Im Fachbereich Physik der Freien Universität Berlin eingereichte Dissertation zur Erlangung des Grades Dr. rer. nat.

Ideas on System Identification

Janina Rustaee Khoshkbijari

2016

Gutachter: **Prof. Dr. Jens Eisert**
Prof. Dr. Felix von Oppen

Dissertation am 20.02.2017

Statement

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person, except where due acknowledgement is made in the text. I also declare that the intellectual content of this thesis is the product of my own work, even though I may have received assistance from others on style, presentation and language expression.

Erklärung

Hiermit versichere ich, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe. Die Arbeit enthält, nach meinem besten Wissen und Gewissen, weder bereits von anderen Personen veröffentlichte Materialien noch wurden jegliche Teile dieser Arbeit bereits einer anderen Universität oder ähnlichen Bildungsstätte zur Erlangung eines Abschlusses oder Titels vorgelegt, außer entsprechendes Material ist als solches gekennzeichnet.

Berlin, 1st of September 2016

Acknowledgements

I would like to express my deep gratitude to Professor Jens Eisert, my research supervisor, for his patient guidance, enthusiastic discussions and useful critiques of this research work. I would also like to express my thankfulness to Dr. Carlos Riofrío, Dr. Mathis Friesdorf, Stephan Wäldchen and Dr. Earl T. Campbell with whom I collaborated on the research projects presented in this thesis. Special thanks to Carlos for being not only an office mate but also a good friend in tough times. Thanks to the whole QMIO group in Dahlem, with Annette Schumann-Welde leading the way for always having our back. Many thanks to the the team of ZEDV, for countless entertaining lunches and for protecting our data from loss of information. This work was kindly funded by the BMBF (QuORep, Q.com).
I would like to extend my thanks to Jörg Behrmann—who also spent hours proofreading—and Nils Krane, you are great friends. I will keep you and our university breakfasts in my mind forever. Finally, I want to thank my parents, my brothers, and my husband for their endless encouragement and mental support throughout my studies. Thank you for always being by my side and continuously making me want to become a better person.

Contents

Zusammenfassung

Quantenmechanik erlaubt uns, das Verhalten physikalischer Systeme in einer Vielzahl experimenteller Situationen verlässlich vorherzusagen. In diesem Sinne ermöglicht sie uns, eine Intuition für das Verhalten mikroskopischer Systeme zu entwickeln, so wie die Newtonsche Mechanik uns das Verständnis der Prinzipien makroskopischer Bewegungen erlaubte. Wir wissen, dass Manipulationen und Messungen quantenmechanischer Systeme diese direkt beeinflussen und die Ergebnisse verändern können. Auf der anderen Seite sind zur Identifikation und Steuerung der mikroskopischen Quantensysteme hochkomplexe experimentelle Aufbauten nötig, die das genaue Verständnis der Vorgänge bei der Messung erschweren.

Diese Arbeit hat das Ziel, Ideen zur effizienten Kontrolle von Quantensystemen zu evaluieren und zu erweitern. Der Fokus liegt dabei auf dem mit der Messung einhergehenden Informationsverlust und dem möglichen Einfluss von Annahmen über das zu messende Quantensystem. Neuartige Wege der Beschreibung und der Rekonstruktion von Quantenzuständen und Quantenoperationen werden auf reale Messergebnisse angewendet um einige der bestehenden Gräben zwischen theoretischen und experimentellen Herangehensweisen an die Rekonstruktion von Quantensystemen zu schließen, und dabei neue Hindernisse zu identifizieren.

Eine weiteres Ziel dieser Arbeit ist das Bewerben eines informationstheoretischen Ansatzes bei der Beschreibung von Quantensystemen. Eine klare mathematische Beschreibung der im System vorhandenen, gesuchten Information erlaubt nicht nur die Analyse der Transformationen und Verluste dieser Information, sondern ermöglicht dadurch auch die Bestimmung der für verschiedene Messvorgänge erlangbaren Erkenntnisse. Diese Ansätze könnten auch zu neuen Ideen der Identifikation und Transformation klassischer Systeme und Netzwerke beitragen.

Entsprechend dieser Motivation eröffnen wir neue Perspektiven auf Messtechniken sowie die Interpretation von Messergebnissen in der Anwendung und Erweiterung neuer Methoden zur Umkehrung von mathematischen Faltungen auf fehlerbehaftete Datensätze aus Messungen von Atomen in optischen Gittern. Wir führen weiterhin sinnvolle Wege der Messung spezieller Eigenschaften von Quantenzuständen—ihre Nähe zu einem Gaussschen Zustand—aus unvollständigen Messdaten ein. Zuletzt führen wir ein neuartiges Protokoll ein um die einem stark korrelierten System innewohnende Verschränkung möglichst nachhaltig nutzbar zu machen.

Es ist besonders dieser letzte Ansatz, der zu der Vision der Realisierung eines bedingungslos sicheren Kommunikationsnetzwerkes beitragen kann. Dieses ermöglichte uns im “Quantum Internet of Things” ein sicheres Netzwerk zwischen intelligenten Geräten aufzubauen und brächte uns so der Welt der Science Fiction ein kleines Stück näher.

Abstract

Quantum mechanics allows us to reliably predict the behaviour of physical systems in a plethora of experimental situations. In this sense, it enables us to build an intuition for microscopic systems as Newtonian mechanics helped us understand the principles of macroscopic motion. We know that our actions and measurements on quantum mechanical systems directly influence our results and have the ability to change outcomes. On the other hand, highly complex experimental setups are necessary to identify and manipulate these microscopic quantum systems, which makes a complete understanding of the measurement process a hard process.

This work aims to test and expand ideas on how this control of quantum systems can be done in careful and efficient ways. It focuses on the loss of information during measurements and the possible influence of presumptions about the quantum state in question. Modern ways of describing and reconstructing quantum states and quantum operations are applied to real experimental situations to fill in some of the gaps—and of course point at new gaps—between theoretical and experimental approaches to present day quantum systems.

This work further aims to advertise an information theoretic mindset in the approach to quantum mechanics. A clear description of the wanted information contained in the system in question not only allows to investigate the transformations and the losses this particular information undergoes, but also helps determining which part of the wanted information is theoretically accessible for different kinds of measurement. As system sizes grow and new technologies like the internet of things introduce the need for faster and faster processing, principles from quantum information theory can offer important leads to new ideas on system identification and transformation.

We develop a number of ideas following this mindset. Firstly, we offer new perspectives on measurement techniques as well as the interpretation of results by applying several blind deconvolution techniques to experimental data of atoms in optical lattices. We then introduce meaningful measures to gain insight on special properties of quantum states—their closeness to a Gaussian state—from only very limited partial measurement data. Lastly we introduce a novel protocol to transform the entanglement content in a highly correlated state to be able to use this resource in the most sustainable way.

It is this last part in particular, that points at a possible contribution of this work to the vision of realising an unconditionally secure communication network. This would enable us to use quantum communication in a quantum internet of things—a secure communication network between intelligent devices—taking us one step closer to our wildest science fiction dreams.

Part I.

Introduction

1. Introduction

> . . . no tool appears better calibrated for a direct assault [on the foundations of quantum mechanics] than quantum information theory. Far from a strained application of the latest fad to a time-honoured problem, this method holds promise precisely because a large part—but not all—of the structure of quantum theory has always concerned information. It is just that the physics community needs reminding. (Fuchs, 2002)

The experimental revolutions in the field of quantum mechanics, achieved in the last century and especially the last decades, are mind-blowing. Aspects, that were once seen as paradoxical and are still hard to grasp, like Schrödinger's cat or the Einstein Podolski Rosen paradox, are verified over and over again and taken as true predictions of quantum mechanics [9, 25, 35].
The field of quantum optics served and continues to serve as a catalyst, as the ability to control the quantum state of light provides an ideal starting point for promising experiments that can test the properties and predictions of quantum mechanics where we lack the imagination [87]. Despite these exhaustive tests, quantum mechanics proved to be correct and no borders to its validity have been found, even though there remain tons of open questions [45, 93].
As our trust in quantum mechanics grows, we can explore the new concepts that this framework offers us. Exploring quantum communication and exploiting the possibilities of quantum measurement to identify the most efficient and economical way within the boundaries of quantum mechanics brought to life the field of quantum information theory and quantum cryptography [50].
Quantum information theory uses the methods of information theory, which have developed over the past decades in parallel to computer science. Already in the late 1940's, Claude Shannon contributed to the questions of how information can be defined, quantified or transmitted over a channel of either time (memory) or space [88]. Following in the footsteps of Shannon, quantum information theory introduced quantum analogues of Shannon's theorems by looking into how strongly quantum information can be compressed, and how noise affects the transmission of quantum information [69].
There is a deep fascination in learning more about systems than we—on the first glance—should be able to. In 2014, a method of nanoscopy, that makes it possible to distinguish

molecules below the diffraction limit was rewarded the Nobel prize. Instead of searching for information, where there is none, special properties of the system and new approaches to measurement were used that introduce a new kind of information to the system, which is then used to get a much higher resolution [57]. In a similar way, new approaches to measurement and manipulation of quantum states revolutionise the efficiency and reliability of operations in the quantum world beyond imagination.
This work aims to test and expand ideas on how quantum systems can be identified and controlled in more efficient ways in order to approach real life applications in labs. It also focuses on where information is lost during our approaches to learn about quantum systems, and where measurement or reconstruction approaches might destroy or—even worse—introduce information into unknown quantum systems. Several modern ways of describing and reconstructing quantum states and quantum operations are applied to real experimental situations to fill in some of the gaps—and of course point at new gaps—between theoretical and experimental approaches to present day quantum systems. Overcoming these gaps is crucial in order to be able to "benefit from the advances in quantum technologies" into our everyday lives, a goal not only listed in the recently released "Quantum Manifesto" but also bringing us on step closer to our wildest dreams of science fiction.

After this general introduction, we will introduce different possibilities in the descriptions of quantum states. The focus herein lies in describing Gaussian states as well as low-rank states and Matrix Product states and operators, as they will all play a major role in the later chapters.

In the following chapter three, different ways of identifying quantum systems and reconstructing quantum states are described. The chapter also features an introduction of the compressed sensing method, which will be central to chapter five. In the last introductory chapter, we will give a short overview on the tools of quantum and classical information theory that will appear in this thesis. We will introduce the concepts of information content, entropy and especially relative entropy.

In chapter five, compressed sensing techniques are used to perform a non-linear deconvolution of fluorescence peaks to efficiently and reliably reconstruct atom positions. This work was done in collaboration with experimentalists from Bonn, performing fluorescence imaging on optical lattices in order to distinguish nearest neighbour atoms. We use several variants of compressed sensing and discuss how compressed sensing can in general be applied to these and similar experimental situations. We perform reconstruction on simulated as well as experimental data and present benchmark results on reconstructabil-

ity containing a newly introduced error model. We also discuss possibilities of improving the distinguishability of close-by atoms and reducing information loss during the measurement process.

In chapter six, instead of reconstructing a low-rank estimate of a certain quantum state, we discuss the possibility of reconstructing more general properties of quantum states. This work was done in collaboration with experimentalists from Munich, performing time-of-flight measurements on cold atoms originating from optical lattices. We investigate the loss of information during time-of-flight measurements and discuss to what extent this information is sufficient to reconstruct properties of the measured states. We introduce the useful property of local deviation from Gaussianity and derive bounds on it that can be estimated with only very limited information.

In chapter seven and eight, rather than performing tomography and learning about a state, we dive into methods to manipulate quantum states to arrive at a wanted outcome state with given properties. Entanglement is the key resource in many quantum information protocols such as quantum key distribution or quantum communication. As local operations and classical communication cannot enlarge the amount of entanglement between two parties, we investigate the possibility of transforming the entanglement between subsequent pairs from a entangling source into wanted entanglement between the two pairs. We investigate whether the rates of generating highly entangled pairs can be improved by treating the entanglement that is already present in the system in the most sustainable way. In chapter seven we therefore present entanglement distillation protocols, the postselective recurrence protocol and the deterministic 5-qubit code, an error-correcting protocol. We prepare the translation of these protocols into the world of tensor networks, which we introduce in chapter eight. There we show how sources of entangled pairs can naturally be described by Matrix Product Operators and translate protocols from quantum optics into this new language. We investigate the steps of the entanglement distillation protocols and use renormalisation techniques. We discuss physical models and present parameter spaces, where correlations between subsequent pairs can indeed be of advantage.

In chapter nine, we summarise our conclusions and present our outlook on future applications of treating information present in a system like the precious resource it is.

To make the main text more readable, the mathematical proofs for chapter eight as well as parts of the numerical code developed for the work presented in chapters five and six is commented and presented in the appendix.

Part II.

State of the Art

2. Descriptions of Quantum States

One of the reasons using quantum mechanics in future communications and computations is, that the Hilbert space underlying the quantum states grows exponentially and allows for complex states like highly entangled states. This can introduce exponential speed-ups in quantum computation and quantum simulation. However, this possibility also bears the serious disadvantage that due to this complexity it can be very hard to describe a generic quantum system efficiently. The large amount of degrees of freedom makes a full tomography and a full description of the state a difficult and inefficient task. However, there exist plenty of quantum systems, where this description can be done more efficiently. By concentrating on a smaller corner in Hilbert space, one can already cover a huge class of physically meaningful settings and states. In this chapter, we concentrate on three different descriptions, that allow us to describe or reconstruct states in different corners of Hilbert space. The first set of states we will focus on is that of Gaussian states. These are the states that have a Gaussian Wigner function and can therefore be efficiently described in phase space. The second class of states are low-rank states, which are quantum states which can be represented by density matrices that have only a few eigenvalues different from zero. By finding the non-trivial eigenspace of these states, much less information is needed for a full description. The last set of states is a huge class, that can in its limits cover the whole of Hilbert space. The class of tensor network states, or the special case of matrix product operators provide a natural description of a system that has mainly local interactions. The so-called bond dimension can be understood as a backdoor to allow for more complicated interactions, but small bond dimensions already cover a large set of meaningful states.

We will now focus on these subspaces in more detail. Afterwards, we will investigate how using additional knowledge on the subspace of quantum states can be used to improve the efficiency of reconstructing information about the quantum state from measurements.

2.1. Finite Dimensions

Every quantum state corresponds to a density operator satisfying three characteristics: it is positive semi-definite, Hermitian, and its trace is one.

$$\rho \geq 0, \qquad \rho = \rho^\dagger, \qquad \mathrm{Tr}\,\rho = 1$$

In this density matrix picture, all state operations can be thought of as maps transforming one density matrix into another one. These operations are called quantum channels and have to conserve the three properties that define a quantum state. One possible allowed operation is the convex combination of quantum states. The outcome of such a convex combination on pure input states is called a mixed state. Every quantum state can be decomposed into a basis of pure states. This decomposition is however not unique, which is one of the reasons why certain properties like entanglement are difficult to define when moving from the world of pure to mixed quantum states. One often has to implement additional knowledge about the system in question, such as a physical reason for a natural decomposition into specific subspaces, to find a reasonable measure for entanglement. We will treat different possible measures of entanglement in a following chapter.

Before we consider several different decompositions of quantum states, we will mathematically define a quantum channel, as we will use this concept in a later section. Quantum channels are completely positive, trace-preserving, linear maps, as they map quantum states to quantum states. They can generally be formulated as

$$\rho \mapsto T(\rho) = \mathrm{Tr_E}[U(\rho \otimes \rho_E)U^\dagger]. \tag{2.1}$$

where the system is enlarged to include an environment, unitarily transformed and reduced to the original system by tracing out the environment again. Any quantum operation can thus be seen as a unitary transformation on a bigger system which is not completely accessible. This quantum channel picture makes it thus possible to cover all kinds of possible quantum transformations such as unitary operations, changes in the system size which include compositions with other systems, partial traces as well as projective measurements, in one single mathematical formulation. This is exceptionally useful in the description of quantum information protocols like entanglement distillation, as it allows the observation of the transformation of information in every step. For continuous variable systems, one has to use a different language though, as the Hilbert space in question will become infinitely dimensional. We will consider this in the next section.

2.2. Gaussian States

If we move away from the realm of qubits and finite-dimensional quantum systems and consider continuous variables instead, most properties like entanglement are in general

hard to bound let alone to describe. A large number of protocols and manipulations of quantum states can however be described by the evolution of the moments of the distributions of the states. One special set of states, the so-called Gaussian states are fit to represent a huge class of physical states. Coherent states or squeezed states, but also thermal or Gibbs states are famous examples for Gaussian states. They all share the property of having a Gaussian Wigner function, which itself is the Fourier transform of the characteristic function. We will introduce this setup, as we will use it in chapter six, where we develop a method to bound the closeness of a specific quantum state to the set of Gaussian states using only very limited information about the quantum state in question. The topic of Gaussian states is widely represented in the literature, we thus refer to Nielsen and Chuang [70] for a more thorough introduction than is given in the following.

The most common example for a system with continuous variables and an infinitely dimensional Hilbert space is the quantum harmonic oscillator. Its coordinates are the continuous functions for position X and momentum P, which can be described using annihilation and creation operators, the product of which forms the number operator, which shares its eigenbasis $|n\rangle$ with the Hamiltonian of the quantum harmonic oscillator. The first example of a Gaussian state is the coherent state, which is nothing but the eigenstate of the annihilation operator, $a\,|\alpha\rangle = \alpha\,|\alpha\rangle$:

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

If we now consider not only one degree of freedom, say a single mode, but N degrees of freedom, we can describe them as N harmonic oscillators and describe the state in the phase space of these oscillators, which is of dimension $\mathbb{R}^{2N}$. As we are dealing with harmonic oscillators, we know that there exist commutation relations in every mode which gives us additional knowledge on a property of the phase space.

The reason why this phase space picture is helpful is, that as observables of different modes commute, we can reduce the complexity of the description of the system by considering the single mode problems independent of each other. To describe either states or operations in phase space, one makes use of the Wigner function $\mathcal{W}$, a normalised and real-valued function. It is in general not a probability distribution, as its values can be negative, but can be used to determine expectation values of products of the canonical coordinates, called moments of the distribution.

$$\left\langle X^a P^b \right\rangle_\rho = \int_\infty^\infty \mathrm{d}x\,\mathrm{d}p\, x^a p^b \mathcal{W}_\rho(x,p).$$

The Wigner function is derived as the Fourier transform of the characteristic function, which is itself defined as the expectation value of the displacement operator in phase

space.

There are certain functions, for which it is exceptionally useful to describe them only in terms of their moments and the evolution of these. Wick's theorem, based on Isserli's formula, tells us, that for example Gaussian states are completely defined by their first and second moments alone [59, 102]. The first moments are the expectation values of the canonical coordinates, i.e. the displacement in phase space. The second moments contain all possible combinations of two $R = X, P$ from all modes, which can be cast in a real symmetric $2N \times 2N$ matrix, called the covariance matrix γ.

$$\gamma_{i,k} = 2\mathrm{Re}\left\langle (R_j - \langle R_j \rangle_\rho)(R_k - \langle R_k \rangle_\rho) \right\rangle_\rho, \quad j,k = 1,\ldots,N.$$

The fact that $2N$ is always even has an important physical implication. Williamson's theorem [103] states, that any n-dimensional real symmetric matrix with an even dimension can be diagonalised using a so-called symplectic transformation. Symplectic transformations are of significant physical interest, as they are transformations of the canonical operators which preserve the commutation relations. Symplectic diagonalisation of the before-mentioned covariance matrix γ leads to its symplectic eigenvalues, which appear as the eigenvalues of the matrix $i\sigma^{-1}\gamma$, where σ is the symplectic form.

$$\sigma = \bigoplus_{k=1}^{n} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{2.2}$$

For more detail, we refer to the constructive proof of this theorem by Narcovich [68].

