.7 we again show the average fidelity of the protocol, now with an added curve showing the results of machine learning. One can see that while we were able to increase average fidelity significantly we were not able to overcome the past quantum state method. This is not surprising as the past quantum state method is based on a detailed probability calculation using the precise knowledge of the density matrix of the system, ρ, and the quantum operation, E, at any given time, t. In other words there is clearly a maximum of information leaking out of the cavity that sets a natural bound that even advanced image recognition neural network can not overcome. This said, we expect that increasing the size of the learning step and potentially size of the neural network would lead to increase of average fidelity and an eventual saturation of the PQS bound.

In this case we were able to make a direct comparison because we assumed all the noise to be Markovian and the master equation description therefore to be valid. Whenever this is not the case, we are left with the neural network analysis alone, and the development of the neural networks suited for classification of experimental noise specific to superconducting qubits is presently under investigation.

6.8 Outlook and conclusion

We have presented a scheme for quantum teleportation between two qubits taking turns interacting with a resonant cavity mode. Subsequent application of a continuous, homodyne measurement of the signal leaking from the cavity constitutes an approximate Bell state measurement. The accomplishment of this readout is significantly restricted by the measurement noise. Using both Bayesian analysis, in the form of the past quantum state formalism, and artificial neural neural network we show that it is possible to increase the fidelity of the protocol over schemes based on the value of the integrated signal alone. The protocol is not exact, and its high fidelity

[1] The only difference are two technical parameters: training step and batch size, see Appendix E for details.

limit merges naturally with the limit where also the integrated signal yields a successful outcome. We thus imagine that the main improvement will be for parameters where, e.g., decoherence and decay of the qubit components forbid high signal-to-noise probing of the system. Our results suggest that similar detailed analyses may improve other protocols implemented in leaky continuously monitored quantum systems.

Our protocol was exemplified in a circuit QED setup with an experimentally motivated preparation of the initial entangled state. We note, however, that the preparation of this initial state is not unique. For example, state-of-the-art remote entanglement schemes [150, 151] could be adapted to perform the state preparation. Similarly, after the state preparation, the target qubit state may be transferred to travelling microwave photons [152–154] without changing the rest of the protocol presented here. Moreover, the method can be easily translated to any other platform with a cavity and two-level systems. Our technique can be generalized to experimental realization of entanglement swapping [155], where a mixed state (one half of maximally entangled pair in particular) is teleported instead of a pure one.

Admittedly, quantum teleportation as a quantum protocol was originally formulated with the state transfer over very long distances in mind. In a circuit QED setting we can only teleport the state between two qubits with the relatively small state separation inside of the resonator. However, the length of a resonator can be up to centimeters, so it can still cover considerable distance on a chip. A network of resonators and qubits could be used for implementation of a quantum repeater [156] for arbitrary entanglement swapping on the large chip. While there are other techniques available for remote entanglement distribution [150, 151, 157], the teleportation scheme presented here provides a tool to explore a much wider range of quantum computation and communication protocols in a circuit QED system. Quantum teleporation for instance provides a universal resource for quantum computing [111] and plays a role in quantum error correction schemes [158], in particular in measurement based quantum computing [159–161]. Teleportation based techniques are also essential for modular approaches to both distributed quantum computing [162] and quantum error correction [163].

In conclusion, we believe that results of this Thesis provide a bright outlook for further efforts in combining quantum measurement with Bayesian and machine learning methods.

Chapter 7

Outlook and conclusions

The main focus of this Thesis is quantum measurement theory for realistic detection scenarios. We offered two main lines of thought: adaptation and improvement of existing quantum optics methods and an alternative neural network based approach.

We showed that the post-processing of the experimental currents based on the elements of quantum measurement theory can lead to significant improvement of probability predictions based on experimental records. Using the conditioning on the measurement record both preceding and following the time for which we want to make the probability prediction (the so-called past quantum state method) we are able to modify the Born rule and obtain better probability prediction. We used this method for both readout of quantum devices and quantum algorithms. We were furthermore able to improve parameter estimation methods for continuously monitored systems. We also extended these tools by introducing a novel parameter estimation method based on machine learning. In particular, we showed that it is possible to train an artificial neural network to recognize the physical parameters from experimental currents.

There are several directions in which this Thesis can be continued and extended. We believe our results presented in Chapter 3 and [4] can be readily applied to spin-based quantum computing [69, 71], but also in NV-center based sensing and magnetometry [73, 74] (see also 3.7). The results shown in Chapter 5 and [5] have direct applications for any quantum protocol containing operations conditioned on measurement outcome, as these can be decided and executed using our approach (see also Section 5.5). Chapter 6 and [3] shows one example of using the neural network tools developed in the work in the quantum protocol context. The approach

shown there can be generalized to more comlicated scenarios, like quantum repeaters [164, 165], or a quantum variational eigensolver for quantum optimization and quantum chemistry [166].

A more complex possibility, combining the methods described above, is to extend our artificial intelligence approach to full modeling of monitored open quantum systems. This idea combines machine learning with insights from traditional quantum optics theory.

In particular, our ability to recover unknown physical parameters in the cases where the standard quantum optics mathematical framework does not apply leads us to the question about the extent of new insights machine learning can provide in the field of quantum optics and its practical application. For this purpose the notion of the neural network state, proposed in [91], could be extended to describe the states and measurements in open quantum systems. The expectation for this research direction is that the traditional tools for description of open quantum systems and measurements can be replaced by neural network tools in order to to span larger sets of problems (i.e. mathematical assumptions needed to derive master equations could be dropped and we could model any quantum system regardless of the source of the noise, not just estimate its parameters). Such complex neural network apparatus would have immediate application in quantum metrology [98], quantum sensing [97], and quantum error correction [158, 167].

An additional approach to explore could take inspiration from quantum machine learning (QML) algorithms [86]. The QML algorithms are already good at classifying data with imperfections and noise (with a potential speed-up over classical methods), so there is a reasonable expectation that we can employ them for classification of the information in a phase estimation protocol that is in the core of most quantum metrology methods.

On the conceptual side, there are also general open questions about performance of machine learning for quantum tasks. For instance, in quantum parameter estimation we have analytical tools, like the Cramer-Rao bound [100], that specify precisely the success of a given method. For (quantum) machine learning methods no such measure exists to date. The same applies to possible application of artificial neural networks to quantum information processing and error correction in general, as the practical successes of the method might become hard to quantify without deeper understanding of its convergence.

Appendix A

Baum-Welch method for quantum systems

The Appendix provides the details about the quantum modified Baum-Welch estimation discussed in Chapter 3 and it overlaps with [4].

The dynamics of a quantum system can be inferred from the measurement data via a stochastic master equation [19, 20]. The average inferred behavior may be at variance with the values assumed for the transition rates in the master equation. This forms the basis for the classical Baum-Welch parameter re-estimation algorithm which iteratively extracts the rates from the inferred dynamics and re-applies them in the inference process, until the values have converged.

As discussed in Chapter 2, the extraction of the rates is based on the estimation of the joint probability,

$$C_t(i,j) = P(X_{t+dt} = j, X_t = i | D_1, ..., D_N), \tag{A.1}$$

that the system occupies the state i at the time t and the state j at the next discrete time step $t + dt$ (conditioned on the overall measurement record $D_1, \ldots, D_N$).

In the quantum case this joint classical probability has to be replaced by the probability that projective measurements at the two subsequent times yield the outcomes $|i\rangle$ and $|j\rangle$, respectively. The quantum theory of measurements allows evaluation of this probability by taking into account the effect of the evolution of the system between such projective measurements.

The past quantum state formalism (3.10), readily generalizes to yield the expression

$$C_t(i,j) = \mathrm{Tr}(|j\rangle\langle j|e^{(Ldt)}\{|i\rangle\langle i|\rho(t)|i\rangle\langle i|\}|j\rangle\langle j|E(t+dt)), \tag{A.2}$$

where $\exp(Ldt)$ is a superoperator denoting the propagator of the master equation for the short time step dt between the (hypothetical) measurements on the system [1].

Let us explain why this is the case. The expression $C_t(i,j)$ can be factorized using the conditional probability rules as

$$C_t(i,j) = P(X_{t+1} = j|X_t = i, D_1 \dots s_N) \cdot P(X_t = i|D_1 \dots D_N), \tag{A.3}$$

i.e. as a product of transition probability between states i and j in the consecutive time steps conditioned on the measurement record and the probability of starting from state i conditioned on the measurement record. Let us see whether we can rewrite this expression on terms of quantum Born rules to understand what these conditional probabilities mean in the case of quantum system.

First of all we note that the variable X_t corresponds to the state of the system (applied on Chapter 3 this means it can attain the values $\uparrow$, $\downarrow$, and 0) and D_t stands for experimentally measured current I (see Fig. (3.3)). As for expression (A.3) itself, its second part, $P(X_t = i|D_1 \dots D_N)$ corresponds to the probability of the system being in a given state, i, conditioned on the overall measurement record, or in other words, exactly for what generalized Born rule (3.10) stands for, i.e.

$$P(X_t = i|D_1 \dots D_N) \quad \mapsto \quad \frac{\mathrm{Tr}(|i\rangle\langle i|\rho_t|i\rangle\langle i|E_t)}{\sum_i \mathrm{Tr}(|i\rangle\langle i|\rho_t|i\rangle\langle i|E_t)}. \tag{A.4}$$

The first part of the product (A.3), $P(X_{t+1} = j|X_t = i, D_1 \dots D_N)$, describes the system in a given state, j, conditioned on *both* the measurement record and the state in the previous time step. Indeed, we can rewrite this probability in terms of the generalized Born rule as well

[1] This procedure estimates the entire transfer probability between the states, i.e., both contributions associated with the rate processes we want to estimate and contributions from unitary, coherent processes. Due to the projection on the eigenstates in the expression (A.2), however, coherent processes represented by $\exp(Ldt)$ will only contribute to second order in dt. We have verified this by omitting the coherent part of $\exp(Ldt)$ in our calculations without any appreciable change of the result.

$$P(X_{t=1} = j|X_t = i, D_1 \dots D_N) \quad \mapsto \frac{\mathrm{Tr}(|j\rangle\langle j|e^{(Ldt)}[|i\rangle\langle i|]|j\rangle\langle j|E_{t+1})}{\sum_j \mathrm{Tr}(|j\rangle\langle j|e^{(Ldt)}[|i\rangle\langle i|]|j\rangle\langle j|E_{t+1})}, \tag{A.5}$$

Grouping Eqs. (A.4)-(A.5) using Eq. (A.3) we straightforwardly obtain Eq. (A.2).

Now, the Baum-Welch procedure obtains the transition rate γ_{ij} from the average of this correlation function over the entire measurement record, normalized to the average population of the initial state $|i\rangle$ [37, 38]

$$\gamma_{ij} dt = \frac{\sum_t C_t(i,j)}{\sum_t (\sum_j C_t(i,j))}. \tag{A.6}$$

Appendix B

3-box paradox

In this Appendix we address a conceptual issue related to the contextuality of quantum mechanics and its relation to the single-electron quantum dot experiment described above. We have seen how the measurement back-action influences future probing results. When we evaluate the generalized probability (3.10) we are conditioning on both past and future evolution of the quantum state. Here we show that while the probability (6.12) to find a quantum system in a subspace spanned by two orthogonal states is just the sum of the probabilities of finding the system in either of the states, this is not true for the PQS retrodicted probabilities. This can be well explained on the analogy of the so-called 3-box paradox that is related to the pre- and post-selection in quantum mechanics. This Appendix has been to a high degree taken over from Chapter 3 of [6].

The 3-box paradox was introduced in [66] and roughly speaking it tells us that when having 3 boxes with particle in one of them, if we pre- and post- select the (unentangled!) wave function of the particle in a certain way we find that the particle is with certainty two of the boxes at the same time.

As pointed it out in [67] this puzzling behavior is not only similar to the dilemma presented in double-slit experiment, but what is worse, here we actually find out that the particle is in two out of three boxes with certainty. It is careful pre- and post-selection of the wave function that leads to this puzzling behavior and it has been recently experimentally tested [68].

Here we show that we can arrive to the same situation in the experiment described in the previous section using a past quantum state analysis. In particular, pre- and post selection come about as a natural part of

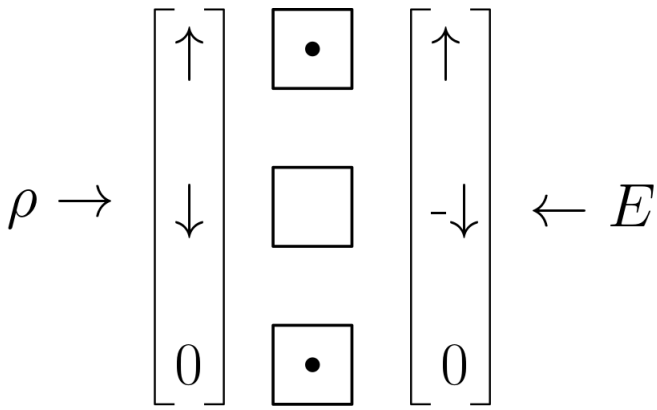

Figure B.1: Scheme of the setup of the 3-box paradox in the context of quantum dot experiment

performing our experiment and we will see that we can understand to the 3-box paradox in terms of the back-action of the given POVM on our system.

Let us start by description of the paradox [67]: we have one particle and three separate boxes. Let us assume that the state of the particle is pre-selected as

$$|\Psi\rangle = \frac{1}{\sqrt{3}}(|A\rangle + |B\rangle + |C\rangle) \tag{B.1}$$

and post-selected as

$$\langle\Phi| = \frac{1}{\sqrt{3}}(\langle A| + \langle B| - \langle C|). \tag{B.2}$$

Here we assume that $|A\rangle$, $|B\rangle$, and $|C\rangle$ are mutually orthogonal. Between pre- and post-selection the state the observation of either box A or box B is being made. For the (projective) observation of box A we have POVM consisting of $\{P_A = |A\rangle\langle A|, \mathbb{1} - P_A = |B\rangle\langle B| + |C\rangle\langle C|\}$. Similarly, POVM for observation of box B reads $\{P_B = |B\rangle\langle B|, \mathbb{1} - P_B = |A\rangle\langle A| + |C\rangle\langle C|\}$. Now, if the particle is not found in A its state is projected onto $\frac{1}{\sqrt{2}}(|B\rangle + |C\rangle)$, which happens to be orthogonal to the post-selected state and leads us to conclusion that the particle has to be in A. Now we can, however, repeat the same logic for B: not finding a particle there would give a state $\frac{1}{\sqrt{2}}(|A\rangle + |C\rangle$, which is, again, orthogonal to the post-selected state $\langle\Phi|$.

How does 3-box paradox fit into the experimental setup and data analysis described in above? As in our model QPC is insensitive to the

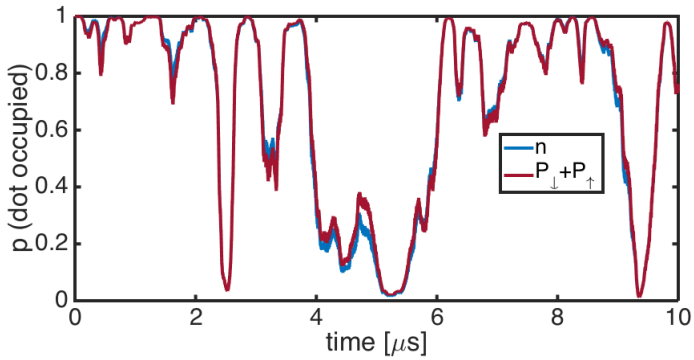

Figure B.2: The different back-action dependent on the measurement for the case of the trajectory of this note: blue curve (violet) represents the final probability curve for the dot occupancy in case the spins were measured separately (the number operator was measured).

spin state inside of the dot, the measurement operator we are interested in our analysis is simply the number operator $\hat{n} = |\uparrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow|$. If we were to evaluate the probability of the dot occupancy according to usual $Tr(M\rho M^\dagger)$ with $M = \hat{n}$ we would find that

$$\mathrm{Tr}(\hat{n}\rho\hat{n}^\dagger) = \mathrm{Tr}(P_\uparrow\rho P_\uparrow^\dagger) + \mathrm{Tr}(P_\downarrow\rho P_\downarrow^\dagger) \tag{B.3}$$

just by applying basic algebraic properties of projectors. Here we denote $P_\uparrow = |\uparrow\rangle\langle\uparrow|$ the projector on spin-up state and similarly $P_\downarrow = |\downarrow\rangle\langle\downarrow|$ the projector on spin-down state.

If we, however, use the generalized Born rule (3.10) we get

$$\mathrm{Tr}(\hat{n}\rho\hat{n}^\dagger E) \neq \mathrm{Tr}(P_\uparrow\rho P_\uparrow^\dagger E) + \mathrm{Tr}(P_\downarrow\rho P_\downarrow^\dagger E) \tag{B.4}$$

as the left hand side of the equation above also contains mixed terms $\mathrm{Tr}(P_\uparrow\rho P_\downarrow^\dagger E)$ and $\mathrm{Tr}(P_\downarrow\rho P_\uparrow^\dagger E)$. In other words, in terms of the measurement back-action it *makes a difference whether we project on the number operator subspace or measure spin-up and spin-down states separately.*

This also means that the state of the quantum dot reconstructed from master equation will be different if we use a point contact insensitive to the spin state to calculate the dot occupancy at a given time or, if we had a sensitive one and we recorded the signals for spin-up and spin-down separately and summed them up.

All this happens because of different measurements have different back-action on our system. In particular, it has to do with the contextuality of quantum mechanics: in one instance we obtain state *in context* of projecting on the {spin-up, spin-down}-subspace, in the other the state of our system is *in context* of projecting on the different spin subspaces separately.

Let us now understand to this back-action in the 3-box paradox terms. The three boxes now corresponds to the three different states of the of the dot $\{|\uparrow\rangle, |\downarrow\rangle.|0\rangle\}$. The pre- and post-selection of the state in this context is due to the fact that the measurement operator $\hat{n}$ that has rank 2 and spans both spin-up and spin-down state, ρ pre-selects randomly the phase C of $|\uparrow\rangle + C|\downarrow\rangle$, while E again randomly post-selects a different one. Essentially, after applying measurement operator $\hat{n}$ in order to find out whether there is a spin in the dot at all (regardless its state) we end up not knowing whether we have the state $|\uparrow\rangle + |\downarrow\rangle$ or $|\uparrow\rangle - |\downarrow\rangle$ and at the given time point different superposition can be 'picked' by ρ and E.

The new situation can be seen on the Fig. B.1. Our pre- and post-selected states are $|\Psi\rangle = \frac{1}{\sqrt{3}}(|\uparrow\rangle + |\downarrow\rangle + |0\rangle)$ and $\langle\Phi| = \frac{1}{\sqrt{3}}(\langle\uparrow| - \langle\downarrow| + \langle 0|)$ respectively. If we now ask whether the dot is in the state $|\uparrow\rangle$, following the logic above, we find out that negative answer would contradict the resulting state and similarly for the question whether there is no spin in the dot at all, i.e. a state $|0\rangle$. It is easy to see that just by choosing opposite phase of our post-selected state we could also rephrase this for the states $|\downarrow\rangle$ and $|0\rangle$. In this sense we could always associate a measurement operator with rank > 1 with the three box paradox using past quantum state.

The difference between choosing one POVM or the other can be seen at the Fig. B.2, where we plot the resulting probability curve for the dot occupancy first for the case where the spin-up and spin-down probability were recorded separately and secondly for the case where we simulate the measurement of the number operator. We can see that the second curve is a little bit more noisy, which is agreement with our intuition of having more back-action coming from the cross-terms $\mathrm{Tr}(P_\uparrow \rho P_\downarrow^\dagger E)$ and $\mathrm{Tr}(P_\downarrow \rho P_\uparrow^\dagger E)$.

To conclude, in this section we have shown more fundamental implication of past quantum state. Namely we have seen that at the first sight very artificial paradox arose just via the use of the retrodiction. We have described the discrepancy that appeared in the simulation data and explained it in the terms of the three box paradox.

Appendix C

Correlation functions and past quantum state

In this Appendix we discuss a PQS take on correlation functions in quantum mechanics. Its content is taken over from Chapter 5 of [6] and overlaps significantly with [1].

Since the observation by Hanbury-Brown and Twiss of temporal correlations in photodetection signals [168] (bunching and anti-bunching), it became clear that dynamics on the quantum level is significantly different from the classical one.