This transformation into independent modes allows a very efficient description of Gaussian states. Their importance in many experimental setups in the description of coherent or thermal states makes them an important building block in the description of quantum systems.

A general Gaussian state is defined by its Gaussian Wigner distribution

$$\mathcal{W} = C \cdot \exp\left(-\frac{1}{2}(\mathbf{r}-\mathbf{d})^T \gamma^{-1} (\mathbf{r}-\mathbf{d})\right). \tag{2.3}$$

C is a normalisation constant and $\mathbf{d}$ is the displacement vector, e.g. the expectation value of the position. It is the first moment of the distribution. Using the symplectic transformation one can show that there always exists a canonical basis such that the Gaussian state factorises into single modes belonging to different temperatures.

One last special property of Gaussian states we want to introduce is their von Neumann entropy. Within the set of all density operators with fixed finite first and second moments, the respective Gaussian state maximises the entropy. This can be proven by evaluating the relative entropy between the Gaussian state σ and any other state ρ of the set with

fixed first and second moments.

$$D(\rho|\sigma) = \mathrm{Tr}(\rho \ln \rho) - \mathrm{Tr}(\sigma \ln \sigma) = -S(\rho) + S(\sigma) \geq 0 \tag{2.4}$$

We will use this feature later to characterise a given distribution and evaluate their distance from the set of Gaussian states.

2.3. Low-Rank States and Matrix Product States

Even though the Hilbert space of a few ions, or even a single particle can be enormous, the vast majority of quantum states are not of physical interest. Most naturally occurring states lie in a very small corner of this Hilbert space. They are either low-rank or even pure (classical) states, ground states of local Hamiltonians that can be described as Tensor Network states, or states with special symmetries as the Gaussian states we introduced in the last section.
One physical example for low-rank states are low entropy states, that are supported on a r-dimensional subspace of the underlying d-dimensional Hilbert space. Those states arise as pure classical states subject to a local source of noise. We will see in the next section, how these states often require significantly fewer measurement settings in order to perform a full tomography than a full rank state, even if the sparse subspace is not known.
The problem of an exponentially growing Hilbert space gets even worse when regarding quantum many body states. This space contains a large number of highly entangled non-local states. However, if we restrict the states to having only local interactions—in the sense that they can only be correlated to states that are physically close to them—the correlations between them decay exponentially with their distance. One special case of these Tensor Network States (TNS) are the one-dimensional Matrix Product States (MPS), where each state is only connected to two neighbouring states and they form a one-dimensional chain of quantum states. This setting lies at the heart of the density matrix renormalisation group (DMRG) approach, and every one-dimensional ground state of a local, gapped Hamiltonian can be represented using an MPS with a constant bond dimension, the dimension that allows for correlations between neighbouring states [75]. If the bond dimension is sufficiently large, any state can be represented as a matrix product state, so MPS are a representation of states rather then a special class. However, one typically chooses the representation of MPS if the bond dimension is small, meaning that it does not grow with the number of states. A typical MPS is defined by a set of matrices $\{A_i^{[k]}\}$ with $k \in 1, \ldots, N$ being the site and a $D_k \times D_{k+1}$ matrix A corresponding to each of the kth site. A pure quantum state $\psi \in \mathcal{C}^{\otimes d^N}$ characterising N sites each of

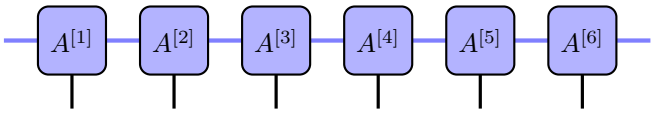

Figure 2.1.: A matrix product state in the graphical representation. Each blue block represents a set of matrices $A^{[k]}$, the blue lines represents the matrix multiplication between these matrices over the bond dimension and the black downward lines correspond to the physical indices at the respective sites.

which belongs to a d-dimensional Hilbert space can thus be written as

$$|\psi\rangle = \sum_{i_1,\ldots,i_N=1}^{d} \mathrm{Tr}\left[A_{i_1}^{[1]} A_{i_2}^{[2]} \ldots A_{i_N}^{[N]}\right] |i_1, i_2, \ldots i_N\rangle \,. \tag{2.5}$$

The bond dimensions D_k are the dimensions that appear in the product of two neighbouring matrices. In the end, the trace over the remaining matrix product defines the prefactors in a specific basis. Normalisation and other expectation values are computed by contraction of the tensors. To do that, we can use a convenient identity:

$$\mathrm{Tr}[A^* B^* C^*]\,\mathrm{Tr}[ABC] = \mathrm{Tr}[(A^* \otimes A)(B^* \otimes B)(C^* \otimes C)]. \tag{2.6}$$

Since to normalise $|\psi\rangle$, we need its scalar product with its complex conjugation, applying the former identity to the MPS representation leads to

$$\langle\psi|\psi\rangle = \sum_{i,j=1}^{d} \mathrm{Tr}\left[\prod_{k=1}^{N} \underbrace{\sum_{\{i\},\{j\}} A_i^{*[k]} A_j^{[k]}}_{E^{[k]}}\right] \langle i|j\rangle \,. \tag{2.7}$$

The matrix E is called the transfer matrix. As this formulation of the MPS representation has the obvious downside of being cumbersome and introducing a large amount of indices, MPS are most of the time represented in a graphical calculus. The translation to this graphical calculus is done by replacing matrices or tensors by boxes, and contractions of indices as bonds between those boxes, see Figure 2.1. A simpler class, which we will mostly consider in this thesis, are translationally invariant MPS, where all matrices A

are identical and thus the bond dimension D is also identical between all sites. Every translationally invariant state has such a translationally invariant MPS representation, as shown in [75]. These translationally invariant MPS allow us to store the state using only d matrices of size $D \times D$, which is very efficient compared to the original size of the Hilbert space. Furthermore, translationally invariant MPS are closely related to completely positive maps $\mathcal{E}$ acting on the space of $D \times D$ matrices: $\mathcal{E}(X) = \sum_i A_i X A_i^\dagger$. We will make use of this concept in a later chapter in the context of renormalisation of matrix product operators, the natural extension of MPS to mixed states.

2.4. Matrix Product Operators (MPO)

The introduction of MPS makes it possible to describe a large range of pure quantum states. However, to describe mixed quantum states in a similar fashion, the more complex representation of Matrix Product Operators has to be used. Here, we use it to describe a density matrix, but it can also be applied to describe parent Hamiltonians or local interactions [36, 99]. They can also be represented graphically, see Figure 2.2.

MPO representing mixed states can be obtained from MPS with a special structure by taking a partial trace, representing an individual interaction with an environment of every particle forming up the MPS. Those MPO are called locally purifiable and could be represented by an MPS with the same bond dimension but with a higher physical degree of freedom of the single particle. This can be seen as including the environment into the state, which is then represented by an additional freedom. Tracing out the environment retrieves the mixed state represented by an MPO. Graphically depicted, this process can be seen as reordering the free indices of the MPO all to one side to get a MPS with a higher number of physical indices.
We will now show, how the representation of a purifiable MPO can be derived from the MPS representation introduced in the last chapter.

$$\begin{aligned}\rho = |\psi\rangle\langle\psi| = &\sum_{i_1,\dots,i_N=1}^{d}\sum_{j_1,\dots,j_N=1}^{d} \mathrm{Tr}\left[A_{i_1}^{[1]}A_{i_2}^{[2]}\dots A_{i_N}^{[N]}\right]\mathrm{Tr}\left[A_{j_1}^{*[1]}A_{j_2}^{*[2]}\dots A_{j_N}^{*[N]}\right]\\ &\times |i_1, i_2, \dots i_N\rangle\langle j_1, j_2, \dots j_N|\\ = &\sum_{i_1,\dots,i_N=1}^{d}\sum_{j_1,\dots,j_N=1}^{d} \mathrm{Tr}\left[M_{i_1,j_1}^{[1]}M_{i_2,j_2}^{[2]}\dots M_{i_N,j_N}^{[N]}\right]|i_1, i_2, \dots i_N\rangle\langle j_1, j_2, \dots j_N|\end{aligned}$$

From now on, we will mostly assume, that the MPS are translationally invariant and drop the upper index $[k]$ for the sake of readability and only keep the physical indices. The MPO matrix M replaces the MPS matrices A and A^*. Not every possible matrix

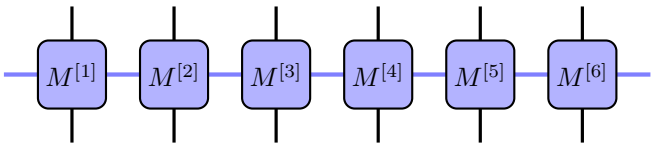

Figure 2.2.: A Matrix product operator in the graphical representation. Again, each blue block represents a set of matrices, now called $M^{[k]}$. Compared to the MPS, there are twice as many free indices.

product operator can be cast into this purifiable form. It is a computationally hard problem to decide, whether an MPO leads to a positive semi-definite state. Kliesch et al showed in [62] that it is NP-hard—it needs exponentially many calculation steps in the number of sites—and even undecidable if the chain is infinitely long. This special case of MPO however allows us to naturally describe correlations between pairs of states, offering us a new way of describing entanglement distillation protocols, as discussed in a later chapter.

3. Identification of Quantum Systems

As hinted on in the last chapter, general quantum systems can be hard to identify. Instead of doing a full tomography on a system, one often only solves an estimation problem to fit a possible outcome to the outcome of the experiment. In general these estimation problems do not have a unique solution. If however, the state that is to be reconstructed is close to a pure state or only has a low-rank, there exist techniques to approach the limits on efficient reconstructions [20, 33].

3.1. Quantum Tomography

As in classical physics, a tomography of a quantum state signifies a full description of the state's properties under measurements. In the same way distributions are generalised to quantum states, stochastic operations reflecting operations on the quantum system or measurements are replaced and generalised to quantum channels—the quantum version of communication channels sending quantum information from one party to another. By definition, such channels are required to map quantum states onto quantum states. It turns out, that a slightly stronger notion than that of positive linear operators is required: Quantum channels are trace-preserving, completely positive maps. Mathematically, they can be cast into the form

$$\rho \mapsto T(\rho) = \sum_j K_j \rho K_j^* \tag{3.1}$$

with $\sum_j K_j^* K_j = \mathbb{1}$. Not only perfect measurements or operations on the channel, but any operation, noisy or not, will be reflected by a linear map of this form. By virtue of the Choi-Jamiolkowski-isomorphism, quantum channels mapping the N-dimensional quantum states onto themselves are isomorphic to quantum states in N^2-dimensions. Measurements on quantum systems are incorporated by resolutions of the identity $P_1, \ldots, P_K$ (called positive operator valued measures: POVMs) with the property that $\sum_{k=1}^K P_k = \mathbb{1}$. Upon measurement, the probability of obtaining the specific outcome k is given by $p_k = \mathrm{Tr}(\rho P_k) = \langle \rho, P_k \rangle$. For comprehensive material on quantum channels, see for example Refs. [101, 104].

Key to the problem of performing tomography on quantum states and channels is to exploit the sparse signal paradigm, methods of compressed sensing and related ideas,

based on two principles, sparsity and incoherence.

3.2. Compressed Sensing

Compressed sensing, a classical idea from the context of matrix completion of sparse matrices, is an intriguing example how to reconstruct information in seemingly mysterious ways from incomplete data. In 2005 and 2006, Candes and Tao describe the problem as a problem of error correction, where the error vector is unknown [20]. They prove, that there are situations, in which the original signal can be reliably reconstructed. Following the same approach, Donoho proposes classes of measurement designs, tight frames, where convex optimisation leads to a reconstruction of the original system with a considerably low amount of needed measurements [33]. Later, Gross and Liu, Flammia, Becker and Eisert translated the method to the world of quantum information theory, where they apply it on the context of fairly pure states and offer a significant performance improvement on the reconstruction of quantum states [54]. All these impressive theories on reconstruction take into account the importance of reliability and offer reliable bounds on the reconstruction. The methods allow feasible reconstructions, that would otherwise be impossible to do on a classical computer.

Suppose that the objects of interest x (for example N-dimensional vectors or $N \times N$-dimensional matrices) can be well-approximated by superpositions $\sum_r c_r \psi_r$ of a few fixed pure states $\{\psi_r\}$. As compared to the full linear span of all states, this smaller set of objects has then substantially reduced complexity. However, to efficiently exploit this additional compressible structure, non-linear signal processing methods have to be used. Assume that the objects x are superpositions of exactly r different states, so they are r-sparse

$$x = \sum_{k \in \mathcal{K}} c_k \psi_k \quad \text{with} \quad |\mathcal{K}| = r. \tag{3.2}$$

Once the supporting set $\mathcal{K}$ is known, conventional signal processing methods can be used to recover x. However, for unknown $\mathcal{K}$, the whole signal set lacks linear structure and is only the union of subspaces referring to different $\mathcal{K}$'s. The classical approach consists of taking sufficiently many samples to reconstruct any x in the full span and then discarding in an adaptive way most of the samples since x was r-sparse. With the advent of compressed sensing [20, 33, 43], it has become clear that it is possible to reconstruct a sparse signal x exactly from a seemingly incomplete set of linear measurements $\langle \phi_m, x \rangle$ (e.g. the scalar product in N-dimensional vector space) and that this can be done efficiently via convex optimisation.

Since a random $N \times N$ matrix has N^2 degrees of freedom, a full tomography of this

matrix takes order $\mathcal{O}\left(N^2\right)$ measurements. However, for sparse matrices with rank r, ideal measurements would only require $\mathcal{O}(r \times N)$ to recover the full information. In compressed sensing, the number of measurement settings is remarkably close to the ideal setting. Only $\mathcal{O}(rN \log N)$ random measurement settings are needed for a reconstruction with a failure probability below $\frac{1}{2}$. By repeating the reconstruction with further copies of the same state, the reconstruction can be brought to a constant error with only $\mathcal{O}\left(r^2 N^2 \log N\right)$ copies of the state. Furthermore, the reconstruction from the outcomes of these measurements can be done using an efficient algorithm. Classical compressed sensing is widely used in image enhancement, where a sparse solution to an underdetermined system is found. This sparse solution can for example reduce unwanted noise effects in pictures. In quantum compressed sensing, the tomography of a quantum state is translated to a matrix completion problem, which can be solved by trace norm minimisation subject to convex constraints which contain the measurements.

The reason for this reduction of measurement settings to be even possible lies in the special symmetries of the underlying space of matrices. The manifold of rank-r matrices in $\mathbb{C}^{N \times N}$ can be embedded into $\mathcal{O}(rN \log N)$ dimensions with only a small distortion in the two-norm [72]. One realisation of this embedding is using the expectation values of a random subset of $\mathcal{O}(rN \log N)$ Pauli matrices. There always exists a fixed set of Pauli matrices that can be used to reconstruct any $N \times N$ matrix with rank r. In fact, most random sets fulfil this property.

In the standard form of the compressed sensing problem, linear noisy observations in a N-dimensional vector space are assumed. Let Φ be the $M \times N$ measurement matrix with the M rows $\{\phi_m\}_{m=1}^{M}$ and x be an N-dimensional vector containing the original signal. We then obtain the standard estimation equation for compressed sensing as

$$y = \Phi x + e, \tag{3.3}$$

where Φ is called the compressed sensing operator and e is a noise vector [64]. y is the M-dimensional vector representing the measured signal. While this problem is underdetermined, the goal is to recover a r-sparse vector x, i.e., a vector with only $\|x\|_0 \leq r$ non-zero entries. A key step in sparse optimisation has been achieved in Ref. [20] where it was shown that under suitable assumptions on Φ, the reconstruction is possible using convex optimisation methods. Meanwhile, this problem is well investigated in the compressed sensing community. Exemplarily, it can be shown that if $M \geq (r \log(N/r))$ and Φ is a random matrix with i.i.d. $\{+1, -1\}$-Bernoulli or sub-Gaussian distributed entries and $\|e\|_2 \leq \epsilon$, any unknown r-sparse vector x can recovered with exponential probability by solving the quadratically constrained ℓ_1-minimisation problem

$$\tilde{x} = \arg\min \|x\|_1 \quad \text{s.t.} \quad \|y - \Phi x\|_2 \leq \epsilon. \tag{3.4}$$

Here, the ℓ_1-norm appears as the convex relaxation of $\|\cdot\|_0$ (by taking the convex hull of sparse vectors). A well-known sufficient criterion on Φ to ensure sparse reconstruction by such type ℓ_1-programs is given by the restricted isometry property (RIP) [19], i.e., Φ is r-RIP if there exists $0 \leq \delta_r < 1$

$$\left| \|\Phi x\|_2^2 - \|x\|_2^2 \right| \leq \delta_r \|x\|_2^2. \tag{3.5}$$

Then, provided the unknown vector x is r-sparse it can be recovered if Φ is $2r$-RIP with $\delta_{2r} < \sqrt{2} - 1$ [19], which was improved to $\delta_{2r} \leq 3/(4 + \sqrt{6}) \approx 0.4652$ [44].

The minimisation of the ℓ_1-norm is shown to have the same optimal outcome as the minimisation of the rank, which is an NP-hard problem. The duality is connected to the trace norm being the largest convex function that puts a lower bound on the rank. [78]

The trace norm optimisation problem can be solved using standard convex optimisation techniques and results in the closest low-rank estimate of the quantum state that was measured.

When $r \ll N$, the number of measurement settings can further be reduced by applying a technique called direct fidelity estimation. Instead of performing a full tomography, the fidelity of the reconstructed low-rank estimate $\tilde{\rho}$ of the original state ρ can be estimated within an ε-region [72].

Let us consider a state $\rho \in \mathcal{H}$ describing a system of d qubits and let thus N be 2^d. Let further $\mathcal{P}$ be the set of all Pauli operators, meaning

$$\mathcal{P} = \{P | P = \sigma_1 \otimes \sigma_2 \otimes \cdots \otimes \sigma_d; \sigma_i \in \{\mathbb{1}, \sigma_x, \sigma_y, \sigma_z\}\}. \tag{3.6}$$

We now choose m different $P \in \mathcal{P}$ by sampling independently and uniformly at random. For each $i \in m$, we measure P_i on several copies of ρ and average over the outcomes to obtain an estimate for $\mathrm{Tr}(P_i \rho)$. More formally, we can define a sampling operator (a linear map) $\mathcal{A} : \mathcal{H} \mapsto \mathbb{R}^m$ with

$$(\mathcal{A}(\rho))_i = \sqrt{\frac{n}{m}} \mathrm{Tr}(P_i \rho),$$

where we already normalised. The measurement outcomes are then given by

$$y = \mathcal{A}(\rho) + z,$$

where z represents statistical noise from the finite amount of measurements. The reconstruction of $\tilde{\rho}$, the low-rank estimate of ρ is then performed as follows. We search for a matrix $x \in \mathcal{H}$ that fits the measurement results y while at the same time minimising the trace norm $|x|_{\mathrm{tr}}$.

$$\tilde{\rho} = \arg\min_x \frac{1}{2} |\mathcal{A}(x) - y|_2^2 + \mu |x|_{\mathrm{tr}} \tag{3.7}$$

μ is a parameter that has to be fitted to the amount of noise from the measurements. The minimisation of the trace norm of x encourages low-rank outcomes in an efficient way, as minimising the rank itself is a hard problem. This low-rank estimate is then checked using direct fidelity estimation.

If the system to be estimated is not discrete but continuous, the measurement setting has to be generalised. Instead of sampling the measurement operators from Pauli measurements, they are taken from a non-orthogonal basis, a tight frame, which is an overcomplete, non-orthogonal generalisation of operator bases. The orthogonal projectors $\mathcal{P}$ are then measured from this tight frame, while satisfying certain incoherence properties to ensure uniqueness of the recovered solution. Fourier type measurements often fulfil such conditions. One example is homodyne measurement, where the expectation values of the displacement operators form a tight frame for certain constraints on the phase angle. See [72] for a more thorough overview on continuous variable compressed sensing.

Compressed sensing, apart from image processing and quantum state tomography, can also be used in the reconstruction of quantum processes. In process tomography, the process ϵ is translated into a state using the Choi-Jamiolkowski isomorphism. The tomography is then performed on the state ρ_ϵ and the process is reconstructed from the reconstruction of the quantum state.

$$\rho_\epsilon = (\epsilon \otimes \mathbb{1}) \left(|\psi_0\rangle\langle\psi_0| \right), \quad |\psi_0\rangle = \frac{1}{\sqrt{d}} \sum_{j=1} |j\rangle \otimes |j\rangle \tag{3.8}$$

The rank of ρ_ϵ is equal to the Kraus rank of ϵ. Variants of the compressed sensing techniques introduced in this chapter will be applied in a following chapter in the deconvolution of non-linear signals from measurements of atoms in optical traps.

4. Quantum Information Theory

To complete our introduction, we will shortly introduce classical and quantum information theory, as they moulded the mindset of a sustainable treatment of the resources of information or entanglement. In measurements of quantum systems, the conducted measurement or the applied transformation becomes a part of the whole system that is to be identified. In order to understand the impact of these operations, tools developed in information theory are absolutely necessary. We will introduce the concepts of entropy, the classical formulation and its quantum counterpart, as well as the relative entropy which is a tool that allows comparisons of probability distributions.

4.1. Classical Information Theory

> Information theory has, in the last few years, become something of a scientific bandwagon [...] Although this wave of popularity is certainly pleasant and exciting for those of us working in the field, it carries at the same time an element of danger. While we feel that information theory is indeed a valuable tool [...] it is certainly no panacea for the communication engineer or, a fortiori, for anyone else. Seldom do more than a few of nature's secrets give way at one time. It will be all to easy for our somewhat artificial prosperity to collapse overnight when it is realised that the use of a few exciting words like information, entropy, redundancy, do not solve all our problems. (Shannon, 1956)

Indeed, the wave of popularity of information theory did not come to an end. One of the reasons might be, that the concept of information is highly intertwined with quantum mechanics itself. The discussion of information theory shifts the point of view from the outcomes of a measurement to the process of the measurement itself. There are many great reviews, books and even philosophical texts about this topic, see for example [70, 91, 94], and we will only concentrate on a few topics.

In classical information theory, the information content of a statement highly depends on the information that is already available about the system. If a desk has only two possible colours—white and black—the statement "The desk is not white" reveals the

full knowledge about the colour of the desk, while in a world where the set of possible colours contains eight or 16 colours, the gained knowledge is way smaller. In that sense, a useful measure for information should quantify how much information is gained when the information is revealed. If a random variable X has x possible outcomes with probabilities $p(x)$, the Shannon entropy

$$S(X) \equiv S(p(x)) = -\sum_x p(x) \log_2 p(x) \geq 0 \tag{4.1}$$

provides such a measure. It gives the average of the information content of a possible statement. The amount of information is hereby modelled by the negative of the logarithm of the probability distribution. As this measure is lower for a low number of possible outcomes with relatively high probabilities, as for the desk which is either black or white, the Shannon entropy is often described as the amount of uncertainty or unpredictability in a certain statement. For the desk, that is either black or white, depending on the probabilities of either outcome, the Shannon entropy gives a maximum of one. If only one eighth of the desks are white, it even drops down to 0.54, as the uncertainty in a possible statement goes down—the statement will probably be, that the desk is black. Likewise, if we allow for more possible outcomes (or colours), the unpredictability rises, and so does the Shannon entropy. For only one single possible outcome the Shannon entropy is equal to zero, reflecting that there is no uncertainty in the statement to come. Similar to the entropy in the case of a single random variable, a measure can be defined for two random variables X and Y. This conditional entropy is defined using the conditional probability $p(y|x)$. This probability reflects the probability of event Y, given that event X already occurred. We say that, introducing the simultaneous or joint probability $p(x, y) = p(x)p(y|x)$

$$S(Y|X) = -\sum_x p(x) \sum_y p(y|x) \log_2(p(y|x)) = -\sum_x \sum_y p(x, y) \log_2 p(y|x).$$

Note that if the two probability distributions are independent of each other, the joint probability of x and y is a simple product of the single probabilities, $p(x, y) = p(x)p(y)$, as $p(y|x) = p(y)$, whenever $p(x) \neq 0$. In this case, the conditional entropy is equal to the entropy of Y alone, $S(Y|X) = \sum_x p(x)S(Y) = S(Y)$. However, if the distributions are not independent, this equality doesn't hold anymore, and $S(Y|X) < S(Y)$. The uncertainty of outcome Y is less, as part of the information is already contained in the occurrence of outcome X. We can define a mutual information content of X and Y as $I(X : Y) = -S(Y|X) + S(Y)$

$$I(Y : X) = -\sum_x \sum_y p(x, y) \log_2 \frac{p(y|x)}{p(x)p(y)} = I(X : Y).$$

This symmetric measure quantifies how independent X and Y are from each other and it is related to the conditional as well as the joint entropy, as $I(X:Y) = -S(Y|X) + S(Y) = -S(X,Y) + S(X) + S(Y)$. We see that for independent distributions, the mutual information content is zero, as $S(Y|X) = S(Y)$ in these cases. This on the other hand means that $S(X,Y) = S(X) + S(Y)$, so the joint entropy is additive in the case of independent distributions. If two independent systems are combined, the entropy of the total system is the sum of the single entropies. Whenever the systems are dependent, the joint entropy is always less than the sum of the single entropies. This behaviour is called subadditivity. This subadditivity is a direct consequence of using the logarithm as a measure for the amount of information in the system. The subadditivity entails, that performing data processing, which contains collection and manipulation, can never increase the amount of information in a system. This is known as the data processing inequality, which holds for all measures of information. It can also be understood as the impossibility of introducing new information through post-processing. Data processing can be seen as a stochastic process and the data processing inequality states that using data processing to arrive at a new variable Z cannot carry more information about an original distribution X then the result of a first data processing Y. Both mutual informations are again bounded from above by the the entropy, or information content of X: $S(X) \geq I(X:Y) \geq I(X:Z)$. While this classical inequalities are plausible and can be understood intuitively, the quantum equivalents are more involved.

4.2. Quantum Information Theory

The objects of interest in quantum information theory are quantum states of physical objects. As introduced in the last chapters, these quantum states can be represented by density operators ρ that are positive semi-definite operators, $\rho \geq 0$, over a Hilbert space (a complex vector space with additional properties), and satisfy $\mathrm{Tr}(\rho) = 1$. For finite-dimensional quantum systems, these operators $\rho \in \mathcal{C}^{N \times N}$ are Hermitian, positive semi-definite matrices over an N-dimensional vector space, normalised to unit trace: They generalise probability distributions over a finite alphabet. The main diagonal elements of ρ will also always constitute a probability distribution. Any quantum state can be described by such an object.

Von Neumann introduced the first analogue to the classical Shannon entropy defined for any density operator ρ with $\rho = \rho^\dagger \geq 0$ with trace 1:

$$S(\rho) = -\operatorname{Tr} \rho \log_2 \rho.$$

If ρ is decomposed into projectors $|\phi_i\rangle \langle\phi_i|$ with probabilities p_i, it can be shown that $S(p_i) \geq S(\rho)$. Equality holds if the pure states ϕ_i are pairwise orthogonal, for

example eigenstates of ρ. The reason for this bound lies in the distinguishability of quantum states. While classical outcomes can always be compared and distinguished, non-orthogonal quantum states cannot be distinguished with certainty. This reduces the information content. The straightforward way of calculating the von Neumann entropy is by using the eigenspectrum of ρ. With eigenvalues λ_i, we get $S(\rho) = -\sum_i \lambda_i \log_2 \lambda_i$.

The quantum entropy $S(\rho)$ is zero if and only if ρ represents a pure state, having only a single non-zero eigenvalue $\lambda = 1$. On the other hand, $S(\rho)$ is maximal for maximally mixed states. As the trace is invariant under the change of basis, the von Neumann entropy shares this property and $S(\rho) = S(U\rho U^\dagger)$ for any unitary transformation U. Similar to its classical equivalent, the quantum entropy can only grow when two or more quantum systems are combined, a property which is called concave. Another property it shares with its classical counterpart is the subadditivity, which bounds this growth as $S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B)$, where ρ_A and ρ_B are the reduced density matrices of a general state ρ_{AB}. The full proofs of these properties is covered in many books, for example Nielsen and Chuang [70] and we will not repeat them here. It is not surprising that both concepts share a lot of properties, as the definitions resemble strongly. There is however one very useful difference between the classical and the quantum entropy. While in classical systems, the entropy of a subsystem is never greater than the entropy of the whole system, there exist quantum states where this is the case. One example is the pure state $\rho_{AB} = |\phi\rangle\langle\phi|$ with $|\phi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ with $S(\rho_{AB}) = 0$. The reduced states however both have an entropy of 1. This is why quantum entropy can be used to detect entanglement between two subsystems. Whenever $S(\rho_{AB}) < S(\rho_A)$, the subsystems A and B are entangled. This property might be surprising, as the reduction to subsystems and the combination of different systems seem to be very similar. The use of quantum entropy reveals this non-classical behaviour in quantum states. As a consequence, the subadditivity can be extended on the left hand side resulting in a triangle inequality: $\|S(\rho_A) - S(\rho_B)\| \leq S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B)$. The left hand side can be interpreted as the possibility of the entropies of the subsystems cancelling each other, if the state is entangled, as shown in the example of the Bell state stated above. If the two subsystems have different amounts of entropy, the entropies can however only partially cancel each other, leaving some entropy in the total system. The role of entanglement as a valuable resource in quantum communication will be studied in chapter seven.

4.3. The Relative Entropy

A useful measure of the closeness or similarity of two probability distributions $p(x)$ and $q(x)$ is given by the relative entropy, or Kullback–Leibler divergence. As we will use it later in the attempt to find the closeness of a measured state to the set of Gaussian

states, it deserves its own section. Its classical formulation is defined as

$$D(p(x)|q(x)) = \sum_x p(x) \log\left(\frac{p(x)}{q(x)}\right) \tag{4.2}$$

and it qualifies as a measure of distance, as it is non-negative and zero if and only if the two probability distributions are equivalent for all x. Notice, that the relative entropy is non-symmetric and measures how much information is lost if $q(x)$ is used to approximate $p(x)$. In that sense, it can qualify whether $q(x)$ is a good model for $p(x)$. The quantum version of the relative entropy, the probability distributions get replaced by states:

$$D(\rho|\sigma) = -\operatorname{Tr}\rho \ln\left(\frac{\rho}{\sigma}\right) = -\operatorname{Tr}\rho\ln\rho + \operatorname{Tr}\rho\ln\sigma. \tag{4.3}$$

A very useful property of the relative entropy that follows from its convexity is the relative entropy of a state ρ with respect to a reference state σ which is a tensor product of lower-dimensional states σ_i:

$$D(\rho|\bigotimes_i \sigma_i) \geq \sum_i D(\rho_i|\sigma_i). \tag{4.4}$$

So if σ is a separable state and can be written as a tensor product of σ_i in subspaces $\mathcal{H}_i$, the relative entropy of the whole system is an upper bound to the sum of relative entropies of the projections on the subspaces $\mathcal{H}_i$. This can be seen as a potential loss of information through the projection of ρ, which can only lower the bound on the distinguishability of ρ and σ.

A more extensive description of the property of (relative) entropies can be found in Nielsen and Chuang [70]. We will now proceed to use these quantum information tools together with the concepts and methods of describing and reconstructing quantum states in several situations.

Part III.

System Identification

5. Deconvolution of Fluorescence Peaks Using Compressed Sensing

In the following chapter we describe our approach to solve the non-linear problem of finding amplitudes and positions of atoms in an optical lattice using a compressed sensing method in a single reconstruction step. This work was done in collaboration with Carlos Ríofrio and Jens Eisert. We closely worked together with Andrea Alberti and Dieter Meschede from the institute of applied physics (IAP) in Bonn, who provided experimental data from fluorescence imaging of atoms in optical lattices [3, 61].
After discussing the general approach to non-linear systems using compressed sensing, we describe the experimental setup used in fluorescence imaging and identify the measurement map in the context of compressed sensing. We use a trace norm minimisation to ensure sparsity of the source function describing the lattice occupation derived from compressed sensing techniques. By investigating the restricted isometry property of the experimental measurement map, we find the limits for the application of compressed sensing in this and similar measurement setups. We derive fundamental error bars on any reconstruction based on the information content in the measurement data and compare the compressed sensing reconstruction to the two-step reconstruction method used by the group in Bonn. Using a simulation of the measurement process, we also investigate possibilities to reduce the error bar in the reconstruction by expanding or changing the measurement setup.

5.1. Non-linear Systems

In a non-linear system, the answer to an exterior excitation is not proportional to its magnitude. Neither additivity nor homogeneity hold, which entails that the superposition principle does not apply. Equations cannot be solved by finding linear combinations of known solutions, which makes these systems difficult to solve. In fact, most systems that occur in nature are non-linear, and some of them are not even linearisable. This is one of the reasons why weather forecasting is hard and not always right.

5.1.1. Linearisation

Many other problems however, can be transformed to a linear form by for example separation of variables or linearisation. Sometimes these transformations only hold in a small domain, like in the case of the pendulum under influence of gravity, which can be brought to the form of a simple harmonic oscillator for small angles. In linearisation, the output of $y = f(x)$ at any $x = a$ is approximated based on the value and the slope of $f(x)$ at $x = b$, given f is continuous on $[a, b]$ and a is close to b. We now want to focus on this concept of linearisation in the context of compressed sensing.

5.1.2. Compressed Sensing

As introduced in section 3.2, Compressed Sensing is a technique to reconstruct low-rank states with a small amount of measurement settings. Classical compressed sensing is used to find the unique sparse solution to an underdetermined system. In the quantum version of compressed sensing, only $\mathcal{O}(Nr \log N)$ measurement settings are needed to perform a tomography on an N-dimensional rank r state ρ. To get a constant error in the reconstruction of the state, only $\mathcal{O}\left(r^2 N^2 \log N\right)$ copies of the state are needed. The reconstruction is done via two-norm minimisation and can be performed efficiently through convex optimisation. A big upside of the reconstruction through compressed sensing is the possibility of reliable error bounds, not only in the reconstruction guarantee but also through direct fidelity estimation. Compressed sensing is best applicable, whenever the signal of interest is an approximately sparse, finite N-dimensional real or complex vector $\vec{x}$. Sampling is done using a linear mapping ϕ to an M-dimensional observation space resulting in a distribution $\vec{y}$

$$\vec{y} = \phi\vec{x} + \vec{e}, \tag{5.1}$$

where $\vec{e}$ represents measurement noise with $\|e\|_2 = \varepsilon$. Not every mapping ϕ allows for an efficient and reliable reconstruction of $\vec{x}$ from $\vec{y}$. ϕ has to fulfil the so-called *restricted isometry property* (RIP), which can be understood as bounding the effect small deviations in the source distribution $\vec{x}$ have on the resulting distribution $\vec{y}$. Another possible interpretation is that if the map ϕ is close to an Euclidean map, it will almost preserve distances, which is equivalent to ϕ being Lipschitz. The Lipschitz property is very closely related to fulfilling the RIP, as can be seen in the following definition: A distribution ϕ fulfils the RIP if

$$(1 - \delta_s)\|x\|_2^2 \leq \|\phi x\|_2^2 \leq (1 + \delta_s)\|x\|_2^2 \tag{5.2}$$

for s-sparse x and $\delta_s < 1$. δ_s is defined as the smallest constant for which RIP holds for all x with fixed rank s. There are several algorithms which can be chosen to solve

the reconstruction problem $\min_x \|y - \phi x\|_2^2$ subject to x being of a fixed sparsity. All of them use that the minimisation problem is isomorphic to minimising the rank of x subject to $\|y - \phi x\| \leq \varepsilon$. However, they differ in the way the constraint is introduced in the reconstruction problem.

One method, which is basically a gradient search with a constraint enforcing sparsity, is called *Iterative Hard Thresholding* (IHT) [14]. It introduces a stepsize or learning rate $\mu > 0$ and is an iterative procedure. The IHT is then given by

$$x_0 = 0, \quad x^{[n+1]} = H_s\left(x^{[n]} + \mu\phi^T(y - \phi x^{[n]})\right). \tag{5.3}$$

$H_s(a)$ is a non-linear operator that sets all but the largest s elements to zero. For $\|\phi\|_2 < 1$, this expression converges to a local minimum of $\min_x \|y - \phi x\|_2^2$ subject to $\|x\|_0 \leq s$ (where $\|x\|_0$ is the number of non-zero elements in x). This method is relatively robust against noise, it degrades only linearly with noise, and gives reliable error bars. The number of needed observations is optimal and grows linearly with s. The only condition is that the map ϕ and its transpose ϕ^T are linear and known. In the beginning of this chapter we discussed, that there are many cases in natural system settings, where this assumption is wrong. We will now focus on how the reconstruction has to be adapted to the case of a non-linear map.

5.1.3. Non-linear Compressed Sensing

If a system is not too non-linear, there is always hope that the approaches used for linear systems can be adapted to the new situation. In non-linear compressed sensing, this adaption is done by introducing a new step in the reconstruction procedure. The non-linear map ϕ is first linearised and afterwards the linear compressed sensing reconstruction method is performed. Of course, this linearisation introduces a new source of error, which changes the error bounds and the conditions on reconstructability. Two main questions immediately arise. The first is, whether we can still find a convex objective function to minimise the problem $\min_x \|y - \phi(x)\|_2^2$ subject to x being sparse, where $\phi(x)$ is now a non-linear map $\phi : \mathcal{H} \mapsto \mathcal{B}$. It is not clear how to find such a function in general, but it turns out one can use an adapted version of IHT. The second question is how to deal with the notion of sparsity in a general space $\mathcal{H}$. For this we can introduce a known set $\mathcal{A} \subset \mathcal{H}$ and define the projection $P_\mathcal{A} : \mathcal{H} \mapsto A$. We then introduce $x_\mathcal{A} = P_\mathcal{A} x$ as follows

$$\|x - x_\mathcal{A}\|^2 \leq \inf_{\tilde{x} \in \mathcal{A}} \|x - \tilde{x}\|^2 + \varepsilon. \tag{5.4}$$

The positive ε entails for the situation that x_{opt} itself might not be in $\mathcal{A}$. By replacing H_s in (5.3) by $P_\mathcal{A}$, we find a new version of IHT by replacing $\phi(x)$ with its linearisation in x^*

$$x_0 = 0, \quad x^{[n+1]} = P_\mathcal{A}\left(x^{[n]} + \mu\phi_{x^n}^T(y - \phi_{x^n} x^{[n]})\right). \tag{5.5}$$

This algorithm converges, whenever ϕ_{x^*} satisfies RIP

$$\forall_{x_1,x_2,x^*\in\mathcal{A}}\alpha\|x_1-x_2\|_2^2 \leq \|\phi_{x^*}(x_1-x_2)\|_2^2 \leq \beta\|x_1-x_2\|_2^2 \tag{5.6}$$

while $\phi(x)$ and $\phi_{x^*}x$ satisfy a linearisation criterion.

$$\|\phi(x_1)-\phi(x_2)-\phi_{x^*}(x_1-x_2)\|_2^2 \leq C\|(x_1-x_2)\|_2^2 \tag{5.7}$$

Whenever the constants satisfy $\beta < \frac{1}{\mu} < 1.5\alpha - 4C$, IHT converges to a x^* which is ε-close to x. The error bound can be directly deduced by bounding $\left\|(x-x^{[n+1]})\right\|_2$ using (5.6) and (5.7) with the appropriate choice of the constants.