Correlation functions play an important role in the analysis of time dependent processes with applications ranging from demographics and finance to engineering and physics. Stochastic realizations of different signals can be characterized by, e.g., their mean values and variances while their temporal dynamics is captured by correlations between their values at different times. One may thus evaluate products of two signals at different times, $m1(t1)m2(t2)$, and define the temporal (two-time) correlation function of the signals as the average of this quantity, $\overline{m1(t1)m2(t2)}$, over different realizations of the process. While the expression deals with classical signals $m1(t)$ and $m2(t)$, they may have their origin in measurements on a quantum system, and hence we must apply quantum theory to analyze and make predictions for the signal correlations.

The description of how the detection signal is obtained via the absorption process in the detector is provided by Glauber's photodetection theory. Its central objects are correlation functions. In [1] we studied amplitude-intensity correlation functions. In the optical settings these represent joint photon counting and homodyne detection. In particular, in the

experiments [169, 170] the homodyne amplitude signal shows correlations also with later photon counting events. Mathematically this correlation function is defined as

$$g^{(1.5)}(\tau > 0) \propto \langle \sigma^\dagger(t)(e^{i\Phi}\sigma^\dagger(t+\tau) + e^{-i\Phi}\sigma(t+\tau))\sigma(t)\rangle, \tag{C.1}$$

where σ is lowering operator for the atomic transition and Φ is the phase component measured by the homodyne detection. We use the proportionality sign as for the sake of compactness we have omitted the normalization. Using the normal ordering of Glauber we arrive to similar expression for the negative time delay τ

$$g^{(1.5)}(\tau < 0) \propto e^{i\Phi}\langle \sigma^\dagger(t)\sigma^\dagger(t-\tau)\sigma(t-\tau)\rangle + e^{-i\Phi}\langle \sigma^\dagger(t-\tau)\sigma(t-\tau)\sigma(t)\rangle. \tag{C.2}$$

A well-known tool for the evaluation of correlation functions is the so-called *quantum regression theorem* (QRT) [22] which yields the following solution for the two-times correlation function (assuming $\tau > 0$)

$$\langle \hat{O}_1(t)\hat{O}_2(t+\tau)\hat{O}_3(t)\rangle = \mathrm{Tr}(\hat{O}_2 e^{\mathcal{L}\tau}[\hat{O}_3\rho_{st}\hat{O}_1]), \tag{C.3}$$

where $\mathcal{L}$ is the Liouvillian of the system we introduced in (2.3) and ρ_{st} is the steady state of the system. The quantum regression theorem is the standard tool for evaluation of correlation functions. We can now use QRT to evaluate the expressions (C.1) and (C.2).

Note, however, that while the expression (C.1) admits for the simple interpretation using conditioned evolution of quantum state ($G^{(2)}(\tau) = \bar{p}_e \mathrm{tr}\left\{|e\rangle\langle e|e^{\mathcal{L}\tau}[|g\rangle\langle g|]\right\}$ with $\bar{p}_e = \langle e|\rho_{st}|e\rangle$), for the expression (C.2) this is not possible as the QRT time evolution would be applied to the objects $\sigma\rho_{st}$ and $\rho_{st}\sigma^\dagger$ which do not permit a density matrix interpretation.

In order to find physical interpretation of the negative time delay correlation function we can, however, with advantage use PQS. In particular, if the positive delay correlation function is nothing else but conditioned time evolution governed by a master equation

$$\frac{d\rho}{dt} = -\frac{i}{\hbar}[H,\rho] + \sum_\mu \hat{c}\rho\hat{c}^\dagger - \frac{1}{2}(\hat{c}^\dagger\hat{c}\rho - \rho\hat{c}^\dagger\hat{c}). \tag{C.4}$$

Similarly we can view the negative delay correlation function we the condition time evolution in the negative direction of time. This evolution is governed by an adjoint equation

$$\frac{dE}{dt} = \frac{i}{\hbar}[H,E] + \sum_\mu \hat{c}^\dagger E\hat{c} - \frac{1}{2}(\hat{c}^\dagger\hat{c}E - E\hat{c}^\dagger\hat{c}). \tag{C.5}$$

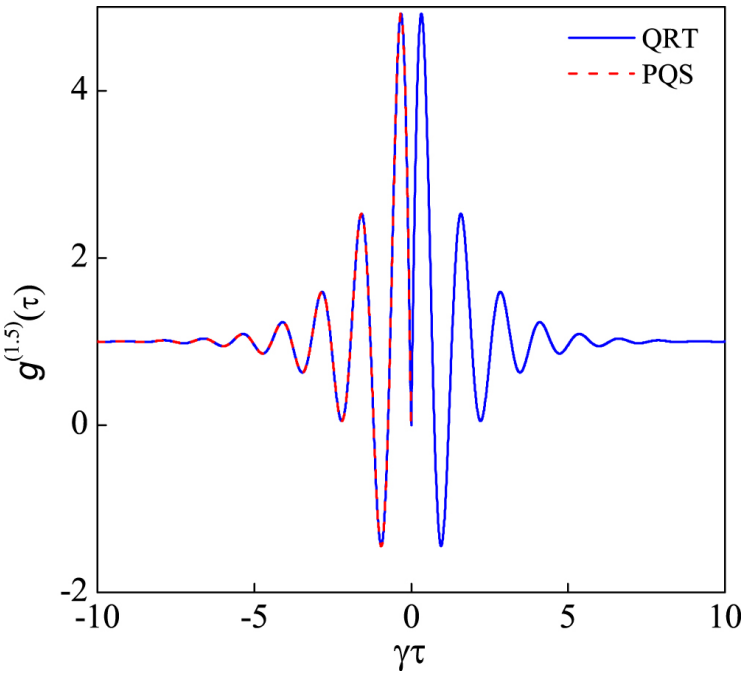

Figure C.1: Amplitude–intensity correlation function (C.1) and (C.1) for a resonantly driven two-level atom determined by the quantum regression theorem (blue solid line) and by the past quantum state (red dashed line). The parameters are: $\Omega = 5\gamma$ and $\varphi = \pi/2$.

In Figure C.1, we show the value of the normalized correlation function $g^{(1.5)}(\tau)$ for a two-level atom for both positive and negative τ, and it is interesting that despite the very different formal expressions, the evaluation of (C.1) and (C.2) yields a symmetric correlation function around $\tau = 0$. In addition to that we also show the respective conditioned past quantum state evolution given by (C.5) that exactly reproduces the negative delay correlation function (C.2).

Symmetry of the amplitude-intensity correlation for two level system is indeed a very special case. Already in the case three-level system we can find example exhibiting strong asymmetry [171]. Let us consider two possible 3 level schemes depicted in the Figure C.2: a three level ladder system and a V-system.

Let us take a look at these in more detail and use PQS to provide an

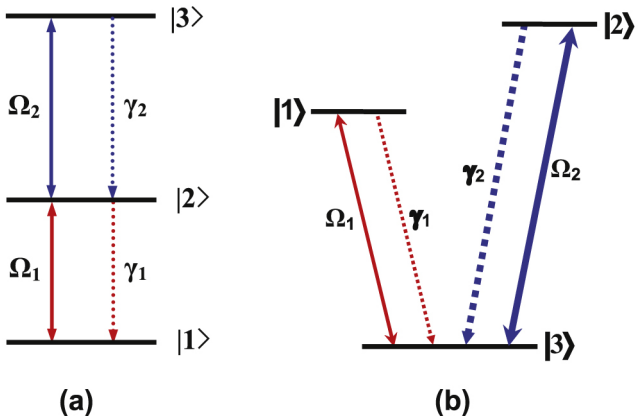

Figure C.2: A three-level atom driven by two resonant laser fields. In the ladder configuration (a), the fields are of comparable strength while the upper state $|3\rangle$ has longer lifetime than the intermediate state $|2\rangle$. In the V-configuration (b), the driving field and the decay rate on the $|2\rangle \leftrightarrow |3\rangle$ transition are much stronger than those on the $|1\rangle \leftrightarrow |3\rangle$ transition.

intuition about the behavior of the forward and backward evolution.

We consider three-level ladder atom with energy levels as drawn in the Fig. C.2. Its Hamiltonian can be written as

$$\begin{aligned} H &= \hbar[\Delta_1\sigma_{22} + (\Delta_1 + \Delta_2)\sigma_{33}] \\ &\quad -\frac{\hbar}{2}(\Omega_1\sigma_{21} + \Omega_2\sigma_{32} + \text{H.c.}), \end{aligned} \tag{C.6}$$

where $\sigma_{ij} = |i\rangle\langle j|$, $(i, j = 1 - 3)$, $\Omega_{1,2}$ are the laser Rabi frequencies, and $\Delta_1 = \omega_{21} - \nu_1$ and $\Delta_2 = \omega_{32} - \nu_2$ are the detunings, with ω_{21} and ω_{32} atomic transition frequencies and $\nu_{1,2}$ frequencies of the probe field for the lower and higher transition respectively. Quantum jump operators are $C_1 = \sqrt{\gamma_1}\sigma_{12}$ and $C_2 = \sqrt{\gamma_2}\sigma_{23}$ for damping (with the radiative damping rates $\gamma_{1,2}$), and $C_3 = \sqrt{\gamma_{\text{ph}}}(\sigma_{33} - \sigma_{22} - \sigma_{11})$ for dephasing (with the rate of phase decay γ_{ph}).