One example for a system that is very close to linear is a measurement setup where a signal undergoes a linear measurement system Φ which satisfies the RIP with some constants α and β and is then captured by detectors with a non-linear response f, such that the total measurement map is given by $\Phi f(x)$ with a non-linear $f(x) = x + h(x)$. The linearisation is then given by $\phi_{x^*} = \Phi + \Phi\mathbb{1}h'(x^*)$, the linear part of the measurement map corrected by the the derivative of h. It turns out that it is enough for $\|h'(x)\|$ to be bounded from above by one for RIP to hold, because this entails the linearisability condition (5.7) for Φ and ϕ_{x^*}.

We will now turn on an explicit example of a measurement setup that is non-linear and perform the linearised version of compressed sensing to recover atom positions in an optical lattice with reliable error bars.

5.2. Fluorescence Imaging

In Fluorescence Imaging, atoms in a deep lattice confinement are illuminated using a coherent light source. The fluorescence light emitted by the atoms is collected using a microscope and CCD detectors. From the resulting image, the positions and fluorescence amplitudes, i.e. reflectivity, of the atoms have to be reconstructed. In this specific experiment, the optical 1D lattice contains caesium atoms. The experiment has been conducted at the group of Prof. Meschede in Bonn [3, 61].

The original source distribution S of atoms at position x in the 1D optical lattice is given by

$$S(x) = a_0 + \sum_{k=1}^{N} a_k\delta(x-\xi_k), \tag{5.8}$$

where ξ_k is one of the N possible positions in the optical lattice, a_k is the fluorescence amplitude of the kth atom and a_0 is stray light background introduced by the measurement. Tunneling is suppressed and we want to recover the positions ξ_k and the

fluorescence amplitudes a_k in the lattice. At the same time we want to get a reliable error bar on the reconstruction to be able to make predictions on which resolution would be needed in the measurement to be able to distinguish individual lattice sites. As enhancing the resolution in the measurement setup is very costly, an investigation of the influence of other parts of the measurement setup on the error bars is very promising. In the experiment, fluorescence imaging is used to measure a discrete distribution $I(s_i)$ which is related to the continuous source distribution $S(x)$ through convolution with the so-called *line spread function* (LSF) $\mathcal{L}$

$$I(s_i) = \int_{-\infty}^{\infty} \mathcal{L}(s_i - u)S(u)\mathrm{d}u + e(s_i) \tag{5.9}$$

$$= \sum_k a_k \mathcal{L}(s_i - \xi_k) + \mathcal{L} \cdot a_0 + e(s_i) \tag{5.10}$$

$$= \phi(a_k, \xi_k) + \tilde{e}(s_i), \tag{5.11}$$

with a non-linear measurement ϕ. If there are only few atoms in the optical lattice, so if $N \ll R$ with R being the image resolution, we can define a sparse signal vector x with $x = (a_1, \ldots, a_N)$, where a_i is non-zero if and only if there is an atom present at position i in the lattice.

This distribution I described in (5.11) is obtained from the source distribution S introduced in (5.8) through several imaging steps. First, the source distribution is transformed by optical diffraction in the imaging microscope. The originally sharp points are transformed into blurred spatial distributions. Mathematically, this can be described by the convolution of the source distribution with the line spread function. After this imaging procedure, the distribution is sampled by a CCD detector, resulting in a discrete version of the former continuous distribution. In general, $I(x_i) \neq I[x_i]$, as the discrete version $I[x_i]$ is the projection of sampling spacings in the object plane to one single pixel through a rectangular function. In this process, information can be lost, as the pixels and the lattice spacings do not have the same size and there are offsets between the two grids. In the last step of the imaging process, each pixel or bin is impaired by noise sources like photon shot noise. This noise varies from pixel to pixel and is represented by $e(s_i)$ in (5.11). The pictures we will discuss are collected over one second of exposure time. The image is then integrated over the time axis to result in a higher signal strength. Several post processing steps have to be performed to be able to perform the reconstruction on the image. Noise background has to be reduced, the number of atoms per image is estimated and the pictures are cut into smaller sections, so-called *regions of interest* (ROI), that contain the least amount of area without atoms, which would only contribute to the error in the reconstruction. Before we discuss how the pictures are processed, we will discuss whether reconstruction using compressed sensing

techniques is possible in this setup.

There are several ways how to approach the problem of reconstructing a_k and ξ_k from (5.11). In the current setup used in [3, 61], the positions are guessed from the Fourier transform (which turns the convolution into a simple product) and used as input values for an iterative least squares procedure that fits the best positions and amplitudes to the image distribution. The downside of this approach is, that there is no reliable information on error bars from this reconstruction. We will follow the approach of non-linear compressed sensing introduced in the last section. In order to do so, we will have a closer look on the measurement map ϕ and the line spread function $\mathcal{L}$.

To model the measurement map, we assume that the source distribution $S(i)$ gets mapped to an image I. The map consists of a convolution with the so called line spread function $\mathcal{L}$, caused by the optical system, followed by a discretisation caused by the finite resolution of the CCD detecting the reflected photons. The number of pixels will be denoted as N. This results in the discrete distribution I introduced in (5.11). $\tilde{e}(s)$ is an error term containing both the stray light background and Gaussian noise at every discrete position s. To be able to reconstruct unknown positions and amplitudes from the measured signal, we have to reformulate the problem.

5.3. Preparing the Compressed Reconstruction

We reformulate the problem by introducing a matrix L which is a matrix of dimension $N \times M$, where M is the resolution up to which the linespread function is known, which is typically much higher than the image resolution. The matrix L contains shifted versions of the linespread function in every row. For $M = N$, L is a Toeplitz matrix. We will address the construction of L in a following section. Using this matrix, the problem can be stated as $I_k = \sum_i L_{ki} x_i + e_k$ with $k \in 1, \cdots, N, i \in 1, \cdots, M$. The vector $\vec{x} \in \mathbb{R}^M$ contains the fluorescence amplitudes a_i in each position. Since only few atoms are present in the lattice, most of these amplitudes will be zero. Therefore, $\vec{x}$ is a sparse vector. This will allow us to use compressed sensing techniques to reconstruct $\vec{x}$ from I. We will now discuss whether this linearisation satisfies the RIP condition needed in order to be able to apply non-linear compressed sensing techniques.

5.3.1. Restricted Isometry Property

To be able to use compressed sensing techniques, we have to find the bounds on the restricted isometry property for the transformation matrix L. We do this by proving the bounds for which a Toeplitz sensing matrix always fulfils the property. As this is true for all Toeplitz matrices, we will choose a different name for the proof and call the

matrix A. Let this $n \times n$ matrix be A such that $A_{ij} = a_{i-j}$

$$A = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \cdots & a_{1-n} \\ a_1 & a_0 & a_{-1} & \cdots & a_{2-n} \\ a_2 & a_1 & a_0 & \cdots & a_{3-n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n-1} & a_{n-2} & a_{n-3} & \cdots & a_0 \end{bmatrix}. \tag{5.12}$$

In our particular case, given the fact that A is the matrix representation of the line spread function, we have the following properties

$$a_0 > a_{\pm 1} > a_{\pm 2} > \cdots > a_{\pm(1-n)} \geq 0. \tag{5.13}$$

For the restricted isometry property to hold, we need to find the minimum and maximum of $\|Ax\|_2^2$, where x is a sparse binary vector, i.e. its components are either 0 or 1.

Note that

$$\|Ax\|_2^2 = x^{\mathrm{T}} A^{\mathrm{T}} A x = \sum_{j,k=1}^{n} \left(\sum_{i=1}^{n} A_{ij} A_{ik} \right) x_j x_k \tag{5.14}$$

where $^{\mathrm{T}}$ denotes matrix transposition, and the term in the parenthesis is

$$B_{kj} = \sum_{i=1}^{n} A_{ij} A_{ik} = \left(A^{\mathrm{T}} A \right)_{kj}. \tag{5.15}$$

Since x is binary, we can write

$$\|Ax\|_2^2 = \sum_{j,k} B_{kj} \tag{5.16}$$

where $j, k \in \mathcal{S}$ with $\mathcal{S} = \{\alpha_1, \alpha_2, \ldots, \alpha_s\}$ is the set of indices α_j for which $x_{\alpha_j} = 1$. The number of elements of $\mathcal{S}$ is the sparsity s.

In order to minimise or maximise this last equation, we need to know the structure of matrix B defined in Eq. (5.15). Since A is Toeplitz, we can write B as

$$B_{kj} = \sum_{i=1}^{n} a_{i-j} a_{i-k}. \tag{5.17}$$

In our problem, the elements of A decay very fast to zero. This means that there exist indices β and γ for which

$$a_0 > a_{-1} > a_{-2} > \cdots > a_{1-\beta} = a_{-\beta} = \cdots = a_{1-n} = 0 \tag{5.18}$$

and

$$a_0 > a_1 > a_2 > \cdots > a_{\gamma-1} = a_\gamma = \cdots = a_{n-1} = 0. \tag{5.19}$$

Noting this, we can write the matrix A explicitly

$$A = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \cdots & a_{2-\beta} & 0 & 0 & 0 & \cdots & 0 \\ a_1 & a_0 & a_{-1} & \cdots & a_{3-\beta} & a_{2-\beta} & 0 & 0 & \cdots & 0 \\ a_2 & a_1 & a_0 & \cdots & a_{4-\beta} & a_{3-\beta} & a_{2-\beta} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{\gamma-2} & a_{\gamma-3} & a_{\gamma-4} & \cdots & a_{\gamma-\beta} & a_{\gamma-\beta-1} & a_{\gamma-\beta-2} & a_{\gamma-\beta-3} & \cdots & 0 \\ 0 & a_{\gamma-2} & a_{\gamma-3} & \cdots & a_{\gamma-\beta+1} & a_{\gamma-\beta} & a_{\gamma-\beta-1} & a_{\gamma-\beta-2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & & & & & & \\ 0 & 0 & 0 & \cdots & & & & & & \\ 0 & 0 & 0 & \cdots & 0 & a_{\gamma-2} & a_{\gamma-3} & \cdots & a_1 & a_0 \end{bmatrix}. \tag{5.20}$$

The matrix B is Toeplitz in a middle region, whenever every row of A is decaying to zero in a "fast" way, such that for every $b > N$: $A_{i,i+b} = A_{i,i-b} = 0$. The number of non-zero entries in a row of A is thus $2N + 1 > \dim(A) = M$. Consider now the following

$$\begin{aligned} B_{j,k} &= \sum_{\xi} A^T_{j,\xi} A_{\xi,k} \\ &= \sum_{\xi} A_{\xi,j} A_{\xi,k} \\ &= \sum_{\xi} A_{\xi+n,j+n} A_{\xi+n,k+n} \\ &= \sum_{\xi} A^T_{j+n,\xi+n} A_{\xi+n,k+n} \\ &= B_{j+n,k+n}, \end{aligned} \tag{5.21}$$

where the last equality sign needs the following discussion to hold. Since the summation always runs from $\xi = 0, \ldots, M$ we need to ensure, that all the relevant terms are still included in the summation. As stated above, we know that A decays to zero after some number of steps N. We thus know, that the only indices where $A_{j,\xi}$ is non-zero is for $j - N < \xi < j + N$. The same holds of course for k. We can derive the same constraint for $\xi + n$, where we also use that $0 < \xi < M : 0 < n < j - N < \xi + n < j + N < M + n$. We can also bound the indices from above by M, since it is the dimension of A. Doing that, we get two constraints, one for the indices j and k and one for n

$$n < M - 2N \tag{5.22}$$

$$N < j < M - N \tag{5.23}$$

$$N < k < M - N. \tag{5.24}$$

If j and k fulfil these conditions, the last equality sign above holds and B is Toeplitz in this domain. The dimension of the Toeplitz part is reduced by $2N$ leaving us with $M - 2N$.

We want to argue that the relevant part of the matrix B is the Toeplitz region. We observe that, for $M \to \infty$, B approaches a fully Toeplitz matrix. The deviation from the Toeplitz form is only a boundary effect.

Bounds in RIP

In the following, we concentrate on the Toeplitz part of B and look for an upper and lower bound in the Toeplitz domain. For simplicity, we stick to the name B. Our problem still is to bound

$$\|Ax\|_2^2 = x^{\mathrm{T}} A^{\mathrm{T}} Ax = x^{\mathrm{T}} Bx = \sum_{j,k} B_{j,k} \tag{5.25}$$

from below and above. In the last equality sign we used that x has a given sparsity s and only binary entries. The sum runs over all indices for which $x_j = x_k = 0$. The complexity of the problem can be simplified further by realising, that we can introduce the distance between the indices $d = k - j$. Then $\sum_{j,k} B_{j,k} = \sum_j \sum_d B_{j,j+d}$, where d runs over all the distances between non-zero entries in x. Since B is Toeplitz, we see that this problem is independent of j and we only have to solve it for one special j. We also know one additional property of B, namely monotonicity: $B_{i,i} > B_{i,i+d}$ and $B_{i,i+d} > B_{i,i+e}$ for $d > e$.

Theorem 1. *The smallest possible distance (meaning next neighbours around the diagonal) yields the upper bound for the RIP.*

Proof. Assume the RIP would be maximised by a distance in the indices that was bigger then one, so we would not include $B_{i,i+e}$ but $B_{i,i+e+1}$ in the summation. Then we can find a higher value by replacing $B_{i,i+e+1}$ by the not yet included $B_{i,i+e}$.

By summing the s highest entries in a row of B we thus get the upper bound for the RIP. □

Theorem 2. *Maximising the distance yields the lower bound for the RIP.*

Proof. This is a bit more subtle. For small sparsities, we can make use of the property mentioned above, where we see that the entries of a row in B decay to zero after some distance N from the diagonal. There it is easy to see that we can minimise $\sum_d B_{i,i+d}$ by taking only $d > N$ apart from $d = 0$. However, if the sparsity is too big, we will have contributions from the next entries as well, since our matrix is not infinitely long. Making

one distance bigger will thus reduce another. We can solve that problem by looking at the gradient of a row of B. If the decay in the entries is bigger closer to the diagonal, the biggest distance optimises the problem: Assume we would have an optimal value for distances a, b, where $a < b$. Then we would get a better bound using $a+1$ and $b-1$, since $B_{i,i+a} - B_{i,i+a+1} > B_{i,i+b-1} - B_{i,i+b}$ and thus $B_{i,i+a} + B_{i,i+b} > B_{i,i+a+1} + B_{i,i+b-1}$.

By maximising the distance and taking equidistant non-zero entries in x we get the lower bound for the RIP, as long as the LSF decays faster close to the diagonal. As we will see in the next part, this is the case in the experiment mentioned above, where the LSF is close to a Gaussian. We can thus indeed use the compressed sensing techniques introduced in the beginning of this chapter to reconstruct the measurement data. □

5.3.2. Simulation of the Measurement

In order to test our reconstruction techniques and to predict which changes in the measurement setup are promising for achieving single site resolution, we want to be able to simulate own measurement data similar to the one obtained from real experiments. To achieve this goal, we modelled the different imaging steps described in the experimental section and used the parameters from the actual measurements. The resulting images resemble the images from the experiment very closely, as can be seen from Fig. 5.1. The linespread function used to simulate the pictures is the one present in the experiment. Note, that the experimental picture has a much higher amplitude. This is because the experimental picture is taken and later integrated over a certain period of time, while the simulated picture shows a single shot variant. In a reconstruction, this will result in higher amplitudes being reconstructed.

The full code of this simulation is available in appendix B.2. We will guide through it shortly and show different examples for simulated pictures in Figures 5.2 and 5.3

With these artificial images, we are able to check the reliability of our reconstructions, which we also performed with actual pictures. Additionally, we compared our reconstructions to the reconstructions from [61] and [3].

5.3.3. Linearisation and Error Bars

It is already possible to use this description of the measurement process to derive error bars on any reconstruction method taking into account the amount of noise that is present in the picture. The noise accumulates from photon shot noise, errors in the detectors as well as the discretisation. Any reconstruction method that does not make use of additional information can only distinguish two possible outcomes where the resulting picture differs more than the average noise in this area. We are mostly interested in the error bar with respect to the position in superresolution. To translate the noise present

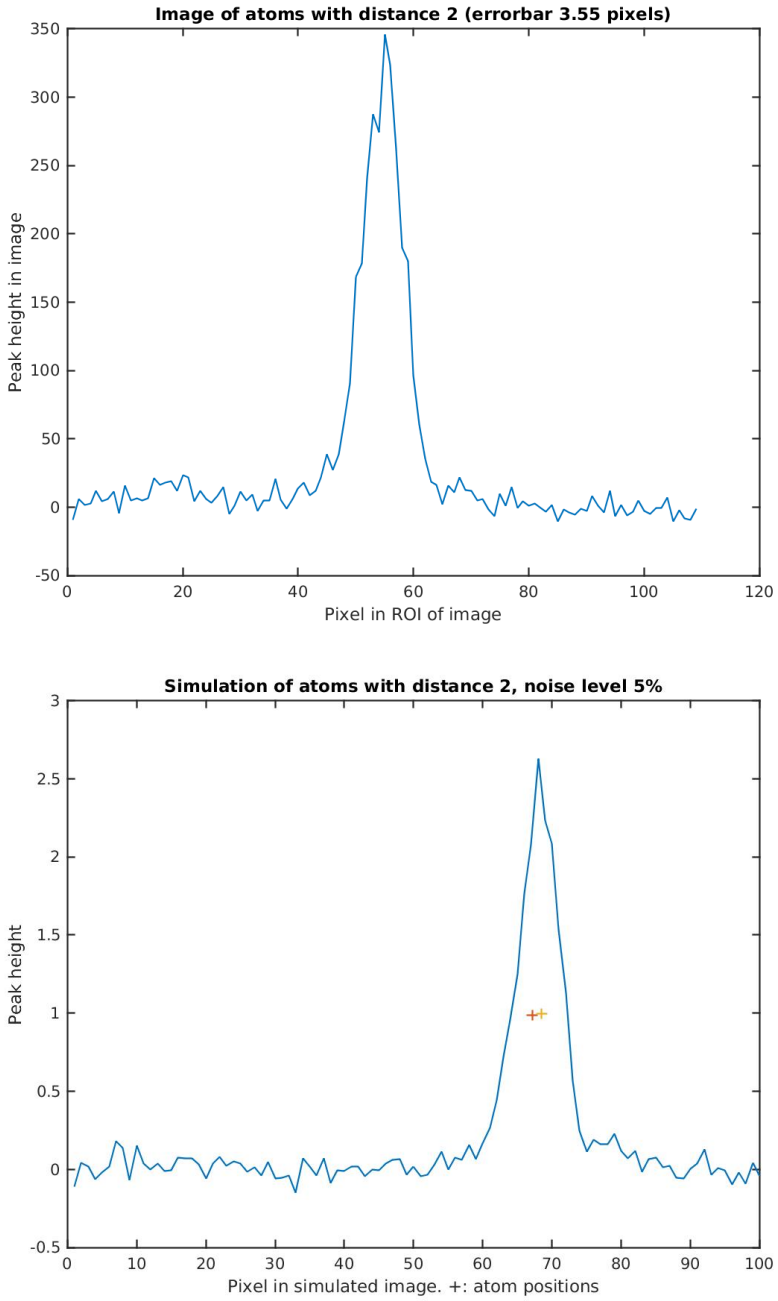

Figure 5.1.: Both images show atoms of distance 2. The top image is an image taken in the lab (image 8 of atoms with distance 2), while the bottom image shows a simulated image with a noise ratio of 5 %, which is the average amount of noise in the experimental pictures.

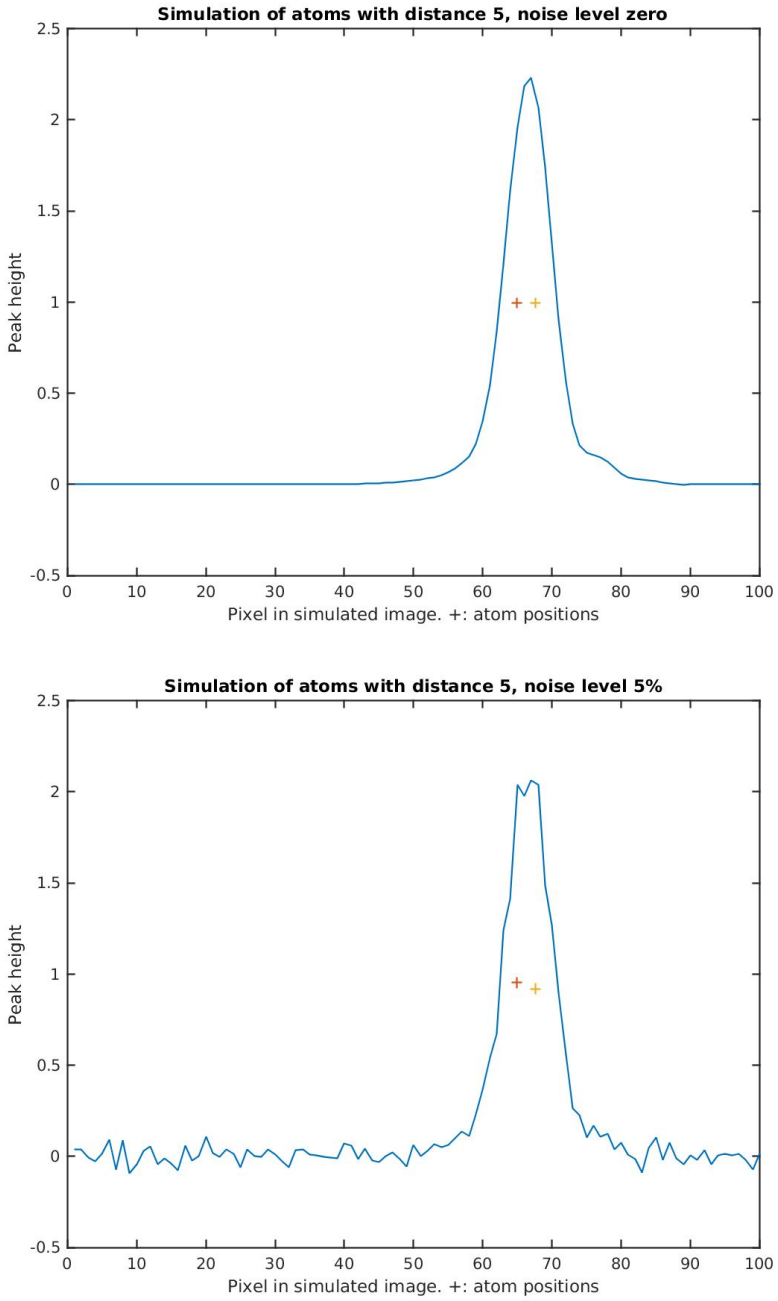

Figure 5.2.: Simulations of different atom configuration and noise setups. In the upper image, the overlap of two pure linespread functions is shown, as the noise level is zero. In these settings, the right positions can always be reconstructed.

in the picture to this error bar, we compare noiseless images $I_k^{(1,2)} = \sum_i L_{ki} x_i^{(1,2)}$, where both $x^{(1,2)}$ only contain only a single non-zero entry which is shifted by n with respect

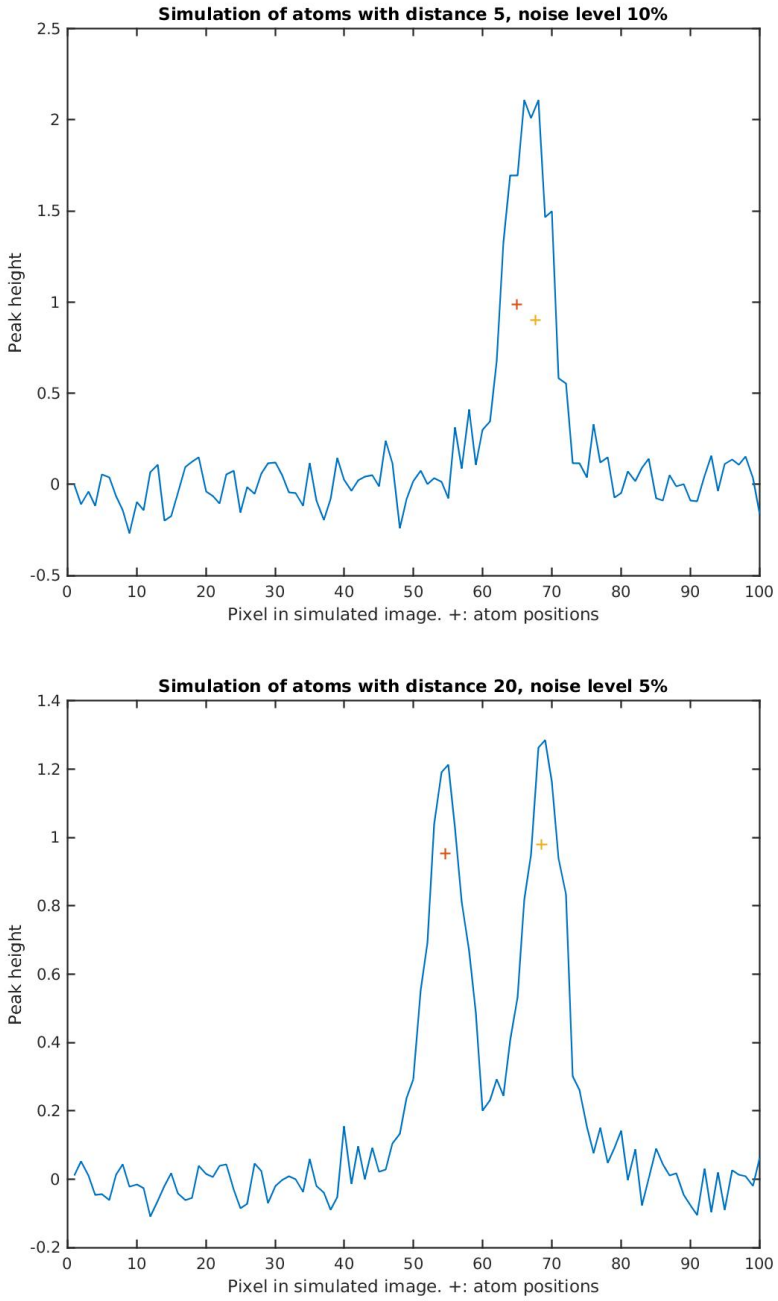

Figure 5.3.: In the upper image, a noise level of 10 % is simulated to benchmark the reconstructability. The lower image shows atoms that are 20 lattice sites apart. Still, their peaks are connected and they lie in one region of interest. The height of the peaks drops, as the two maxima do not overlap anymore.

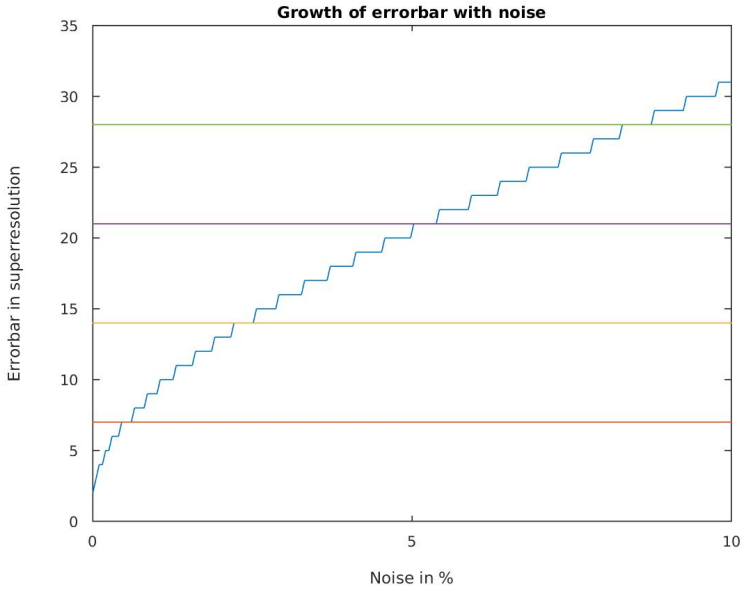

Figure 5.4.: Dependence of the size of the error bar with the amount of noise in %. The size of the error bar is given in superresolution—the pixels are cut in 10 sections each. The lattice distance corresponds to approximately 7 steps in superresolution, or 0.7 px, and are indicated by coloured lines.

to each other $x_i^{(1)} = x_{i+n}^{(2)} = 1$. We know that we cannot distinguish two images, if the

$$\left\| I^{(1)} - I^{(2)} \right\|_2 \leq \epsilon\sqrt{N}, \tag{5.26}$$

where ϵ is the standard deviation of the noise present in the image and N is the resolution (or length) of I. We obtain the error bar by finding the maximal n, such that Eq. (5.26) is still fulfilled. Our error bar in position is then exactly n (in superresolution).

This highly depends on the linespread function. For the datasets that were provided, these error bars are about 2-3 pixels in normal image resolution. As this compares to 3-4 lattice sites, it is impossible to distinguish these configurations using the linespread function at hand, see also Fig. 5.4.

5.3.4. Image Preparation

To perform a reconstruction of images taken from the experiment, the pictures have to be prepared to reduce the size of the error bars. In this preparation, two main goals are achieved. In a first step, the background noise is removed. The origin of this background noise is the light that is scattered randomly during the exposure and hits the detectors.

The second part of the preparation consists of detecting so-called Regions of Interest (ROI). These regions of interest are the parts of the image, that actually contain atom peaks. Since the lattice is filled sparsely, most parts of the image only contain background noise. If these parts would be added to the reconstruction, the total amount of noise would grow linearly thus weakening the error bars. Furthermore, by reducing the effective size of the system to be reconstructed, we improve the speed of the reconstruction.
For further discussion, let I be the array that contains the measured picture. To remove the baseline, we average over the entries of I that are lower than a certain threshold, namely the minimal absolute value of I plus two times an error margin that is estimated from the variance of an empty part of the picture. This average value is then subtracted from the array I. The resulting picture is fluctuating around zero. This way, the additive noise in the picture is very low.
The estimation of ROIs is more involved. In a first step, the cumulative sum of I is calculated. Whenever an atom peak is present in the picture, this cumulative sum has a high slope, while staying constant whenever there is no atom present. This is of course only true if the baseline is already removed. To detect high slopes, the gradient of the cumulative sum of I is estimated, which results in a smoothed version of I, where the noise influence is less strong. Now, the maxima of this smoothed I are evaluated up to a low multiple of the background noise and contribute to the ROIs. The high contributions that originate from one area of the image are included in one single ROI. Empty ROIs, that don't contain a contribution of an atom and are detected from noise artefacts are removed due to low cumulative sums. Afterwards, the remaining ROIs are enlarged by the size of the LSF, such that the positioning of the atoms in the pictures with reduced sizes is not imposed and still done in the reconstruction itself. The algorithm searches for lower peaks in the image by removing the peaks detected first from the sample and searching for peaks again. When all peaks are detected, their distance is evaluated and their ROIs are either combined to one big ROI, if they overlap or the peaks are separated in two distinct ROIs. From regions that are not covered by the ROIs, the noise estimated is estimated again and compared to the original estimate. Also, ROIs that contain peaks that are not fully part of the image, because the original atom was on the border of the area covered by the detectors, are removed. The positions of these atoms are not reconstructed. However, the dataset contained only a few (1 %) of these cases. The preparation procedure also detects the empty pictures, that don't contain any atoms.

The procedure is performed on both simulated as well as measured pictures, as the simulation produces images that resemble the measured pictures closely and also contain background noise. Both types of pictures are treated the same way and the preparation procedure is necessary as its output format is used in the reconstruction procedure, which will be described in the following section.

5.4. Reconstruction

The principal idea in our compressed sensing approach is to use the sparsity constraint as additional information. We want to find the solution that fits the measured distribution best under the constraint, that there is a only certain, small, number of particles in the lattice. We don't restrict possible solutions to lattice sites, but the estimated solutions should lie on lattice sites up to our estimation error which makes the lattice distance recoverable as a consistency check.

We analysed the datasets at hand using our trace norm minimisation compressed sensing algorithm and the linespread function provided in the data. We compared our results to the provided benchmark results (for the data sets of close-by atoms) and estimated the lattice distance from the pictures with distinct peaks from far apart atoms.

5.4.1. Iterative Hard Thresholding

Our first reconstruction uses a simple gradient method where in every step we only keep the positions p that contribute with the highest amplitude, where p is our sparseness constraint. In this case the number of atoms in the lattice, which enters as an initial guess. The iteration stops, if the error in two-norm between the reconstructed picture for the estimated positions and the actual image is below some small constant, e.g. 10^{-2}. For a typical image, this is normally achieved after approximately 50 steps of iteration.

To select the highest amplitudes, we sort for size and additionally forbid two peaks to be in the area that is covered by the width of the linespread function.

This reconstruction however only succeeds in the case of atoms that are relatively far separated, otherwise it always underestimates the distance.

5.4.2. Trace Norm Minimisation

We use a reconstruction method where we minimise the l_1-norm of our estimation of a signal vector $\vec{x}$ while bounding the l_2 norm of the deviation of the measured image and the estimation $L\vec{x}$ by an error depending on the standard deviation of the noise in the image to be reconstructed. Using this algorithm, we reconstruct the possible positions of the atoms, that are—taking into account the error bars—in agreement with the positions of the reconstructions given in the benchmark with a high rate. (100 % in the case of the dataset for distance 2, 98 % in the case of the dataset of distance 1.)

These reconstructions were done on the subset of pictures with atom number two, estimated by integrating over the region of interest. We find these estimates in a one step algorithm. It has to be noted, that since the error bars are of the size of about three times the lattice constant, distance 1 and 2 cannot be distinguished with certainty using any reconstruction method as the information is not in the picture. The trace norm minimisation always underestimates the distance in superresolution. This could be overcome by implementing information about the lattice distance, which our current reconstruction method does not use. However, standard compressed sensing methods do not come to their full potential here, as the linespread function is very coherent. A more incoherent map would spread the information over the whole image and would allow for significantly smaller error bars, that would make it possible to distinguish the exact distance of nearby atoms.

One possible way of implementing said incoherence without changing the current measurement setup too much is by implementing masks or filters in the optical setup. This option has yet to be investigated.

5.4.3. Moving the Camera

Additionally, an approach which could help with making the reconstruction more reliable is moving the camera in a subpixel way while taking the picture of the atoms. This would reduce the error in the offset between the pixels and the lattice sites and thus make the estimate better. Also, blind deconvolution techniques could be applied to reconstruct the LSF with high resolution with error bars. For low noise settings, where the offset between atoms and pixels is the main source of error, this can help in the reconstruction of positions in superresolution. In the present experiment however, the contribution of the offset error is very small. To use this extra information, each position is measured ten times while slowly moving the camera over the lattice. The resulting images are then each reconstructed and the transition of the reconstruction hopping from one pixel to the next indicates the exact position of the atom. This approach works whenever the error bar is lower than one pixel, and could be used in future measurements. The code to perform the reconstruction can be found in the appendix.

5.4.4. Nuclear Norm Minimisation

In [2, 78] Recht et al. introduce an approach on how to reconstruct pure Fourier type images by minimising the nuclear norm. As the measurement setup could be changed easily to capture in the Fourier plane, we followed this approach to test whether the reconstruction would be improved by this change. In the Fourier plane, the convolution

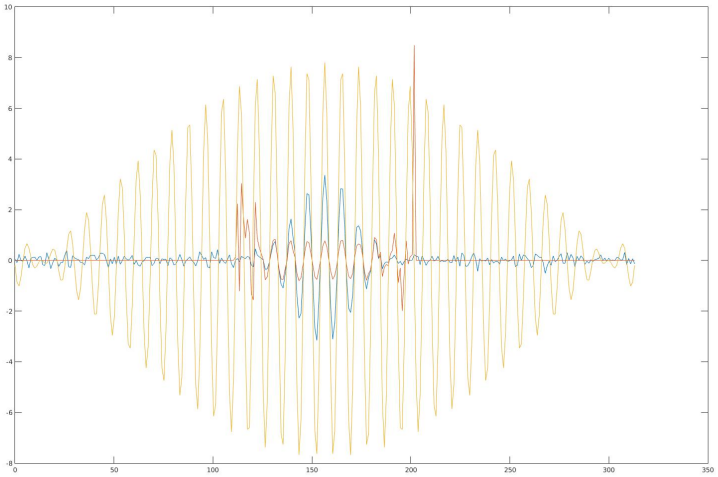

Figure 5.5.: Picture and linespread function in Fourier space. The noisy image is shown in blue, while the linespread function is indicated in red. The frequency of the envelope of the big orange curve reflects the original signal that is to be reconstructed. However, the width of the linespread function allows us to only have access to the centre part of the convolution (as it is zero outside). This forbids the reliable reconstruction of the frequency of the original signal using the atomic norm approach of Recht et al.

of the signal with a linespread function translates to a simple multiplication. To eliminate the influence of the linespread function, we calculated the Fourier transform of a fit to the linespread function. However, the placement of the cutoff of the linespread function proves to have an extreme influence on the resulting picture, as high frequency border regions hide the information in the picture. It is thus not possible to get more information from the measurement without putting in additional information in the first place, which makes this approach less valuable, as a perfect Fourier type image cannot be achieved. For perfect settings, without any noise, the method reconstructs as well as the compressed sensing approach in real space. The code on how to perform the minimisation can be found in the appendix.

5.4.5. Performance Analysis and Benchmarking

To test our algorithm, we set up a simulation of the measurement setup introduced above, which uses the measured data for the linespread function to generate artificial images where the positions and amplitudes are known and can be compared to the reconstructed

positions and amplitudes. In the noiseless case, the reconstruction is of course always perfect. We add statistical noise with a standard deviation similar to the noise in the picture sets to our simulated pictures. For single atoms, this still works remarkably well, also in superresolution, as the RIP-constant for sparsity 1 is reasonably small. For higher sparsities and atoms that are close to each other, the distance is often underestimated (within the relatively big error bars), see Fig. 5.6.
When applied to real experimental images of one atom, the algorithm converges to distinct positions and amplitudes, which are in high agreement to the ones reconstructed using the two step algorithm.
From the images of atoms with larger distances, we reconstructed a lattice distance by sampling over the reconstructed distances (see B.2) of 0.69 px in agreement with the experiment.