Calculating the correlation functions for the lower transition of the three level ladder model using both QRT and PQS we obtain the results plotted in the Figure C.3. We indeed see striking asymmetry. In particular, the part of the correlation function for the negative time delay appears to have double the frequency of the part with the positive time delay. Can

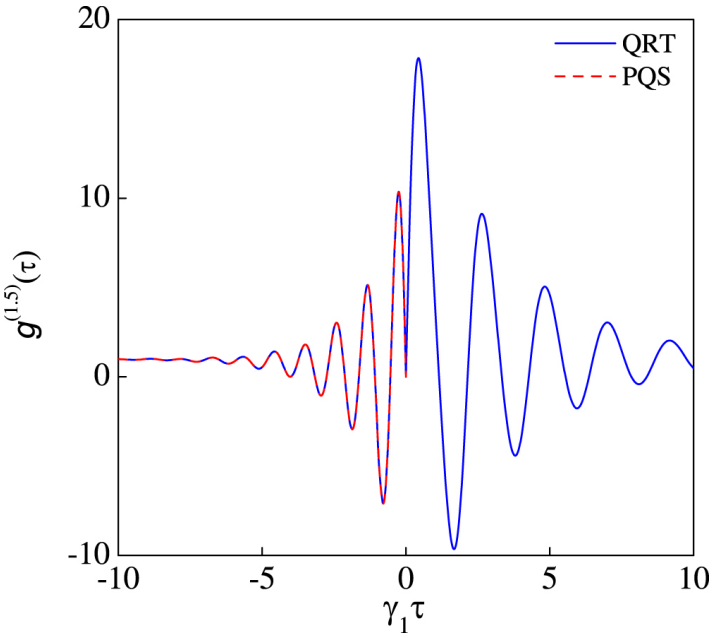

Figure C.3: Amplitude–intensity correlation on the lower transition of a single three-level ladder atom by invoking the quantum regression theorem (blue solid line) and past quantum state (red dashed line). The parameters are: $\Omega_1 = 3\gamma_1$, $\Omega_2 = 5\gamma_1$, $\gamma_2 = 0.2\gamma_1$, $\gamma_{ph} = 5 \times 10^{-5}\gamma_1$, and $\varphi = \pi/2$.

we understand this behavior using PQS? Let us forget about dissipation for a moment and concetrate on the Hamiltonian evolution alone. The Hamiltonian (C.6) has three eigenstates

$$\begin{aligned} |+\rangle &= \frac{1}{\sqrt{2}}(\cos\vartheta|1\rangle - |2\rangle + \sin\vartheta|3\rangle), \\ |0\rangle &= -\sin\vartheta|1\rangle + \cos\vartheta|3\rangle, \\ |-\rangle &= \frac{1}{\sqrt{2}}(\cos\vartheta|1\rangle + |2\rangle + \sin\vartheta|3\rangle). \end{aligned} \tag{C.7}$$

with eigenenergies

$$E_+ = \frac{\hbar}{2}\Omega_\mathrm{R}, \quad E_0 = 0, \quad E_- = -\frac{\hbar}{2}\Omega_\mathrm{R}, \tag{C.8}$$

where we defined the generalized Rabi frequency $\Omega_\mathrm{R} = \sqrt{\Omega_1^2 + \Omega_2^2}$ and the angle ϑ via $\tan\vartheta = \frac{\Omega_2}{\Omega_1}$.

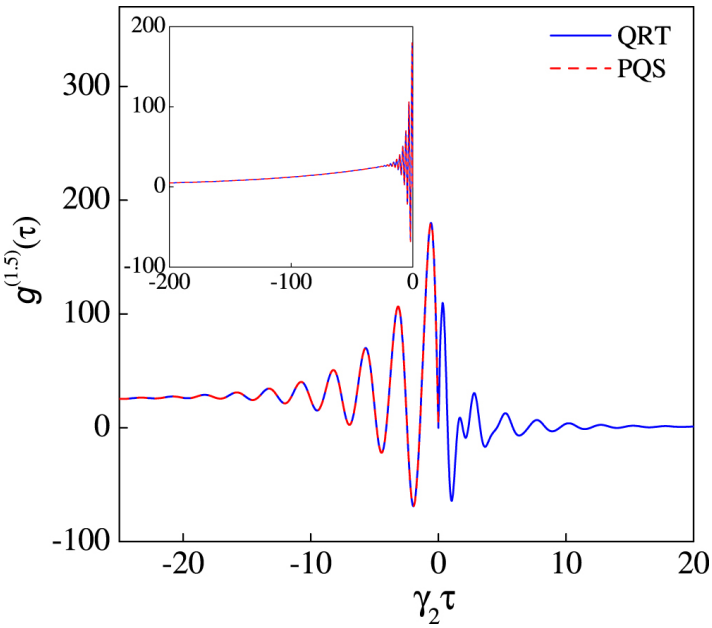

Figure C.4: Amplitude–intensity correlation on the weak probe transition of a single three-level V atom by invoking the quantum regression theorem (blue solid line) and past quantum state (red dashed line). The inset indicates a long (negative) time trace approaching unity on the 'Zeno time scale'. The parameters are:$\Omega_1 = 0.1\gamma_2$, $\Omega_2 = 5\gamma_2$, $\gamma_1 = 0.01\gamma_2$, $\varphi = \pi/2$.

For the forward evolution, starting from $|1\rangle$ the time evolution will clearly be a superposition of time evolution of all three eigenstates (and we see that the relative frequency of their time evolution is $\Omega_R/2$). Backward evolution, on the other hand, starts from the state $|2\rangle$. This is the case because for the adjoint evolution the creation operators are turned into annihilation ones and vice versa. The ground state $|1\rangle$ initial condition was associated with the photodetection click (application of annihilation operator). Therefore the initial condition of the backward evolution will be associated with the adjoint of this operation, which results in the state $|2\rangle$. In this case, however, the evolution will be only superposition of $|+\rangle$ and $|-\rangle$ eigenstates, the relative frequency of which is Ω_R. We could of course make the similar case for the other transition in the model depicted in the Figure C.2.

We can make a similar case for the V-system. Hamiltonian of this system reads

$$\begin{aligned} H &= \hbar(\Delta_1\sigma_{22} + \Delta_2\sigma_{33}) \\ &\quad -\frac{\hbar}{2}(\Omega_1\sigma_{21} + \Omega_2\sigma_{31} + \text{H.c.}), \end{aligned} \tag{C.9}$$

and quantum jump operators $C_1 = \sqrt{\gamma_1}\sigma_{12}$ and $C_2 = \sqrt{\gamma_2}\sigma_{13}$, and $C_3 = \sqrt{\gamma_{\text{ph}}}(\sigma_{33} - \sigma_{22} - \sigma_{11})$.

The resulting correlation function is plotted in the Figure C.4. Here the origin of the asymmetry is fundamentally different from the previous case. For $\tau > 0$, we observe damped dipole moment oscillations, with an additional modulation that we ascribe to the rapid dynamics on the stronger atomic transition. For $\tau < 0$, the oscillations are more regular, however they are damped towards a large almost constant value, and only on a much longer time scale (see insert), the correlation function converges to unity.

We can account for both of these observations by the backward evolution of the matrix $E(t)$: first, we note that E(t) has the excited state $|1\rangle\langle 1|$ as its final state. This state is resonantly driven towards the state $|3\rangle$, which is, however, strongly perturbed by the laser excitation of the $|2\rangle \leftrightarrow |3\rangle$ transition with Rabi frequency Ω_2. This driving causes an AC Stark splitting of the state $|3\rangle$ by Ω_2 and detunes the weak transition by $\Omega_2/2$, which emerges as a generalized Rabi frequency of the weak and oscillatory amplitude transfer. The incoherent rate processes between states $|2\rangle$ and $|3\rangle$ damp these off-resonant oscillations on the time scale of a few $(\gamma_2)^{-1}$. The matrix E(t) has not yet reached its steady state value, and despite the coherent coupling strength Ω_1, it is only for times longer than the time scale $(\gamma_1)^{-1}$, that the correlation function regresses to the uncorrelated product. Within the picture of an evolving matrix $E(t)$, we find this suppression of coherent dynamics similar to the quantum Zeno dynamics [172], observed for evolving density matrices [173].

In this Appendix we reviewed some application of PQS theory in the context of the interpretation of temporal asymmetry of quantum optics correlation functions. The detailed theoretical foundation for this application using POVM language is given in [1].

Appendix D

Quantum Dot: Numerical Simulation

In this Appendix we explain the details of the stochastic master equation simulation as well as the realization of our technique based on the actual experimental current. This appendix partly overlaps with [4].

Let us consider the POVM (3.6). Equivalently, we can write these in terms of the change of tunneling amplitude through the QPC. In particular, using the notation of [19] (where also the detailed description of the tunneling processes in QPC can be found) we call the amplitude in the absence (presence) of the dot τ ($\tau + \xi$). We can related these easily to the rates

$$r_0 = |\tau|^2, \quad r_1 = |\tau + \xi|^2.$$

The POVM (3.6) correspond to the point-wise current through the QPC

$$I(t) = e(r_0 + (r_1 - r_0)\langle n \rangle(t)), \tag{D.1}$$

where $\langle n \rangle$ is the occupation of the dot, which will jump between 0 and 1, and e is the elementary charge. The current will consequently have two values er_0 and er_1 for empty and occupied dot respectively.

The more realistic model is the one, where current is very big compared to the tunneling change induced by the presence of an electron in the dot. In particular, this means that many electrons have to pass through QPC before we can distinquish the occupation of the dot. To derive this limit

we have to make use of so-called *diffusive approximation* [19]. This allows us to transform the master equation for the point process

$$d\rho = dN(t)\mathcal{G}[\sqrt{\mu}\tau + \xi\hat{n}]\rho - \frac{1}{2}dt\mathcal{H}[iH + 2\mu\tau\xi\hat{n} + (\xi\hat{n})^2]\rho$$
$$+ dt(\gamma_L\mathcal{D}[c_\downarrow^\dagger] + \gamma_R\mathcal{D}[c_\uparrow] + (1-\mu)\mathcal{D}[\tau + \xi\hat{n}])\rho, \tag{D.2}$$

where $\mathcal{H}[r]\rho = r\rho + \rho r^\dagger - \mathrm{Tr}[r\rho + \rho r^\dagger]\rho$ and $\mathcal{G}[r]\rho = \frac{r\rho r^\dagger}{\mathrm{Tr}[r\rho r^\dagger]} - \rho$ and μ stands for measurement efficiency, to the case of continuous current

$$d\rho(t) = -i[H, \rho(t)]dt + dW(t)\sqrt{\mu}\xi\mathcal{H}[\hat{n}]\rho(t)$$
$$+ dt(\gamma_L\mathcal{D}[c_\downarrow^\dagger] + \gamma_R\mathcal{D}[c_\uparrow] + \xi^2\mathcal{D}[\hat{n}])\rho. \tag{D.3}$$

What we essentially did here, is the replacement of the precise measurement record specified by dN's that could only attain values 0 and 1 with the Gaussian random variable containing Wiener noise dW, which is used to model a shot noise [20]. The current of which the solution of (D.3) is conditioned is

$$I(t) = e\mu|\tau|^2 + e\mu\tau\xi - e\sqrt{\mu}\tau\frac{dW(t)}{dt}. \tag{D.4}$$

This precisely corresponds to the experimentally accessible current going through the QPC.