5.5. Outlook and Further Ideas

In this chapter, we presented the limits and possibilities of using a Compressed sensing approach to reconstruct positions and amplitudes of atoms in optical lattices from fluorescence images. We used a simulation of the experimental measurement to compare different reconstruction procedures and developed an error model derived from the information content in the images. We performed benchmarking of the compressed sensing reconstruction procedure on experimental images and were able to reconstruct the source distribution in great accordance with the results from the two-step reconstruction method. Using the simulation of images, we investigated different extensions to the measurement setup that improve reconstructability in low noise regimes and for different measurement maps. To remove the error caused by the offset between the discrete lattice and the discrete detector setup, the camera with the detector plane could be moved slowly while the pictures are taken. This results in a set of images containing different subpixel shifts of a single atom. A reconstruction of this set of images allows the placement of the atom with subpixel resolution. This step should be taken as soon as single pixel resolution is achieved. Another idea is to move the measurement plane in the Fourier plane of the microscope. This way, rather than capturing the position distribution, the Fourier transform of the discrete source distribution is captured. Together with a suitable linespread function, this bears the possibility of performing an atomic norm reconstruction with improved error bars. However, the current linespread function is not suitable for this approach, as information is lost during the convolution, resulting in a very small cut of the Fourier transform of the source function which does not contain enough information for a reconstruction.

We have shown that to further reduce the size of the error bars in the setting at

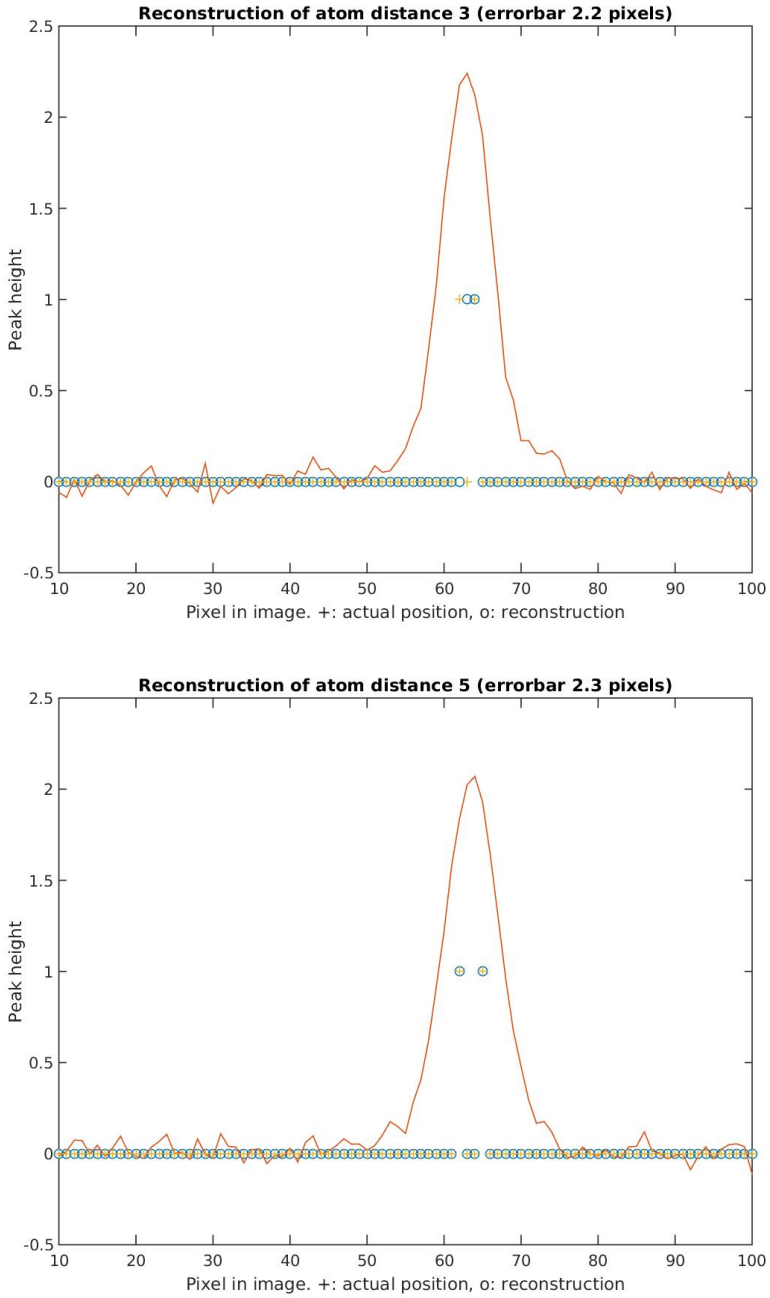

Figure 5.6.: Reconstructions and actual positions of simulated pictures with atoms of distances 3 and 5. The error bar is indicated in the title. The reconstruction reveals the right positions as soon as the distance of the atoms is distinguishable—meaning above two times the error bar. For individual peaks, the right position is reconstructed most of the times.

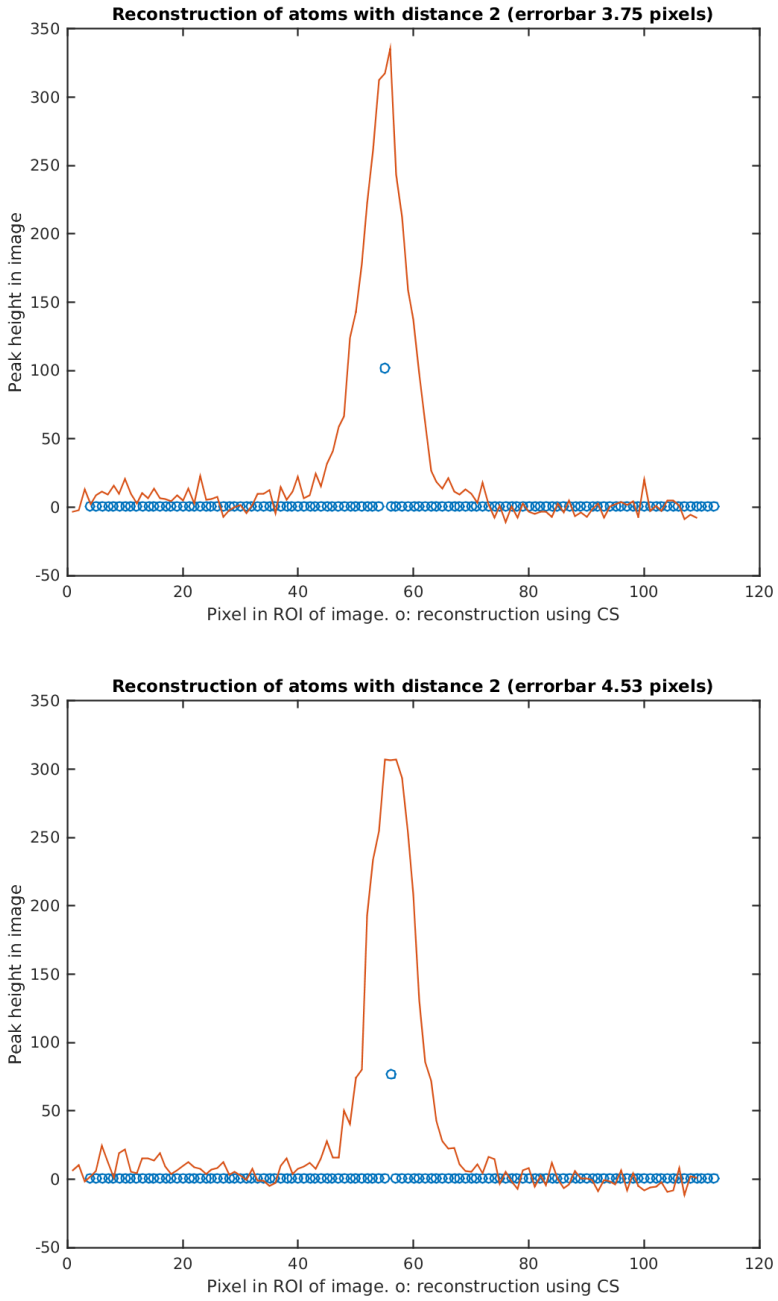

Figure 5.7.: Reconstructions and actual positions of pictures 9 and 10 of a dataset with atoms of distance 2. Again, the error bar is indicated in the title. The two atoms are placed at the same position (less then a pixel apart) and the actual distance of the atoms is underestimated, as could also be seen in the simulated pictures in Fig. 5.6

hand, the background noise has to be reduced, as images of close-by atoms contain the same information for distinct atom positions. Bringing down background noise is a very costly process, as the lenses and filters often have to be manufactured in the labs. However, there could be an easier, even though less intuitive way. In order to improve the distinguishability of atoms in neighbouring lattice positions, these settings have to produce images, that contain different information. One hope to achieve this and to also get the greatest benefit from the reconstruction using Compressed sensing techniques is to enhance the incoherence of the measurement map. Similar to the Fourier approach described above, this incoherent measurement map would be very hard to interpret for a human eye and brain. However, as close-by source signals would be mapped to very distinct images, the information about the positions is less affected and could thus be reconstructed using a computer. One way to enhance the incoherence of a measurement is to place random diffraction filters or masks in the optical setup before the source distribution undergoes the measurement process [53].

Another approach in getting a more incoherent measurement map is to unfocus the microscope used in the fluorescence imaging. This would produce a far field image by blurring the image. However, the blurring might apply to all positions in the same way which doesn't help to distinguish them, as it does not include additional information in the picture. In the atomic norm approach we saw however, that for a Fourier type image the envelope from the linespread function has to be wide enough. Combining the unfocused microscope with a shift of the measurement plane from position to Fourier space has the potential to improve the resolution of close-by atoms using the atomic norm approach.

An intriguing extension of the approach described in this chapter is to reconstruct not only the source distribution containing the positions and amplitudes but to include the measurement map L. This approach is known as blind deconvolution or Compressed Sensing 3.0. To obtain a reliable high precision estimate for the source function, it is crucial to know L to high precision. This reconstruction of the measurement map could be run on certified single atom pictures at different positions. Blind deconvolution could serve as a catalyst to unravel the information about L, that is hidden in these noisy pictures. As different offsets between lattice spaces and detectors would be contained in the different pictures, this could allow us to achieve certified subpixel resolution reconstructions with reliable error bars from a small number of pictures. Especially when applying filters to implement coherence in the measurement map, this approach would make it possible to overcome limitations that appear insurmountable otherwise.

It is important to understand that reconstructability is the same as invertibility of the measurement process. The reason for the loss of information that happens during the convolution and the following discretisation lies in the invertibility of the weighed sum of

unknown values. Using incoherent, e.g. seemingly random prefactors, which are reliably known with great precision is as important as reducing unwanted noise. All these factors have a key role to allow a high resolution reconstruction of the information contained in an image by inverting the influence of the measurement process.

6. Tomography of Properties – Gaussianity in Optical Lattices

The following work was done in collaboration with Mathis Friesdorf, Carlos Riofrío and Jens Eisert at FU Berlin [48]. The data that motivated this work was produced in the quantum optics laboratory in Munich by Ulrich Schneider, Simon Fölling and other colleagues, in the group of Prof. Bloch [42]. Most measurements are nowadays conducted to confirm a theoretical prediction. The experimental realisation aims to reproduce an outcome that can be derived from a simplified model of the actual experimental situation. After the measurement, the experimental outcome is compared to the theoretical prediction and deviations either hint at unwanted effects or noise in the experimental setup or at a wrong or incomplete theoretical model. In this chapter, we will follow a slightly different approach. We will investigate to what extent a measurement result can be interpreted with implying as little as possible on the underlying system. Introducing the Gaussianity of a quantum state, we will show how the information about this property is transformed during the measurement and can be deduced from incomplete measurement data. After introducing our tool of choice to measure Gaussianity, the relative entropy, we will describe how it can be used to efficiently distinguish quantum states. Subsequently, we will discuss Gaussianity in the context of cold atoms in optical lattices. We show results for simulated data and discuss the loss of information during time-of-flight measurements.

6.1. Tomography without Assumptions

Assumptions on properties of systems can change the way we interpret possible outcomes. Especially in the interpretation of measurement data of complex quantum systems, when assuming certain properties of the system in question we also assume that the devices used in the course of the measurement can be controlled perfectly. It would be preferable, however, if we could determine certain properties of quantum systems without introducing assumptive information about the system or the context of the measurement.

One field of study that follows this approach is the one of device independent tomography [5]. There, one characterises the systems solely on their measurement statistics

and the observed correlations and takes the underlying systems and dimensions as unknown. In this chapter, we follow this approach to characterise the Gaussianity of a measured state from partial data only. We rigorously avoid assumptions and find lower bounds using tools from statistical theory as well as quantum information science.

6.1.1. The Relative Entropy

A promising tool to measure the Gaussianity of a measured state, given only its second and fourth moments and without further assumptions on the system is an estimation of a lower bound on the minimal relative entropy between the state ρ and a Gaussian state σ. The minimal relative entropy is given by

$$\min_{\rho,\sigma} \mathrm{D}(\rho||\sigma) = \min_{\rho,\sigma} \left(\mathrm{Tr}(\rho \ln \rho) - \mathrm{Tr}(\rho \ln \sigma)\right). \tag{6.1}$$

Since ρ is a physical state of which we know second and fourth moments, we have to add several constraints to this minimisation problem. Let $r_i = x_1, \ldots, x_n, p_1, \ldots, p_n$ be the $2n$ canonical coordinates for our quantum system. Then the constraints are the following

$$\begin{aligned} \langle r_i \rangle_\rho &= 0 && 0 \leq i \leq 2n \\ \langle \{r_i, r_j\} \rangle_\rho &= \gamma_{ij} && 0 \leq i \leq 2n \\ \langle r_i r_j r_k r_l \rangle_\rho &= m^{(4)}_{ijkl} && 0 \leq i \leq 2n \end{aligned} \tag{6.2}$$

where the first can be fulfilled easily by shifting the variables to a new set of canonical variables with vanishing first moments. Instead of minimising over all possible Gaussian states σ, we constrain this set to the Gaussian state σ_ρ with the same second moments as ρ. By this assumption, we can simplify the minimisation problem, since $\mathrm{Tr}(\rho \ln \sigma) = \mathrm{Tr}(\sigma_\rho \ln \sigma)$. Thus (6.1) can be written as follows

$$\begin{aligned} \min_{\rho,\sigma} \mathrm{D}(\rho||\sigma) &= \min_{\rho,\sigma} (\mathrm{Tr}(\rho \ln \rho) - \mathrm{Tr}(\sigma_\rho \ln \sigma_\rho) + \mathrm{Tr}(\sigma_\rho \ln \sigma_\rho) - \mathrm{Tr}(\sigma_\rho \ln \sigma)) \\ &= \min_{\rho,\sigma} (\mathrm{Tr}(\rho \ln \rho) - \mathrm{Tr}(\sigma_\rho \ln \sigma_\rho) + \mathrm{D}(\sigma_\rho||\sigma)) \\ &= \min_{\rho} (\mathrm{Tr}(\rho \ln \rho) - \mathrm{Tr}(\sigma_\rho \ln \sigma_\rho)) + \min_{\sigma} \mathrm{D}(\sigma_\rho||\sigma) \\ &= \min_{\rho} (\mathrm{Tr}(\rho \ln \rho) - \mathrm{Tr}(\sigma_\rho \ln \sigma_\rho)), \end{aligned} \tag{6.3}$$

since $\mathrm{D}(\rho||\sigma) \geq 0$ and $\mathrm{D}(\rho||\sigma) = 0$ if and only if $\rho = \sigma$. The initial problem is thus simplified to minimise (6.3) under the same constraints as above, with known second moments of the Gaussian state σ_ρ.

6.1.2. Decomposition of Gaussian States

As introduced in previous chapters, a Gaussian state σ is completely characterised by its first and second moments. As its covariance matrix γ is a positive definite, real, symmetric $2n \times 2n$ matrix, there exists a symplectic transformation $S \in Sp(2n, \mathbb{R})$ such that $S^T \gamma S = \mathrm{diag}(\eta_1, \ldots, \eta_n, \eta_1, \ldots, \eta_n)$. For Gaussian states this property leads to the fact that the state can be written as a product of single mode states. This can be seen looking at the Wigner density function of the Gaussian state. The Wigner density function of σ for vanishing first moments is given by

$$W_\sigma(\mathbf{r}) = \sqrt{\frac{1}{(2\pi)^{2n} \det(\gamma)}} \exp\Big(-\frac{1}{2}\mathbf{r}^T \gamma^{-1} \mathbf{r}\Big). \tag{6.4}$$

Since we know that the covariance matrix γ can be diagonalised, we can derive the following property of the inverse of the covariance matrix:

$$S^T \gamma S = \mathrm{diag}(\eta_i) \to \left(S^T \gamma S\right)^{-1} = \mathrm{diag}(\eta_i^{-1}) \to \gamma^{-1} = S\ \mathrm{diag}(\eta_i^{-1}) S^T \tag{6.5}$$

If we now choose new canonical variables (the symplectic transformation leaves the commutation relations invariant) $\tilde{\mathbf{r}} = S^T \mathbf{r}$, the Wigner function can be written as a product over "single mode" Gaussian states:

$$W_\sigma(\tilde{\mathbf{r}}) = \frac{1}{(2\pi)^n \prod_{i=1}^n \eta_i} \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{\tilde{x}_i^2 + \tilde{p}_i^2}{\eta_i}\right) = \prod_{i=1}^n W_{\sigma_i}(\tilde{\mathbf{r}}). \tag{6.6}$$

Thus the corresponding Gaussian state σ can be written as the normalised (C_i) product over single mode Gaussian state.

$$\sigma = \prod_{i=1}^n \frac{C_i}{2\pi\eta_i} \exp\left(-\frac{\tilde{x}_i^2 + \tilde{p}_i^2}{2\eta_i}\right) = \prod_{i=1}^n \sigma_i \tag{6.7}$$

with $C_i = 2\pi\eta_i \sinh(1/2\eta_i)$. We can further simplify this expression by making the transformation to the usual creation and annihilation operators, which is given by

$$\tilde{x}_i = \frac{1}{\sqrt{(2)}} \left(b_i + b_i^\dagger\right) \qquad \tilde{p}_i = -i \frac{1}{\sqrt{(2)}} \left(b_i - b_i^\dagger\right) \tag{6.8}$$

obtaining

$$\sigma_i = \left(1 - e^{-\frac{1}{\eta_i}}\right) e^{-\frac{b_i^\dagger b_i}{\eta_i}}. \tag{6.9}$$

6.1.3. Estimating the Gaussianity

To bound the Gaussianity of the state ρ, we would like to get a lower bound for the relative entropy of the states ρ and σ. In the last section we learnt, that a Gaussian state

σ can be composed into a product of single mode states σ_i. Earlier, we introduced the Gaussian state σ_ρ with the same first and second moments as the state ρ, which we now can compose into single mode states by a transformation depending on the symplectic transformation that diagonalises the covariance matrix γ symplectically. Since the relative entropy is given by

$$\mathrm{D}(\rho||\sigma) = \mathrm{Tr}(\rho \ln \rho) - \mathrm{Tr}(\sigma_\rho \ln \sigma_\rho) = -\mathrm{S}(\rho) + \mathrm{S}(\sigma_\rho), \tag{6.10}$$

S being the von Neumann entropy, D has the same properties as S. As the trace is invariant under unitary transformations, we can transform to the space, where σ_ρ can be written as a product over single mode states. We then can make use of the following property of the relative entropy:

$$\mathrm{D}(\rho||\sigma_1 \otimes \sigma_1) \geq \mathrm{D}(\mathrm{Tr}_2(\rho)||\sigma_1) + \mathrm{D}(\mathrm{Tr}_1(\rho)||\sigma_2) \tag{6.11}$$

If we apply this inequality repeatedly, and apply it on our initial problem and use the unitary transformation belonging to the symplectic transformation S, we get:

$$\mathrm{D}(\rho||\sigma_\rho) = \mathrm{D}(\mathrm{U(S)}\rho\mathrm{U(S)}^\dagger||\otimes_{i=1}^n \sigma_{\rho,i}) \geq \sum_{i=1}^{n} (\mathrm{Tr}_{\{j\},j\neq i}(\mathrm{U(S)}\rho\mathrm{U(S)}^\dagger)||\sigma_{\rho,i}) \tag{6.12}$$

We can thus find a lower bound for the initial problem by minimising the single mode problem for the given moments of ρ. The symplectic transformation S simplifies the second moments for ρ, since its covariance matrix is the same as the covariance matrix of σ. However, one cannot hope to get easier expressions for the higher moments of ρ. The constraints for the higher moments of ρ have to be adapted using the symplectic transformation, which we will see in a following section.
To simplify the problem, we will use that a pinching operation can only increase the entropy. This is the case since from Schur's theorem we know that

$$\mathrm{diag}(\rho) \prec \lambda(\rho). \tag{6.13}$$

The function $x \ln x$ is convex and by that, $\mathrm{Tr}\,\rho \ln \rho$ is Schur convex. We thus know that

$$-\mathrm{S}(\rho) = \sum_i \lambda_i \ln \lambda_i \geq \sum_i \rho_i \ln \rho_i. \tag{6.14}$$

For further simplification, we loosen our constraints regarding the second and fourth moments. By that, we lower the relative entropy, since

$$\min_{\rho \in A^*} \mathrm{Tr}(\rho \ln \rho) \geq \min_{\rho \in A} \mathrm{Tr}(\rho \ln \rho) \quad \text{for} \quad A^* \subseteq A. \tag{6.15}$$

We now minimise over all states ρ subject to second and fourth moments, where the numbers of creation and annihilation operators are the same. This allows us to severely

reduce the set of moments we have to take into account. We will come to the explicit selection of these moments in the next section.