Now, taking into account the diffusive master equation (D.3) we can construct the POVM reflecting the back-action of our monitoring on the time evolution of the system in the spirit of Eq. (3.8)

$$M_I = \mathbb{1} - iHdt - \frac{1}{2}\mu(\xi\hat{n})^2dt - \sqrt{\mu}\xi\hat{n}dtI(t). \tag{D.5}$$

Appendix E

Parameter estimation with neural networks: Numerical simulation

In this Appendix we address technical aspect of artificial neural network used in Chapter 5. This Appendix overlaps with Supplementary Material of [5].

The neural network used in Chapter 5 is a convolutional neural network consisting of an input layer, two convolution layers (followed by pooling), one densely connected layer and an output layer. The activation function for the convolution layers and the densely connected layer is of the ReLU type as explained in Appendix A. The input layer represent here the continuously measured signal and is thus discretized on a time axis into bins of length, Δt. We therefore have an input layer of $N = T/\Delta t$ nodes, where T is the full time of the recorded measurement current. We set $\Delta t = 10\,\text{ns}$ which corresponds to a realistic sample rate for experimental realizations and the simulations are performed for up to $T = 500\,\mu\text{s}$ such that we have $N = 5 \times 10^4$. In other words, the input layer can be represented by a tensor with dimension $[M \times N \times 1]$, where M is the *batch size*, i.e. the number of currents loaded. The input data is scaled to be between 0 and 1.

The batch size is arbitrary as no optimization parameter depends on M. As such, once the network is trained, we can perform the prediction step with a single experimental current, i.e. $M = 1$. For the training we used consecutive batches each with a batch size of $M = 1000$ and for each

batch we perform 250 learning steps (back-propagations) and similarly, in order to have a small statistical error, we evaluated the performance of the network with a single batch with $M = 1000$. The evaluation batch was not used for training. The batches are generated offline, i.e. prior to loading of the neural network, and to save memory usage only one batch is being saved in memory at a time. In the numerical simulation of the physical system, each trajectory used a different random seed and for all trajectories in the batch, the parameters were randomly chosen in order to sample the whole parameter space. For all the generated currents, we also save the parameters used in the simulation which we can use to generate the target probability function and for calculating the distance measured used in the main text. In practice, the target parameters are loaded into memory simultaneous with the currents and saved as a *one-hot tensor* in each parameter that is a tensor representing the probability distribution for each parameter with only one non-zero entry.

The input layer is followed by a convolution layer with a kernel size of 9 such that each neuron is connected to 9 nodes in the input layer. We use 16 convolutional filters. We follow the convolution with a pooling layer with a pool size of 2. The resulting tensor is therefore $[M \times N/2 \times 16]$ and the layer is represented by 9×16 parameters, which are updated and optimized in each learning step. The second convolution layer is similar but with 32 filter resulting in a tensor size of $[M \times N/4 \times 32]$.

After the two convolution layers, the network has a densely connected layer with 1024 neurons with a ReLU activation function. As a result, the tensor is of dimensions $[M \times 1024]$ and the layer has $1024 \times N/4 \times 32$ optimization parameters. During the training, the densely connected layer is followed by a so-called dropout layer such that, for each learning step, a fixed random fraction, 40% in this work, of the neurons are dropped out of the network.

Since we have three parameters that we want to estimate each of which are picked from a set of 6 values, the output layer is of the size $[M \times 3 \times 6]$. A renormalization function is applied to each of the three dimensions such that the entry $[i, j, k]$ denotes the probability that the ith currents jth parameter has the kth value. The output layer is now finally compared to the target tensor. For the learning steps we calculate the cross-entropy for each parameter, while the evaluation steps output the average distance to the target and accuracy of the predictions.

The fidelity is calculated in a straightforward manner by considering the correct parameters for the trajectories in the evaluation batch. Here $p'_{i,j}(k) \in \{0, 1\}$ denotes the probability vector for the jth parameter of

Average distance	$T = 500\,\mu$s	$T = 200\,\mu$s	$T = 50\,\mu$s
Markovian	0.0028 MHz	0.0064 MHz	0.1084 MHz
Non-Markovian	0.0033 MHz	0.0050 MHz	0.1166 MHz

Table E.1: Average distance after 6250 learning steps for the three total measurement times for both the Markovian and non-Markovian case.

the ith trajectory, but as the correct parameter is known $p'_{i,j}(k)$ is only non-zero for a single k. Similarly we denote the probability outcomes of the softmax activation $p_{i,j}(k)$. The function $\text{argmax}_k(p_{i,j})$ returns the k corresponding to the maximal value of $p_{i,j}(k)$. We can therefore define the fidelity as

$$F = \frac{1}{3M} \sum_{i=1}^{M} \sum_{j=1}^{3} \left[\text{argmax}_k(p_{i,j}) == \text{argmax}_k(p'_{i,j})\right], \tag{E.1}$$

with $x == y$ denoting the logical equal operation, which returns 1 if x is equal to y and otherwise 0.

The last figure of merit that we calculate to evaluate the performance of the parameter estimation is the average distance from the estimated value to the real value. The parameter values for the parameter estimation performed in this work are as follow: $\Omega \in \{4,\ 5.2,\ 6.4,\ 7.6,\ 8.8,\ 10\}$ MHz and $\gamma_\uparrow, \gamma_\downarrow \in \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$ MHz. Consider now the vectors $\vec{\Omega}$, $\vec{\gamma}_\uparrow$ and $\vec{\gamma}_\downarrow$ to contain the candidate values for each parameters. The average distance is now calculated as

$$\begin{aligned} \bar{d} = \frac{1}{3M} \sum_{i=1}^{M} \Big(& |p_{i,1} \cdot \vec{\Omega} - p'_{i,1} \cdot \vec{\Omega}| + |p_{i,2} \cdot \vec{\gamma}_\uparrow - p'_{i,2} \cdot \vec{\gamma}_\uparrow| \\ & + |p_{i,3} \cdot \vec{\gamma}_\downarrow - p'_{i,3} \cdot \vec{\gamma}_\downarrow| \Big), \end{aligned} \tag{E.2}$$

with $x \cdot y$ denoting the inner product between x and y.

The learning process was terminated at the point of convergence for the Markovian case with a time trace of $T = 500\,\mu$s at which point the average distance was 0.0028 MHz. Similarly, the learning process for the rest of the simulated cases was terminated after the same amount of time-steps. The resulting average distance at termination is shown in Table E.1.

Bibliography

[1] Qing Xu, Eliska Greplova, Brian Julsgaard, and Klaus Mølmer. Correlation functions and conditioned quantum dynamics in photodetection theory. *Physica Scripta*, 90(12):128004, 2015.

[2] Eliska Greplova and Géza Giedke. Degradability of fermionic gaussian channels. *arXiv preprint arXiv:1604.01954*, 2016.

[3] Eliska Greplova, Klaus Mølmer, and Christian Kraglund Andersen. Quantum teleportation with continuous measurements. *Phys. Rev. A*, 94:042334, Oct 2016.

[4] Eliska Greplova, Edward A. Laird, G. Andrew D. Briggs, and Klaus Mølmer. Conditioned spin and charge dynamics of a single-electron quantum dot. *Phys. Rev. A*, 96:052104, Nov 2017.

[5] Eliska Greplova, Christian Kraglund Andersen, and Klaus Mølmer. Quantum parameter estimation with a neural network. *arXiv preprint arXiv:1711.05238*, 2017.

[6] Eliska Greplova. Past quantum state: Theory and applications, 2016. Part-A Progress Report.

[7] Maximilian Schlosshauer. Decoherence, the measurement problem, and interpretations of quantum mechanics. *Reviews of Modern physics*, 76(4):1267, 2005.

[8] AJ Leggett. The quantum measurement problem. *science*, 307(5711):871–872, 2005.

[9] Stephen Boughn and Marcel Reginatto. A pedestrian approach to the measurement problem in quantum mechanics. *arXiv preprint arXiv:1309.0724*, 2013.

[10] Arthur Fine. Insolubility of the quantum measurement problem. *Physical Review D*, 2(12):2783, 1970.

[11] Hannes Bernien, Sylvain Schwartz, Alexander Keesling, Harry Levine, Ahmed Omran, Hannes Pichler, Soonwon Choi, Alexander S Zibrov, Manuel Endres, Markus Greiner, et al. Probing many-body dynamics on a 51-atom quantum simulator. *arXiv preprint arXiv:1707.04344*, 2017.

[12] J. M. Elzerman, R Hanson, L H Willems Van Beveren, B Witkamp, L M K Vandersypen, and L. P. Kouwenhoven. Single-shot read-out of an individual electron spin in a quantum dot. *Nature*, 430(6998):431–5, 2004.

[13] Rami Barends, Julian Kelly, Anthony Megrant, Andrzej Veitia, Daniel Sank, Evan Jeffrey, Ted C White, Josh Mutus, Austin G Fowler, Brooks Campbell, et al. Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature*, 508(7497):500–503, 2014.

[14] C Neill, P Roushan, K Kechedzhi, S Boixo, SV Isakov, V Smelyanskiy, R Barends, B Burkett, Y Chen, Z Chen, et al. A blueprint for demonstrating quantum supremacy with superconducting qubits. *arXiv preprint arXiv:1709.06678*, 2017.

[15] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M Chow, and Jay M Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*, 549(7671):242–246, 2017.

[16] David Kielpinski, Chris Monroe, and David J Wineland. Architecture for a large-scale ion-trap quantum computer. *Nature*, 417(6890):709–711, 2002.

[17] Nissim Ofek, Andrei Petrenko, Reinier Heeres, Philip Reinhold, Zaki Leghtas, Brian Vlastakis, Yehan Liu, Luigi Frunzio, SM Girvin, L Jiang, et al. Extending the lifetime of a quantum bit with error correction in superconducting circuits. *Nature*, 536(7617):441–445, 2016.

[18] Bas Hensen, Hannes Bernien, Anaïs E Dréau, Andreas Reiserer, Norbert Kalb, Machiel S Blok, Just Ruitenberg, Raymond FL Vermeulen,

Raymond N Schouten, Carlos Abellán, et al. Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres. *Nature*, 526(7575):682–686, 2015.