The full estimation of the relative entropy is summarised in the following inequality:

$$\begin{aligned}
\mathrm{D}(\rho||\sigma) &= \mathrm{Tr}(\rho \ln \rho) - \mathrm{Tr}(\rho \ln \sigma) \\
&= \mathrm{Tr}(\rho \ln \rho) - \mathrm{Tr}(\sigma_\rho \ln \sigma_\rho) \\
&\geq \sum_{i=1}^{n} \mathrm{D}(\mathrm{Tr}_{\{j\},j\neq i}\, \mathrm{U(S)}\rho \mathrm{U}^\dagger(\mathrm{S})||\sigma_{\rho,i}) \\
&\geq \sum_{i=1}^{n} \left(\sum_k \rho_k \ln \rho_k + \mathrm{S}(\sigma_{\rho,i}) \right)
\end{aligned} \tag{6.16}$$

where ρ_k are the diagonal elements of $\mathrm{Tr}_{\{j\},j\neq i}\, \mathrm{U(S)}\rho \mathrm{U}^\dagger(\mathrm{S})$. The entropy of the Gaussian state σ_ρ can be evaluated using the representation of σ_i we obtained earlier, $\sigma_i = \left(1 - e^{1/\eta_i}\right) e^{-b_i^\dagger b_i/\eta_i}$.

$$\begin{aligned}
\mathrm{Tr}(\sigma_{\rho,i} \ln \sigma_{\rho,i}) &= \sum_{k=0}^{\infty} \left(1 - e^{-1/\eta_i}\right) e^{-k/\eta_i} \ln \left[\left(1 - e^{-1/\eta_i}\right) e^{-k/\eta_i} \right] \\
&= \ln \left(1 - e^{-1/\eta_i}\right) - \sum_k \frac{1}{\eta_i} k \sigma_{\rho,i} \\
&= \ln \left(1 - e^{-1/\eta_i}\right) - \frac{1}{\eta_i} \mathrm{M}_2
\end{aligned} \tag{6.17}$$

The M_2 are the transformed version of the second moments of both state ρ and σ and can be obtained from the covariance matrix using the symplectic transformation S. To make life easier, we will clarify the used transformations and variables in the next section before we finally solve the single mode problem.

6.1.4. Transforming the Measurement Data

To keep track of the different variables we use here, this section concentrates on the relationships between $r, \tilde{r}, q, \tilde{q}$

$$\begin{array}{ccc}
\mathbf{r} = \begin{pmatrix} \hat{x}_i \\ \hat{p}_i \end{pmatrix} & \xrightarrow{T} & \mathbf{q} = \begin{pmatrix} a_i \\ a_i^\dagger \end{pmatrix} \\
S^\dagger \downarrow & & T^{-1} S^\dagger T \downarrow \\
\tilde{\mathbf{r}} = \begin{pmatrix} \tilde{x}_i \\ \tilde{p}_i \end{pmatrix} & \xrightarrow{T} & \tilde{\mathbf{q}} = \begin{pmatrix} b_i \\ b_i^\dagger \end{pmatrix}.
\end{array}$$

The transformation T is the linear mapping between the canonical variables and the bosonic operators. The symplectic transformation S diagonalises the covariance matrix

γ, where we chose the normalisation $\gamma = \langle\{r_i, r_j\}\rangle = \langle 2\,\mathrm{Re}(r_i, r_j)\rangle$. In a measurement, one measures densities $\left\langle a^\dagger a \right\rangle$ and density correlations $\left\langle (a^\dagger a)(a^\dagger a) \right\rangle$. To estimate the relative entropy from measured second and fourth moments, we thus have to transform the measured data using both the symplectic transformation and the linear mapping. The linear transformation T is given by:

$$T = \begin{pmatrix} \mathbb{1} & i\mathbb{1} \\ \mathbb{1} & -i\mathbb{1} \end{pmatrix} \qquad \text{with} \quad Tr = q \tag{6.18}$$

Using this transformation, we obtain the covariance matrix γ:

$$\gamma = 2\,\mathrm{Re}(rr^T) = 2\,\mathrm{Re}(T^{-1}qq^T T^{-T}) \tag{6.19}$$

γ is a Hermitian, real, positive matrix and can be diagonalised with the symplectic transformation S. This transformation is constructed following Narcovic [68]. For every positive definite, real, symmetric $2n \times 2n$ matrix γ, any root η of

$$\chi(\eta) = \det(\gamma + i\eta J) = 0 \quad \text{with} \quad J = \begin{pmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{pmatrix} \tag{6.20}$$

is real and non-zero; furthermore, $-\eta$ is also a root. We call the eigenvectors of this eigenvalue problem W_j, so that we have

$$(\gamma + i\eta_i J)\, W_j = 0, \qquad (\gamma - i\eta_i J)\, W_j^* = 0 \tag{6.21}$$

we now recombine these eigenvectors and introduce X_j and Y_j:

$$X_j = \frac{1}{2}(W_j + W_j^*) \quad \text{and} \quad Y_j = \frac{1}{2}(W_j - W_j^*). \tag{6.22}$$

Multiplication of γ with these vectors diagonalises γ and we get the symplectic transformation S:

$$S^T \gamma S = [X, Y]^T\ \gamma\ [X, Y] = \mathrm{diag}(\eta_1, \ldots \eta_n, \eta_1, \ldots \eta_n) \tag{6.23}$$

To obtain the transformed second and fourth moments, we need to get from the q to the $\tilde{q}$, since we used $M_2 = \left\langle \tilde{q}\tilde{q}^T \right\rangle$ and $M_4 = \langle \tilde{q}_i \tilde{q}_j \tilde{q}_k \tilde{q}_l \rangle$ where M_2 is a $2n \times 2n$ matrix and M_4 is a vector containing all $(2n)^4$ fourth moments for all combinations of $0 \leq i, j, k, l \leq 4$. The transformation $\tilde{S}^T = T S^T T^{-1,T}$ maps q to $\tilde{q}$ using $\tilde{q} = \tilde{S}^T q$. The measured moments can thus be transformed by

$$M_{2,\tilde{q}} = \tilde{S}^T M_{2,q} \tilde{S} \quad \text{and} \quad M_{4,\tilde{q}} = \tilde{S}_i^T \tilde{S}_j^T \tilde{S}_k^T \tilde{S}_l^T M_{4,q} \tag{6.24}$$

Since we pinch in the $\tilde{q}$ basis, we are only interested in the second and fourth moments with the same number of creation and annihilation operators. M_2 has the following form:

$$M_{2,\tilde{q}} = \begin{pmatrix} b_i b_j & b_i b_j^\dagger \\ b_i^\dagger b_j & b_i^\dagger b_j^\dagger \end{pmatrix}, \quad i,j \in 0,\dots n \tag{6.25}$$

After pinching, this matrix becomes bidiagonal. We are only interested in the lower diagonal of M_2 where $M_2 = \left\langle b_i^\dagger b_i \right\rangle$. Regarding M_4, we are interested in

$$M_4 = \left\langle (b_i^\dagger b_i)^2 \right\rangle = \left\langle b_i^\dagger b_i b_i^\dagger b_i \right\rangle = \langle q_{n+i} q_i q_{n+i} q_i \rangle, \quad i \in 0,\dots n. \tag{6.26}$$

We thus get n transformed second and fourth moments for the n single mode problems. In the following section, we introduce the method we used for solving the single mode problems.

6.1.5. Solving the Single Mode Problem

We still need to solve the single mode problem. We want to minimise the entropy of the state ρ to find a lower bound on the minimised entropy $D(\rho||\sigma)$ subject to the measured constraints.

$$\begin{aligned} \text{minimise} \quad & \sum_k \rho_k \ln \rho_k + \frac{1}{\eta_i} M_{2,i} - \ln\left(1 - e^{-\frac{1}{\eta_i}}\right) \\ \text{subject to} \quad & \sum_n \rho_n = 1 \\ & \sum_n \rho_n n = M_2 \\ & \sum_n \rho_n n^2 = M_4 \end{aligned}$$

To solve this problem, we look at the Lagrange dual problem. This is the function $g(\mu,\lambda) = \inf_x \mathcal{L}(x,\mu,\lambda)$ where μ and λ are Lagrange multipliers. This function is concave and is a lower bound to the original function. Since our original problem is convex, the duality gap between the two problems is zero and the optimal value for the dual problem is the same as the optimal value for the original problem. In our case, we have three constraints and thus need three Lagrange multipliers. The Lagrangian to our

problem is given by

$$\begin{aligned}\mathcal{L}(\rho,\mu_0,\mu_2,\mu_4) = &\sum_k \rho_k \ln \rho_k \\ &+ \mu_0\left(\sum_k \rho_k - 1\right) \\ &+ \mu_2\left(\sum_k \rho_k k - M_2\right) \\ &+ \mu_4\left(\sum_k \rho_k k^2 - M_4\right) \\ &+ \frac{1}{\eta}M_2 - \ln\left(1 - e^{-\frac{1}{\eta}}\right)\end{aligned} \tag{6.27}$$

The extremal points of this Lagrangian are of the form $\rho_k = e^{-(1+\mu_0+\mu_2 k+\mu_4 k^2)}$ for fixed μ_i. Thus the Lagrange dual function is given by:

$$g(\mu_i) = -\sum_k e^{-(1+\mu_0+\mu_2 k+\mu_4 k^2)} - \mu_0 - \mu_2 M_2 - \mu_4 M_4 + \frac{1}{\eta}M_2 - \ln\left(1 - e^{-\frac{1}{\eta}}\right) \tag{6.28}$$

If we maximise this concave function, we get the optimal value for our former problem. Each single mode problem can thus be reduced to a maximisation of a concave function with three variables.
The convergence of this function is non-trivial, so we performed numerical minimisation of $-g(\mu_i)$ using convex optimisation. We can thus find the optimal value of the dual optimisation problem and arrive at our estimate for the bound on the relative entropy. See the appendix for a description of the used optimisation. This leaves us with a method of efficiently characterising the Gaussianity of a quantum state by estimating the bounds on the relative entropy between the given moments of the measured state and the corresponding Gaussian state. We will now turn on investigating, how this approach can be applied in the situation of bosonic states.

6.2. Gaussianity in the Context of Optical Lattices

Ultra-cold atoms in optical lattices provide one of the most promising platforms for analogue quantum simulations of complex quantum many-body systems. Large system sizes can now routinely be reached and are already used to probe a large variety of different physical situations, ranging from quantum phase transitions to artificial gauge theories. At the same time, measurement techniques are still limited and full tomography for these systems seems out of reach. Motivated by this observation, we present a method

to directly detect and quantify to what extent a quantum state deviates from a locally Gaussian description, based on available noise correlation measurements from in-situ and time-of-flight measurements, as an indicator to what extent strong correlations in ground and thermal states are significant. We connect our findings to equilibration, disordered systems and the suppression of transport in Anderson insulators.

6.2.1. Ultra-cold Atoms

Ultra-cold atoms in optical lattices provide one of the most prominent architectures to probe the physics of interacting many-body systems. The parameters of the resulting Hamiltonians can be largely tuned in different regimes allowing to explore a wide range of physical phenomena in and out of equilibrium [13, 22, 81, 106]. They are one of the most promising platforms for realising quantum simulations and show signatures of outperforming classical computers for certain problems [12, 16, 95, 96]. Using state-of-the-art techniques, large scale systems with several thousand atoms can be controlled [13]; in fact, even states with specific initial configurations and atoms aligned on largely arbitrary shapes can be realised [39]. By modulating the optical lattice in time or by altering its geometry, a wide range of complex physical settings can be explored, ranging from probing quantum phase transitions [16, 96] to realising instances of artificial gauge theories [106].

While a large degree of control over these platforms has been achieved, at present, it is still true that the measurement capabilities are limited to an extent. This appears particularly relevant in the context of quantum simulations, where the result of the simulation has to be read out from the physical experiment. It seems clear that full quantum state tomography is infeasible, both for limits in measurement prescriptions as well as due to the unfavourable scaling of the effort of tomography with the system size. Suitable combinations [71] of tensor network tomography [29, 90] and compressed sensing schemes [54] suggest a way forward towards achieving tomographic knowledge, but at present such ideas have not been realised yet.

In the light of these obstacles, it seems imperative to focus the attention on developing tools to directly detect relevant properties of the quantum state, rather than trying to capture the full density operator – giving rise to information that is often not needed anyway. Among these, entanglement features come to mind that contain valuable information about a quantum state [6, 26, 38, 55], or notions of non-classicality that can be directly detected [66]. Similarly, it is of interest to identify to what extent the state realised corresponds to a ground or thermal state of an interacting model, and hence to what extent the state deviates from a Gaussian state. As one of the main promises of the field of ultra-cold atoms is to precisely study interacting quantum many-body models

deviating from non-interacting theories, this type of information is highly relevant.

In this work, we introduce a new scheme that can be used to directly estimate the local deviation from the Gaussianity of a state, based solely on second and fourth moments of particle number measurements. We apply this tool to the specific context of ultra-cold atoms and show that noise correlations in in-situ measurements are already sufficient for its calculation. In this way, we build upon and relate to the ideas of Refs. [4, 41, 42], but deliver an answer to the converse task: We do not show how interacting models are reflected in noise-correlations, but ask how data can be used to unambiguously witness such deviations from non-interacting models.

Gaussian States of Massive Particles

Interacting many-body quantum systems are exceedingly hard to capture and describe in terms of classical parameters. Non-interacting models – models that have Gaussian ground and thermal states – are an exception to this rule, in that the complexity of their description is low. They are a paradigmatic class of states both for fundamental questions in quantum information as well as for finding ground states of condensed matter models, such as the interaction-free Bose-Hubbard model [13]. At the same time they are states that do not exhibit the intricate structure of interacting quantum many-body models. Again, since one of the main promises of the field is to address such interaction effects, it seems important to have tools at hand to directly detect a deviation from Gaussianity.

6.2.2. Gaussianity of States in Optical Lattices

The bosonic Gaussian states discussed here are characterised by the second moments collected in the correlation matrix γ with entries

$$\gamma_{i,j} = \mathrm{Tr}\Big(b_i^\dagger b_j\, \rho\Big), \tag{6.29}$$

where $b_j^\dagger, b_j$ denote the canonical bosonic creation and annihilation operators, $j = 1, \ldots, n$. Throughout this work, we investigate massive bosons leading to a situation in which all $\mathrm{Tr}(b_i b_j \rho) = 0$ for all $i, j = 1, \ldots, n$. Any such correlation matrix satisfying $\gamma \geq 0$ can be diagonalised with a unitary $V \in U(n)$ as $D = V\gamma V^\dagger$, reflecting a mode transformation

$$b_j = \sum_{k=1}^{n} V_{j,k} \tilde{b}_k \tag{6.30}$$

preserving the bosonic commutation relations. One immediately finds the following convenient maximum entropy property: Gaussian states σ are maximum entropy states

given their correlation matrix, so that

$$\sigma = \underset{\substack{\rho\in\mathcal{D}\\ \mathrm{Tr}\left(\tilde{b}_j^\dagger \tilde{b}_k \rho\right)=\delta_{j,k}D_{k,k}}}{\arg\max}\, S(\rho), \tag{6.31}$$

where $S(\sigma) = -\operatorname{Tr}(\sigma \log \sigma)$ denotes the von-Neumann entropy and $\mathcal{D}$ denotes the set of density matrices. In addition, by invoking the pinching inequality [11], one finds that it is already sufficient to fix the diagonal entries $D_{k,k}$ for $k = 1, \ldots, n$

$$\sigma = \underset{\substack{\rho\in\mathcal{D}\\ \mathrm{Tr}\left(\tilde{b}_k^\dagger \tilde{b}_k \rho\right)=D_{k,k}}}{\arg\max}\, S(\rho). \tag{6.32}$$

Following from this, given the diagonal elements of the correlation matrix in the momentum representation, the following Gaussian state is uncorrelated over the individual modes and given by (6.33)

$$\sigma = \prod_{k=1}^{n} \sigma_k, \quad \sigma_k = \left(1 - e^{-\eta_k}\right) e^{-\eta_k \tilde{b}_k^\dagger \tilde{b}_k}, \tag{6.33}$$

where $\eta_k > 0$ is corresponding to $D_{k,k}$.