[19] H.M. Wiseman and G.J. Milburn. *Quantum Measurement and Control*. Cambridge University Press, 2010.

[20] K. Jacobs and D. Steck. A straightforward introduction to continuous quantum measurement. *Contemporary Physics*, 47:279–303, September 2006.

[21] Klaus Mølmer, Yvan Castin, and Jean Dalibard. Monte carlo wavefunction method in quantum optics. *J. Opt. Soc. Am. B*, 10(3):524–538, Mar 1993.

[22] C. Gardiner and P. Zoller. *Quantum Noise: A Handbook of Markovian and Non-Markovian Quantum Stochastic Methods with Applications to Quantum Optics*. Springer Series in Synergetics. Springer, 2004.

[23] H. Carmichael. *An Open Systems Approach to Quantum Optics: Lectures Presented at the Université Libre de Bruxelles, October 28 to November 4, 1991*. Number v. 18 in An Open Systems Approach to Quantum Optics: Lectures Presented at the Université Libre de Bruxelles, October 28 to November 4, 1991. Springer Berlin Heidelberg, 1993.

[24] Klaus Mølmer and Yvan Castin. Monte carlo wavefunctions in quantum optics. *Quantum and Semiclassical Optics: Journal of the European Optical Society Part B*, 8(1):49, 1996.

[25] KW Murch, SJ Weber, Christopher Macklin, and Irfan Siddiqi. Observing single quantum trajectories of a superconducting quantum bit. *Nature*, 502(7470):211–214, 2013.

[26] Philippe Campagne-Ibarcq, Emmanuel Flurin, Nicolas Roch, David Darson, Pascal Morfin, Mazyar Mirrahimi, Michel H Devoret, François Mallet, and Benjamin Huard. Persistent control of a superconducting qubit by stroboscopic measurement feedback. *Physical Review X*, 3(2):021008, 2013.

[27] Agustin Palacios-Laloy, François Mallet, François Nguyen, Patrice Bertet, Denis Vion, Daniel Esteve, and Alexander N Korotkov. Experimental violation of a bell/'s inequality in time with weak measurement. *Nature Physics*, 6(6):442–447, 2010.

[28] Kater W Murch, Kevin L Moore, Subhadeep Gupta, and Dan M Stamper-kurn. Observation of quantum-measurement backaction with an ultracold atomic gas. *Nature Physics*, 4(7):561, 2008.

[29] Sebastien Gleyzes, Stefan Kuhr, Christine Guerlin, Julien Bernu, Samuel Deleglise, Ulrich Busk Hoff, Michel Brune, Jean-Michel Raimond, and Serge Haroche. Quantum jumps of light recording the birth and death of a photon in a cavity. *Nature*, 446(7133):297–300, 2007.

[30] Jun Jun Sakurai. *Modern Quantum Mechanics*. Addison-Wesley, 1985.

[31] Søren Gammelmark, Brian Julsgaard, and Klaus Mølmer. Past quantum states of a monitored system. *Phys. Rev. Lett.*, 111:160401, Oct 2013.

[32] T. Rybarczyk, B. Peaudecerf, M. Penasa, S. Gerlich, B. Julsgaard, K. Mølmer, S. Gleyzes, M. Brune, J. M. Raimond, S. Haroche, and I. Dotsenko. Forward-backward analysis of the photon-number evolution in a cavity. *Phys. Rev. A*, 91:062116, Jun 2015.

[33] D. Tan, S. J. Weber, I. Siddiqi, K. Mølmer, and K. W. Murch. Prediction and retrodiction for a continuously monitored superconducting qubit. *Phys. Rev. Lett.*, 114:090403, Mar 2015.

[34] P. Campagne-Ibarcq, L. Bretheau, E. Flurin, A. Auffèves, F. Mallet, and B. Huard. Observing interferences between past and future quantum states in resonance fluorescence. *Phys. Rev. Lett.*, 112:180402, May 2014.

[35] D. Tan, M. Naghiloo, K. Mølmer, and K. W. Murch. Quantum smoothing for classical mixtures. *Phys. Rev. A*, 94:050102, Nov 2016.

[36] Søren Gammelmark and Klaus Mølmer. Bayesian parameter inference from continuously monitored quantum systems. *Phys. Rev. A*, 87:032115, Mar 2013.

[37] S. Gammelmark, K. Mølmer, W. Alt, T. Kampschulte, and D. Meschede. Hidden markov model of atomic quantum jump dynamics in an optically probed cavity. *Phys. Rev. A*, 89:043839, Apr 2014.

[38] W.H. Press. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, 2007.

[39] Samuel L. Braunstein and Carlton M. Caves. Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.*, 72:3439–3443, May 1994.

[40] Sergio Boixo, Steven T. Flammia, Carlton M. Caves, and JM Geremia. Generalized limits for single-parameter quantum estimation. *Phys. Rev. Lett.*, 98:090401, Feb 2007.

[41] Greg A. Smith, Andrew Silberfarb, Ivan H. Deutsch, and Poul S. Jessen. Efficient quantum-state estimation by continuous weak measurement and dynamical control. *Phys. Rev. Lett.*, 97:180403, Oct 2006.

[42] David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Phys. Rev. Lett.*, 105:150401, Oct 2010.

[43] John K. Stockton, J. M. Geremia, Andrew C. Doherty, and Hideo Mabuchi. Robust quantum parameter estimation: Coherent magnetometry with feedback. *Phys. Rev. A*, 69:032109, Mar 2004.

[44] Jay Gambetta and H. M. Wiseman. State and dynamical parameter estimation for open quantum systems. *Phys. Rev. A*, 64:042105, Sep 2001.

[45] A. C. Doherty and K. Jacobs. Feedback control of quantum systems using continuous state estimation. *Phys. Rev. A*, 60:2700–2711, Oct 1999.

[46] R. Hanson, L. P. Kouwenhoven, J. R. Petta, S. Tarucha, and L. M. K. Vandersypen. Spins in few-electron quantum dots. *Rev. Mod. Phys.*, 79:1217–1265, Oct 2007.

[47] J. R. Petta, A. C. Johnson, J. M. Taylor, E. A. Laird, A. Yacoby, M. D. Lukin, C. M. Marcus, M. P. Hanson, and A. C. Gossard. Coherent manipulation of coupled electron spins in semiconductor quantum dots. *Science*, 309(5744):2180–2184, 2005.

[48] Marc Cahay. Quantum transport: Immune to local heating. *Nat Nano*, 9(2):97–98, 02 2014.

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

[50] Géraldine Haack, Mathias Albert, and Christian Flindt. Distributions of electron waiting times in quantum-coherent conductors. *Phys. Rev. B*, 90:205429, Nov 2014.

[51] Ivonne Guevara and Howard Wiseman. Quantum state smoothing. *Phys. Rev. Lett.*, 115:180407, Oct 2015.

[52] A N Jordan, Björn Trauzettel, and Guido Burkard. Weak-measurement theory of quantum-dot spin qubits. *Physical Review B*, 76(15):155324, 2007.

[53] D. J. Reilly, C M Marcus, M P Hanson, and A C Gossard. Fast single-charge sensing with a rf quantum point contact. *Applied Physics Letters*, 91:162101, jul 2007.

[54] M C Cassidy, A S Dzurak, R G Clark, K D Petersson, I Farrer, D. A. Ritchie, and C G Smith. Single shot charge detection using a radio-frequency quantum point contact. *Applied Physics Letters*, 110:222104, jul 2007.

[55] N Ares, F J Schupp, A Mavalankar, G Rogers, J P Griffiths, G A C Jones, I Farrer, D. A. Ritchie, C G Smith, A Cottet, G. A D Briggs, and E A Laird. Sensitive Radio-Frequency Measurements of a Quantum Dot by Tuning to Perfect Impedance Matching. *Physical Review Applied*, 5:034011, 2016.

[56] A. Stockklauser, P. Scarlino, J. V. Koski, S. Gasparinetti, C. K. Andersen, C. Reichl, W. Wegscheider, T. Ihn, K. Ensslin, and A. Wallraff. Strong coupling cavity qed with gate-defined double quantum dots enabled by a high impedance resonator. *Phys. Rev. X*, 7:011030, Mar 2017.

[57] M. R. Delbecq, V. Schmitt, F. D. Parmentier, N. Roch, J. J. Viennot, G. Fève, B. Huard, C. Mora, A. Cottet, and T. Kontos. Coupling a quantum dot, fermionic leads, and a microwave cavity on a chip. *Phys. Rev. Lett.*, 107:256804, Dec 2011.

[58] T. Frey, P. J. Leek, M. Beck, A. Blais, T. Ihn, K. Ensslin, and A. Wallraff. Dipole coupling of a double quantum dot to a microwave resonator. *Phys. Rev. Lett.*, 108:046807, Jan 2012.

[59] J. Basset, D.-D. Jarausch, A. Stockklauser, T. Frey, C. Reichl, W. Wegscheider, T. M. Ihn, K. Ensslin, and A. Wallraff. Single-electron double quantum dot dipole-coupled to a single photonic mode. *Phys. Rev. B*, 88:125312, Sep 2013.

[60] Frank H. L. Koppens, C Buizert, Klaas-Jan Tielrooij, Ivo T. Vink, Katja C. Nowack, T Meunier, L. P. Kouwenhoven, and L M K Vandersypen. Driven coherent oscillations of a single electron spin in a quantum dot. *Nature*, 442(7104):766–771, 2006.

[61] Jarryd J. Pla, Kuan Y. Tan, Juan P. Dehollain, Wee H. Lim, John J. L. Morton, David N. Jamieson, Andrew S. Dzurak, and Andrea Morello. A single-atom electron spin qubit in silicon. *Nature*, 489(7417):541–545, 09 2012.

[62] Katja C. Nowack, F. H. L. Koppens, Yuli V. Nazarov, and L M K Vandersypen. Coherent control of a single electron spin with electric fields. *Science (New York, N.Y.)*, 318(5855):1430–1433, 2007.

[63] E A Laird, C Barthel, E I Rashba, C M Marcus, M P Hanson, and A C Gossard. Hyperfine-Mediated Gate-Driven Electron Spin Resonance. *Physical Review Letters*, 99(24):246601, dec 2007.

[64] M. Pioro-Ladrière, T Obata, Y Tokura, Y.-S. Shin, Toshihiro Kubo, K Yoshida, T. Taniyama, and S Tarucha. Electrically driven single-electron spin resonance in a slanting Zeeman field. *Nature Physics*, 4(10):776–779, aug 2008.

[65] C Barthel, M. Kjaergaard, J Medford, M. Stopa, C M Marcus, M P Hanson, and A C Gossard. Fast sensing of double-dot charge arrangement and spin state with an rf sensor quantum dot. *Physical Review B*, 81:161308, 2010.

[66] Y Aharonov and L Vaidman. Complete description of a quantum system at a given time. *Journal of Physics A: Mathematical and General*, 24(10):2315, 1991.

[67] T. Ravon and L. Vaidman. The three-box paradox revisited. *Journal of Physics A Mathematical General*, 40:2873–2882, March 2007.

[68] Richard E. George, Lucio M. Robledo, Owen J. E. Maroney, Machiel S. Blok, Hannes Bernien, Matthew L. Markham, Daniel J. Twitchen, John J. L. Morton, G. Andrew D. Briggs, and Ronald

Hanson. Opening up three quantum boxes causes classically undetectable wavefunction collapse. *Proceedings of the National Academy of Sciences*, 110(10):3777–3781, 2013.

[69] D. Klauser, W A Coish, and D Loss. Nuclear spin state narrowing via gate-controlled Rabi oscillations in a double quantum dot. *Physical Review B*, 73(20):205302, 2006.

[70] Alexandr Sergeevich, Anushya Chandran, Joshua Combes, Stephen D. Bartlett, and Howard M. Wiseman. Characterization of a qubit Hamiltonian using adaptive measurements in a fixed basis. *Physical Review A*, 84(5):052315, 2011.

[71] M D Shulman, S P Harvey, J M Nichol, S D Bartlett, A C Doherty, V Umansky, and A Yacoby. Suppressing qubit dephasing using real-time Hamiltonian estimation. *Nature Communications*, 5(May):5156, 2014.

[72] O E Dial, M D Shulman, S P Harvey, H Bluhm, V Umansky, and A Yacoby. Charge Noise Spectroscopy Using Coherent Exchange Oscillations in a Singlet-Triplet Qubit. *Physical Review Letters*, 110:146804, 2013.

[73] L Childress, MV Gurudev Dutt, JM Taylor, AS Zibrov, F Jelezko, J Wrachtrup, PR Hemmer, and MD Lukin. Coherent dynamics of coupled electron and nuclear spin qubits in diamond. *Science*, 314(5797):281–285, 2006.

[74] V. M. Acosta, E. Bauch, M. P. Ledbetter, C. Santori, K.-M. C. Fu, P. E. Barclay, R. G. Beausoleil, H. Linget, J. F. Roch, F. Treussart, S. Chemerisov, W. Gawlik, and D. Budker. Diamonds with a high density of nitrogen-vacancy centers for magnetometry applications. *Phys. Rev. B*, 80:115202, Sep 2009.

[75] Florian Marquardt. Machine learning for physicists, 2017.

[76] Michael A Nielsen. *Neural Networks and Deep Learning*. Determination Press, 2015.

[77] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. Adaptive computation and machine learning. MIT Press, 2016.

[78] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *ArXiv e-prints*, December 2014.

[79] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533 EP –, 10 1986.

[80] Y LeCun and C Cortes. The mnist database., 2010.

[81] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[82] Tensorflow MNIST tutorial. `https://www.tensorflow.org/get_started/mnist/beginners`.

[83] Fei-Fei Li. Convolutional neural networks for visual recognition, 2017.

[84] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[85] Yoshua Bengio et al. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.

[86] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 09 2017.

[87] Jonathan Romero, Jonathan P Olson, and Alan Aspuru-Guzik. Quantum autoencoders for efficient compression of quantum data. *Quantum Science and Technology*, 2(4):045001, 2017.

[88] Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. 8, 05 1992.

[89] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 01 2016.

[90] Satinder Singh, Andy Okun, and Andrew Jackson. Artificial intelligence: Learning to play go from scratch. *Nature*, 550(7676):336–337, 10 2017.

[91] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017.

[92] Kyle Mills, Michael Spanner, and Isaac Tamblyn. Deep learning and the Schrödinger equation. *Phys. Rev. A*, 96:042113, Oct 2017.

[93] Dong-Ling Deng, Xiaopeng Li, and S. Das Sarma. Quantum entanglement in neural network states. *Phys. Rev. X*, 7:021021, May 2017.

[94] Giacomo Torlai and Roger G. Melko. Neural decoder for topological codes. *Phys. Rev. Lett.*, 119:030501, Jul 2017.

[95] C. Granade, C. Ferrie, I. Hincks, S. Casagrande, T. Alexander, J. Gross, M. Kononenko, and Y. Sanders. Qinfer: Statistical inference software for quantum applications. *ArXiv e-prints*, October 2016.

[96] B. D'Anjou, L. Kuret, L. Childress, and W. A. Coish. Maximal adaptive-decision speedups in quantum-state readout. *Phys. Rev. X*, 6:011017, Feb 2016.

[97] C. L. Degen, F. Reinhard, and P. Cappellaro. Quantum sensing. *Rev. Mod. Phys.*, 89:035002, Jul 2017.

[98] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum metrology. *Phys. Rev. Lett.*, 96:010401, Jan 2006.

[99] V. B. Braginsky and F. Ya. Khalili. Quantum nondemolition measurements: the route from toys to tools. *Rev. Mod. Phys.*, 68:1–11, Jan 1996.

[100] H Cramer. *Mathematical methods of statistics*. Princeton University Press, 1954.

[101] Ronald Aylmer Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222:309–368, 1922.

[102] F. Verstraete, A. C. Doherty, and H. Mabuchi. Sensitivity optimization in quantum parameter estimation. *Phys. Rev. A*, 64:032111, Aug 2001.

[103] Mankei Tsang. Continuous quantum hypothesis testing. *Phys. Rev. Lett.*, 108:170502, Apr 2012.

[104] P Warszawski and H M Wiseman. Quantum trajectories for realistic photodetection: I. general formalism. *Journal of Optics B: Quantum and Semiclassical Optics*, 5(1):1, 2003.

[105] P Warszawski and H M Wiseman. Quantum trajectories for realistic photodetection: Ii. application and analysis. *Journal of Optics B: Quantum and Semiclassical Optics*, 5(1):15, 2003.

[106] P. Warszawski, H. M. Wiseman, and H. Mabuchi. Quantum trajectories for realistic detection. *Phys. Rev. A*, 65(2):023802, February 2002.

[107] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 10 1986.

[108] Vedran Dunjko, Jacob M. Taylor, and Hans J. Briegel. Quantum-enhanced machine learning. *Phys. Rev. Lett.*, 117:130501, Sep 2016.

[109] S Gammelmark and K Mølmer. Quantum learning by measurement and feedback. *New Journal of Physics*, 11(3):033017, 2009.

[110] Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters. Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels. *Phys. Rev. Lett.*, 70:1895–1899, Mar 1993.

[111] Daniel Gottesman and Isaac L Chuang. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature*, 402(6760):390–393, 1999.

[112] Nicolas Gisin, Grégoire Ribordy, Wolfgang Tittel, and Hugo Zbinden. Quantum cryptography. *Rev. Mod. Phys.*, 74:145–195, Mar 2002.

[113] Dik Bouwmeester, Jian-Wei Pan, Klaus Mattle, Manfred Eibl, Harald Weinfurter, and Anton Zeilinger. Experimental quantum teleportation. *Nature*, 390(6660):575–579, 1997.

[114] Akira Furusawa, Jens Lykke Sørensen, Samuel L Braunstein, Christopher A Fuchs, H Jeff Kimble, and Eugene S Polzik. Unconditional quantum teleportation. *Science*, 282(5389):706–709, 1998.

[115] Mark Riebe, H Häffner, CF Roos, W Hänsel, J Benhelm, GPT Lancaster, TW Körber, C Becher, F Schmidt-Kaler, DFV James, et al. Deterministic quantum teleportation with atoms. *Nature*, 429(6993):734–737, 2004.

[116] MD Barrett, J Chiaverini, T Schaetz, J Britton, WM Itano, JD Jost, E Knill, C Langer, D Leibfried, R Ozeri, et al. Deterministic quantum teleportation of atomic qubits. *Nature*, 429(6993):737–739, 2004.

[117] Jacob F Sherson, Hanna Krauter, Rasmus K Olsson, Brian Julsgaard, Klemens Hammerer, Ignacio Cirac, and Eugene S Polzik. Quantum teleportation between light and matter. *Nature*, 443(7111):557–560, 2006.

[118] Lars Steffen, Yves Salathe, Markus Oppliger, Philipp Kurpiers, Matthias Baur, Christian Lang, Christopher Eichler, Gabriel Puebla-Hellmann, Arkady Fedorov, and Andreas Wallraff. Deterministic quantum teleportation with feed-forward in a solid state system. *Nature*, 500(7462):319–322, 2013.

[119] Wolfgang Pfaff, BJ Hensen, Hannes Bernien, Suzanne B van Dam, Machiel S Blok, Tim H Taminiau, Marijn J Tiggelman, Raymond N Schouten, Matthew Markham, Daniel J Twitchen, et al. Unconditional quantum teleportation between distant solid-state quantum bits. *Science*, 345(6196):532–535, 2014.

[120] Shuntaro Takeda, Takahiro Mizuta, Maria Fuwa, Peter van Loock, and Akira Furusawa. Deterministic quantum teleportation of photonic quantum bits by a hybrid technique. *Nature*, 500(7462):315–318, 2013.

[121] E. T. Holland, B. Vlastakis, R. W. Heeres, M. J. Reagor, U. Vool, Z. Leghtas, L. Frunzio, G. Kirchmair, M. H. Devoret, M. Mirrahimi, and R. J. Schoelkopf. Single-photon-resolved cross-kerr interaction for autonomous stabilization of photon-number states. *Phys. Rev. Lett.*, 115:180501, Oct 2015.

[122] Io-Chun Hoi, Anton F. Kockum, Tauno Palomaki, Thomas M. Stace, Bixuan Fan, Lars Tornberg, Sankar R. Sathyamoorthy, Göran Johansson, Per Delsing, and C. M. Wilson. Giant cross–kerr effect for propagating microwaves induced by an artificial atom. *Phys. Rev. Lett.*, 111:053601, Aug 2013.

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

[124] Asher Peres. Delayed choice for entanglement swapping. *Journal of Modern Optics*, 47(2-3):139–143, 2000.

[125] Xiao-song Ma, Johannes Kofler, and Anton Zeilinger. Delayed-choice gedanken experiments and their realizations. *Rev. Mod. Phys.*, 88:015005, Mar 2016.

[126] Andreas Wallraff, David I Schuster, Alexandre Blais, L Frunzio, R-S Huang, J Majer, S Kumar, Steven M Girvin, and Robert J Schoelkopf. Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics. *Nature*, 431(7005):162–167, 2004.

[127] Alexandre Blais, Ren-Shou Huang, Andreas 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, Jun 2004.

[128] M Brune, S Haroche, JM Raimond, L Davidovich, and N Zagury. Manipulation of photons in a cavity by dispersive atom-field coupling: Quantum-nondemolition measurements and generation of "Schrödinger cat" states. *Physical Review A*, 45(7):5193, 1992.

[129] Brian Vlastakis, Gerhard Kirchmair, Zaki Leghtas, Simon E Nigg, Luigi Frunzio, Steven M Girvin, Mazyar Mirrahimi, Michel H Devoret, and Robert J Schoelkopf. Deterministically encoding quantum information using 100-photon Schrödinger cat states. *Science*, 342(6158):607–610, 2013.

[130] Brian Vlastakis, Andrei Petrenko, Nissim Ofek, Luyan Sun, Zaki Leghtas, Katrina Sliwa, Yehan Liu, Michael Hatridge, Jacob Blumoff, Luigi Frunzio, et al. Characterizing entanglement of an artificial atom and a cavity cat state with bell's inequality. *Nature communications*, 6, 2015.

[131] Sangchul Oh, Soonchil Lee, and Hai-woong Lee. Fidelity of quantum teleportation through noisy channels. *Phys. Rev. A*, 66:022316, Aug 2002.

[132] Gabriel G. Carlo, Giuliano Benenti, and Giulio Casati. Teleportation in a noisy environment: A quantum trajectories approach. *Phys. Rev. Lett.*, 91:257903, Dec 2003.

[133] Matteo G A Paris, Mary Cola, and Rodolfo Bonifacio. Remote state preparation and teleportation in phase space. *Journal of Optics B: Quantum and Semiclassical Optics*, 5(3):S360, 2003.

[134] Sebastian G. Hofer, Witlef Wieczorek, Markus Aspelmeyer, and Klemens Hammerer. Quantum entanglement and teleportation in pulsed cavity optomechanics. *Phys. Rev. A*, 84:052327, Nov 2011.

[135] Benjamin Schumacher. Sending entanglement through noisy quantum channels. *Phys. Rev. A*, 54:2614–2628, Oct 1996.

[136] Serge Haroche and Jean Michel Raimond. *Exploring the quantum.* Oxford Univ. Press, 2006.

[137] JP Reithmaier, G Sek, A Löffler, C Hofmann, S Kuhn, S Reitzenstein, LV Keldysh, VD Kulakovskii, TL Reinecke, and A Forchel. Strong coupling in a single quantum dot–semiconductor microcavity system. *Nature*, 432(7014):197–200, 2004.

[138] J. M. Gambetta, A. A. Houck, and Alexandre Blais. Superconducting qubit with purcell protection and tunable coupling. *Phys. Rev. Lett.*, 106:030502, Jan 2011.

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

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

[141] 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, Mar 1989.

[142] MA Castellanos-Beltran and KW Lehnert. Widely tunable parametric amplifier based on a superconducting quantum interference device array resonator. *Applied Physics Letters*, 91(8):083509, 2007.

[143] M. Hatridge, R. Vijay, D. H. Slichter, John Clarke, and I. Siddiqi. Dispersive magnetometry with a quantum limited squid parametric amplifier. *Phys. Rev. B*, 83:134501, Apr 2011.

[144] ZR Lin, K Inomata, WD Oliver, K Koshino, Y Nakamura, JS Tsai, and T Yamamoto. Single-shot readout of a superconducting flux qubit with a flux-driven josephson parametric amplifier. *Applied Physics Letters*, 103(13):132602, 2013.

[145] Leonard E. Baum and Ted Petrie. Statistical inference for probabilistic functions of finite state markov chains. *Ann. Math. Statist.*, 37(6):1554–1563, 12 1966.

[146] Yakir Aharonov, Peter G. Bergmann, and Joel L. Lebowitz. Time symmetry in the quantum process of measurement. *Phys. Rev.*, 134:B1410–B1416, Jun 1964.

[147] Yakir Aharonov, David Z. Albert, and Lev Vaidman. How the result of a measurement of a component of the spin of a spin- *1/2* particle can turn out to be 100. *Phys. Rev. Lett.*, 60:1351–1354, Apr 1988.

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

[149] John Watrous. *QIC 890 - Haar Measure in Quantum Information Theory.* 2014.

[150] N. Roch, M. E. Schwartz, F. Motzoi, C. Macklin, R. Vijay, A. W. Eddins, A. N. Korotkov, K. B. Whaley, M. Sarovar, and I. Siddiqi. Observation of measurement-induced entanglement and quantum trajectories of remote superconducting qubits. *Phys. Rev. Lett.*, 112:170501, Apr 2014.

[151] A. Narla, S. Shankar, M. Hatridge, Z. Leghtas, K. M. Sliwa, E. Zalys-Geller, S. O. Mundhada, W. Pfaff, L. Frunzio, R. J. Schoelkopf, and M. H. Devoret. Robust concurrent remote entanglement between two superconducting qubits. *Phys. Rev. X*, 6:031036, Sep 2016.

[152] Alexander N. Korotkov. Flying microwave qubits with nearly perfect transfer efficiency. *Phys. Rev. B*, 84:014510, Jul 2011.

[153] J. Wenner, Yi Yin, Yu Chen, R. Barends, B. Chiaro, E. Jeffrey, J. Kelly, A. Megrant, J. Y. Mutus, C. Neill, P. J. J. O'Malley, P. Roushan, D. Sank, A. Vainsencher, T. C. White, Alexander N. Korotkov, A. N. Cleland, and John M. Martinis. Catching time-reversed microwave coherent state photons with 99.4% absorption efficiency. *Phys. Rev. Lett.*, 112:210501, May 2014.

[154] M. Pechal, L. Huthmacher, C. Eichler, S. Zeytinoğlu, A. A. Abdumalikov, S. Berger, A. Wallraff, and S. Filipp. Microwave-controlled generation of shaped single photons in circuit quantum electrodynamics. *Phys. Rev. X*, 4:041010, Oct 2014.

[155] M. Halder, A. Beveratos, N. Gisin, V. Scarani, C. Simon, and H. Zbinden. Entangling independent photons by time measurement. *Nature Physics*, 3:692–695, October 2007.

[156] Nicolas Sangouard, Christoph Simon, Hugues de Riedmatten, and Nicolas Gisin. Quantum repeaters based on atomic ensembles and linear optics. *Rev. Mod. Phys.*, 83:33–80, Mar 2011.

[157] Felix Motzoi, K. Birgitta Whaley, and Mohan Sarovar. Continuous joint measurement and entanglement of qubits in remote cavities. *Phys. Rev. A*, 92:032308, Sep 2015.

[158] Daniel A Lidar and Todd A Brun. *Quantum error correction.* Cambridge University Press, 2013.

[159] Robert Raussendorf and Hans J. Briegel. A one-way quantum computer. *Phys. Rev. Lett.*, 86:5188–5191, May 2001.

[160] Debbie W Leung. Two-qubit projective measurements are universal for quantum computation. *arXiv preprint quant-ph/0111122*, 2001.

[161] Michael A Nielsen. Quantum computation by measurement and quantum memory. *Physics Letters A*, 308(2):96–100, 2003.

[162] Michel H Devoret and Robert J Schoelkopf. Superconducting circuits for quantum information: an outlook. *Science*, 339(6124):1169–1174, 2013.

[163] Ying Li and Simon C Benjamin. Hierarchical surface code for network quantum computing with modules of arbitrary size. *arXiv preprint arXiv:1509.07796*, 2015.

[164] H-J Briegel, Wolfgang Dür, Juan I Cirac, and Peter Zoller. Quantum repeaters: the role of imperfect local operations in quantum communication. *Physical Review Letters*, 81(26):5932, 1998.

[165] Nicolas Sangouard, Christoph Simon, Hugues De Riedmatten, and Nicolas Gisin. Quantum repeaters based on atomic ensembles and linear optics. *Reviews of Modern Physics*, 83(1):33, 2011.

[166] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J Love, Alán Aspuru-Guzik, and Jeremy L O'brien. A variational eigenvalue solver on a photonic quantum processor. *Nature communications*, 5, 2014.

[167] Charles H Bennett, David P DiVincenzo, John A Smolin, and William K Wootters. Mixed-state entanglement and quantum error correction. *Physical Review A*, 54(5):3824, 1996.

[168] R. Hanbury Brown and R. Q. Twiss. Correlation between photons in two coherent beams of light. *Nature*, 177:27–29, 1956.

[169] G. T. Foster, L. A. Orozco, H. M. Castro-Beltran, and H. J. Carmichael. Quantum state reduction and conditional time evolution of wave-particle correlations in cavity qed. *Phys. Rev. Lett.*, 85:3149–3152, Oct 2000.

[170] S. Gerber, D. Rotter, L. Slodička, J. Eschner, H. J. Carmichael, and R. Blatt. Intensity-field correlation of single-atom resonance fluorescence. *Phys. Rev. Lett.*, 102:183601, May 2009.

[171] E. R. Marquina-Cruz and H. M. Castro-Beltran. Nonclassicality of resonance fluorescence via amplitude-intensity correlations. *Laser Physics*, 18(2):157–164, 2011.

[172] P. Facchi, H. Nakazato, and S. Pascazio. From the quantum zeno to the inverse quantum zeno effect. *Phys. Rev. Lett.*, 86:2699–2703, Mar 2001.

[173] Xiao-Qiang Shao, Li Chen, Shou Zhang, and Kyu-Hwang Yeon. Fast cnot gate via quantum zeno dynamics. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 42(16):165507, 2009.