As described previously, the Gaussian state corresponding to measured second moments is given by

$$\sigma = \underset{\substack{\rho\in\mathcal{D}\\ \mathrm{Tr}\left(\tilde{b}_k^\dagger \tilde{b}_k \rho\right)=D_{k,k}}}{\arg\max}\, S(\rho). \tag{6.34}$$

This maximum entropy state is achieved by the generalised Gibbs ensemble

$$\sigma = \frac{1}{Z} \prod_{k=1}^{n} e^{-\eta_k \tilde{b}_k^\dagger \tilde{b}_k}\ , \tag{6.35}$$

where Z is the usual partition sum and the η_k are determined by demanding that

$$\mathrm{Tr}\left(\tilde{b}_k^\dagger \tilde{b}_k e^{-\eta_k \tilde{b}_k^\dagger \tilde{b}_k}\right) = D_{k,k}\ . \tag{6.36}$$

This expression can be calculated using the bosonic partition sum

$$Z(\{\eta_k\}) = \mathrm{Tr} \prod_{k=1}^{n} e^{-\eta_k \tilde{b}_k^\dagger \tilde{b}_k} \tag{6.37}$$

$$= \prod_{k=1}^{n} \sum_{j=0}^{\infty} e^{-\eta_k j} \tag{6.38}$$

$$= \prod_{k=1}^{n} \frac{1}{1 - e^{-\eta_k}}. \tag{6.39}$$

With this, we straightforwardly have

$$D_{k,k} = \frac{1}{Z}(-\partial_{\eta_k})Z \tag{6.40}$$

$$= (1-e^{-\eta_k})\frac{e^{-\eta_k}}{(1-e^{-\eta_k})^2} \tag{6.41}$$

$$= (e^{\eta_k}-1)^{-1} \tag{6.42}$$

and hence

$$\eta_k = \ln\left(1+D_{k,k}^{-1}\right) . \tag{6.43}$$

These states can be viewed as an instance of a generalised Gibbs ensemble [21, 27, 80]. Importantly for our context at hand, such Gaussian states also play a prominent role in the context of optical lattices, where they can, for example, be used to capture the superfluid ground state.

In recent years, research on cold atoms in optical lattices has progressed significantly, by now allowing for unprecedented control of interacting quantum many-body system, involving several thousands of atoms. Relying on recent experimental advances, the position of individual atoms can be tracked using single site addressing. Using such techniques, local expectation values of the particle number as well as density-density correlations can be resolved. These local measurements have already provided important insight into the out-of-equilibrium dynamics of quantum many-body systems [13] and allow to access their microscopic properties. It is the main result laid out subsequently to identify tools to detect a deviation from Gaussianity – reflecting a non-interacting system – based on simple particle number measurements, building upon Refs. [46, 47]. For this, we begin with a clarifying discussion to what extent states encountered in optical lattices can be Gaussian.

States in optical lattices describe massive particles. For that reason, the particle number in each experimental run is fixed. This, however, implies that the full state is not Gaussian. The best example for this is the perfect superfluid state, corresponding to the ground state of the 1D free hopping Hamiltonian

$$H_{\text{hop}} = \sum_{j=1}^{n-1}\left(b_j^\dagger b_{j+1} + b_{j+1}^\dagger b_j\right) . \tag{6.44}$$

Using a chemical potential $\mu > 0$ to maintain the expected particle number, thermal states of this Hamiltonian take the form

$$\rho \propto \exp\left(-\beta H - \mu\sum_{j=1}^{n} b_j^\dagger b_j\right), \tag{6.45}$$

where $\beta > 0$ denotes the inverse temperature. This ensemble is Gaussian, as it is the exponential of a quadratic expression of creation and annihilation operators. It, however, only fixes the particle number on average. In order to fix the actual particle number, meaning to ensure that the state lives on a fixed particle number sector, also all higher moments have to be included in the ensemble. However, already including a fixed variance with a Lagrange parameter $\mu_2 > 0$

$$\rho \propto e^{-\beta H - \mu \sum_{j=1}^{n} b_j^\dagger b_j - \mu_2 \left(\sum_{j=1}^{n} b_j^\dagger b_j\right)^2}, \tag{6.46}$$

results in a state that is no longer strictly Gaussian.

Nevertheless, the superfluid can be thought of as being Gaussian in one important sense. Local particle number measurements are indistinguishable from a Gaussian state for sufficiently large system sizes. Thus, many measurements of the state can be captured using a simple Gaussian description. In the following, we present a general method that allows us to estimate the local deviation of a state from Gaussianity, based solely on measurements of second and fourth moments, which will afterwards be applied to states capturing optical lattices.

6.2.3. Estimating the Local Deviation from Gaussianity

In order to approach the local deviation of states from Gaussianity in optical lattices, we first describe how the distance of the global state to the manifold of Gaussian states can be captured and estimated relying only on second and fourth moments. We then use this insight to define a local deviation from Gaussianity and evaluate it for paradigmatic bosonic models.

To quantify the deviation of a state from Gaussianity, we use the relative entropy, also known as Kullback-Leibler divergence, as a natural quantity with a precise statistical interpretation. The relative entropy between two states ρ and σ is defined as

$$S(\rho||\sigma) = \mathrm{Tr}(\rho \ln \rho) - \mathrm{Tr}(\rho \ln \sigma). \tag{6.47}$$

This quantity provides the asymptotic statistical distinguishability of ρ from σ in the situation of having arbitrarily many copies of the state available [97].

Based on this relative entropy, we define the global Gaussianity of a state as

$$G(\rho) := \min_{\sigma \in \mathcal{G}} S(\rho||\sigma), \tag{6.48}$$

where $\mathcal{G}$ denotes the set of all Gaussian states. This can be seen as a measure of the strong correlations present in the state, as quantifying the statistical deviation from a state that could have been the ground or thermal state of a non-interacting model. As

it turns out, the minimum is always achieved for the Gaussian state σ with the same correlation matrix [67]. We denote this special Gaussian state by σ_ρ. Such a global Gaussianity can be lower bounded by relying only on local measurement data, which will be performed in the following.

We begin by using the additivity of the relative entropy which allows us to move to the symplectic eigenbasis of σ_ρ and decompose the estimate into a problem involving individual eigenmodes [97]

$$\begin{aligned} G(\rho) = S(\rho||\sigma_\rho) =& S\left(U(\gamma)\rho U(\gamma)^\dagger||\otimes_{k=1}^n \sigma_k\right) \\ &\geq \sum_{k=1}^n S\left(\rho_k||\sigma_k\right), \end{aligned} \tag{6.49}$$

where $U(\gamma)$ denotes the unitary transformation in Hilbert space (so-called metaplectic representation) reflecting the moment transformation into the eigenmodes of γ, Tr_{k^c} denotes the reduction to the k-th mode and $\rho_k := \mathrm{Tr}_{k^c}\left(U(\gamma)\rho U(\gamma)^\dagger\right)$. For each individual mode, we can use the Gaussianity of σ_ρ and rewrite the estimate in terms of entropies [47]

$$G(\rho) \geq \sum_{k=1}^n \left(S(\sigma_k) - S\left(\rho_k\right)\right). \tag{6.50}$$

These single mode entropy estimates are a major simplification compared to the original problem.

Naturally calculating the entropy of ρ_k is still not a task that can be performed efficiently. Rather, we solely rely on fourth moments $\mathrm{Tr}\left(\tilde{b}_k^\dagger \tilde{b}_k \tilde{b}_k^\dagger \tilde{b}_k \rho\right) = M_{4,k}$ and calculate the smallest possible distance to the manifold of Gaussian states compatible with this data. In this way, we are able to calculate a minimal deviation from Gaussianity, thus showing that the state is, in this precise sense, strongly correlated. We therefore relax the problem as follows

$$G(\rho) \geq \sum_{k=1}^n \left(S(\sigma_k) - S\left(\kappa_k\right)\right), \tag{6.51}$$

$$\kappa_k := \underset{\substack{\kappa\in\mathcal{D} \\ \mathrm{Tr}\left(\tilde{b}_k^\dagger \tilde{b}_k \kappa\right)=D_{k,k} \\ \mathrm{Tr}\left(\tilde{b}_k^\dagger \tilde{b}_k \tilde{b}_k^\dagger \tilde{b}_k \kappa\right)=M_{4,k}}}{\arg\max} S(\kappa)\ , \tag{6.52}$$

where κ_k is the maximum entropy state compatible with the second and fourth moments. Using Schur's theorem [11], which states that the ordered eigenvalues of a matrix majorise the ordered diagonal entries

$$\mathrm{diag}(\rho) \prec \lambda(\rho)\ , \tag{6.53}$$

and the fact that the von-Neumann entropy is Schur convex, the entropy of this state can be upper bounded by only considering the diagonal

$$S(\rho_k) \leq \sum_i \rho_{k,i} \ln \rho_{k,i} \ . \tag{6.54}$$

This allows for an efficient solution of the optimisation problem using Lagrange multipliers. Thus, we have seen that the global Gaussianity defined in Eq. (6.48) can be lower bounded by using second and fourth moments in the symplectic eigenbasis of the state.

Based on this insight, we turn back to states describing massive particles in optical lattices. There, the full state cannot be reconstructed, which necessarily implies that important features of the state have to be addressed. Here we focus on a particularly simple quantity that only relies on measuring the particle number on a single site, which is accessible experimentally using single-site addressing [39]. Using such data, define the local deviation from Gaussianity on site j of a state as

$$G_{\text{local}}(\rho, j) := S(\sigma_\kappa) \underset{\substack{\kappa \in \mathcal{D} \\ \text{Tr}(n_j \kappa) = \bar{n}_j \\ \text{Tr}\left(n_j^2 \kappa - \bar{n}_j^2\right) = \overline{n_j^2}}}{\arg\max} S(\kappa) \ , \tag{6.55}$$

where n_j is the particle number operator on site j and $\bar{n}_j, \overline{n_j^2}$ denotes its experimentally measured expectation value and variance.

This local deviation from Gaussianity is not only easily measurable, but more importantly also yields relevant information about the quantum system at hand. As discussed above, the superfluid state is locally Gaussian in the sense that the quantity in Eq. (6.55) vanishes for large enough system sizes. In this way, calculating the local deviation from Gaussianity, which captures to what extent onsite particle measurements are compatible with a Gaussian state, is a natural way to quantify the distance to a perfect superfluid state. Moreover, for the important case of non-interacting particles, it can also be used to identify the suppression of particle propagation due to disorder. Both these applications are elaborated upon below.

6.2.4. Local Deviation from Gaussianity in the Bose-Hubbard Model

In the following, we present numerical results for bosonic models commonly encountered in optical lattices. We begin with the attractive Bose-Hubbard model

$$H_{\text{BH}} = -\sum_{j=1}^{n-1} \left(b_j^\dagger b_{j+1} + b_{j+1}^\dagger b_j \right) - \frac{U}{2} \sum_{j=1}^{n} n_j (n_j - 1) \ , \tag{6.56}$$

where we have chosen the hopping strength equal to unity and denote the interaction strength by U. The ground state for $U = 0$ is the superfluid state introduced above and

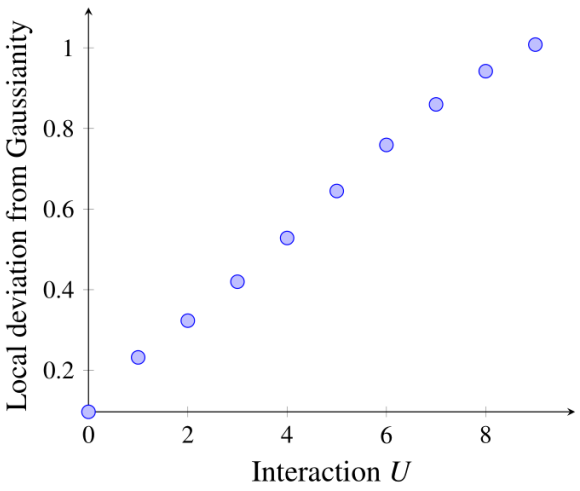

Figure 6.1.: Plotted is the local deviation from Gaussianity based on the measurement of particle number and density-density correlator on a single lattice site of a 1D system for the ground state of the attractive Bose-Hubbard model with filling fraction $\bar{n} = 1$ and interaction strength U. The results are obtained with exact diagonalisation using periodic boundary conditions on $L = 15$ lattice sites.

thus locally Gaussian in our sense, corresponding to a local deviation from Gaussianity of zero. In contrast, when the interaction strength is increased, the system becomes strongly correlated and the size of the local deviation from Gaussianity should increase.

We have confirmed that this behaviour is indeed encountered numerically and find an almost linear behaviour between the local deviation from Gaussianity and the interaction strength (see Fig. 6.1). In this way, one could even see the local deviation from Gaussianity as an experimental probe to directly measure the interaction strength based solely on ground state particle number fluctuations.

Another setting in which the local deviation from Gaussianity is intriguing to investigate is the evolution of free systems, where it can be used as an indicator of disorder and the concomitant suppression of transport.

6.2.5. Disordered systems

It is known that non-interacting systems which exhibit transport in a suitable sense evolve in time in a way that the states tend to be locally Gaussian following out of equilibrium dynamics [37], a feature that is true in surprising generality [27, 28, 51]. Thus, in this setting, the precise initial conditions are forgotten over time and local expectation values can be captured using only the second moments of the initial state,

then fully determining local expectation values.

This applies in particular to the free hopping Hamiltonian introduced in Eq. (6.44). A paradigmatic setting for this is given by an initial product state with one particle on every second site, which can be experimentally prepared employing optical superlattices [85, 95]. This initial state is the ground state of an infinitely strongly interacting Bose-Hubbard model and it is thus far from being locally Gaussian, as described above. During time evolution that has transport, however, the particles distribute evenly over the lattice, thus moving towards the manifold of locally Gaussian states.

For disordered systems, transport is strongly suppressed and the distribution of particles over the lattice thus does not take place. For concreteness, let us consider a simple 1D hopping model

$$H = -\sum_{j=1}^{n-1} \left(b_j^\dagger b_{j+1} + b_{j+1}^\dagger b_j \right) + \sum_{j=1}^{n} w_j b_j^\dagger b_j \ , \tag{6.57}$$

with local potentials w_j drawn uniformly from some interval $[-h, h]$. For $h = 0$ the model reduces to the free hopping and local Gaussification thus takes place. In contrast, when randomness is present, transport breaks down, resulting in a positive local deviation from Gaussianity even for long times. In Fig. 6.2 we see a initial state which is a charge-density wave corresponding to a Fock state with one atom on every second lattice site and the evolution is governed by the Hamiltonian in Eq. (6.57) with w_j drawn randomly from the interval $[-h, h]$ for three different values of the disorder strength. Initially, the local state is very non-Gaussian. For the translationally invariant system showing transport, information spreads through the lattice and the local measurements become compatible with a fully Gaussian description of the state, that is, the system dynamically gaussifies. In contrast, in the disordered case, transport is strongly suppressed and the non-Gaussianity remains locally visible.

6.3. Gaussianity from Time-of-Flight

The same approach discussed here for local deviation from Gaussianity can be applied to global properties of the lattice. In fact, it can be applied to any modes that are defined by a mode transformation

$$\tilde{b}_q = \sum_{j=1}^{n} V_{q,j} b_j. \tag{6.58}$$

In particular, it is applicable to time-of-flight measurements. If the quantum state is not translationally invariant due to the presence of a harmonic trap, an average Gaussianity over the system size is then directly detected in this way. In time-of-flight measurements

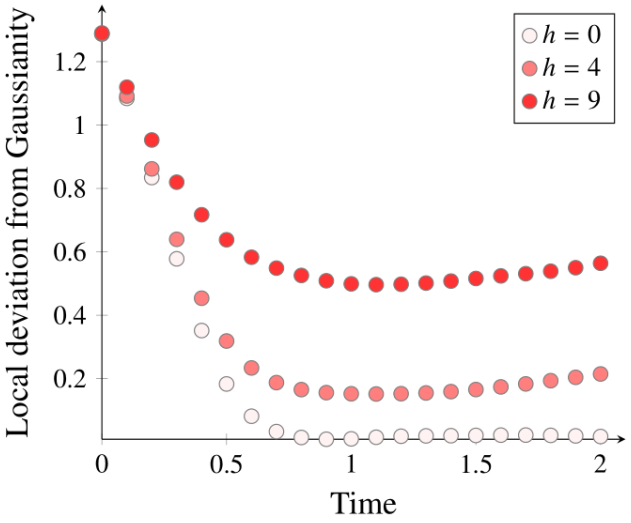

Figure 6.2.: Plotted is the time-dependence of the local deviation from Gaussianity based on the measurement of particle number and density-density correlator on a single lattice site of a 1D system of $L = 20$ sites. Shown is the average over 40 disorder realisations.

with finite accuracy of the camera pixels, only the diagonal elements $\left\{\langle n(q)\rangle = \left\langle \tilde{b}_q^\dagger \tilde{b}_q \right\rangle\right\}$ of

$$\Gamma = V\gamma V^\dagger \tag{6.59}$$

are measured. The resulting distribution is more commonly expressed as a function of q and t,

$$\langle n(q, t_{\mathrm{ToF}})\rangle = |\hat{w}_0(q)|^2 \sum_{j,k} e^{iq(r_j - r_k) - i\frac{c\left(r_j^2 + r_k^2\right)}{t_{\mathrm{ToF}}}} \left\langle b_j^\dagger b_k \right\rangle . \tag{6.60}$$

where t_{ToF} is the time-of-flight and $c > 0$ is a constant derived from the mass and the lattice constant of the optical lattice. The fourth moments of the same modes defined by V are accessible as $\left\{\left\langle \tilde{b}_q^\dagger \tilde{b}_q \tilde{b}_q^\dagger \tilde{b}_q \right\rangle\right\}$ and contained in the very same images from the laboratory, merely by computing higher moments, following a prescription of Ref. [4]. Let it be stressed again that in contrast to this reference, we aim for and provide a direct detection of correlations based solely on second and fourth moments.

Let us concentrate on the measurement process in more detail, and discuss the problems of extracting the expectation values of particle number operators and density-density correlations in lattice space from time-of-flight measurements.

6.3.1. Time-of-Flight Measurements

Over the course of the measurement process, information about the expectation values of the state in question is lost in certain steps. This entails, that the full information cannot be recovered. This means, that the original method of extracting a bound on the relative entropy between the quantum state in the system and the closest Gaussian reference state has to be adapted, since until now, we relied on exact measurements of fourth and second moments. This adaption worsens the tightness of the bounds on the relative entropy.

However, most parts of the measurement process only destroy information about the correlations but do not introduce new correlations in the data, where there are no correlations present in the system. This way our bound gets worse but not wrong. Given small enough errors introduced by the measurement, we could still reliably detect correlations using our method. Since however some parts of the measurement are not completely understood yet, we cannot fully claim that the measurement doesn't introduce new correlations; only that they don't seem to be big, since they are not visible in the measurement results. We will now consider the single processes of the time-of-flight measurement.

In a time-of-flight measurement, atoms that are confined in a optical trap potential are suddenly released. They start to move out of their positions according to their momentum and are allowed to spread freely for a certain time t_{ToF}. After this time-of-flight, their momentum distribution is imaged by taking a picture of the cloud and inverting the data. Each line in the resulting data array is then convolved with itself to get information about density-density correlations. Fig. 6.3 shows two images that are obtained in that way.

Electronic Noise and Photon Shot Noise

The electronic noise in the detector and the photon shot noise of taking the picture introduce a simple, but non-reversable error in the measurement outcome. The electronic noise can safely be neglected, as it is much smaller than the photon shot noise, which is around $\sqrt{n}$. Thus for the relation between the real density distribution in the cloud $n(\vec{r})$ we can say that $n_m(\vec{r}) = n_{PS}(\vec{r}) + n(\vec{r})$ with n_{PS} is randomly distributed of order $\mathcal{O}(n(\vec{r}))$. Since the error is uncorrelated, it won't introduce correlations to the signal, if there are enough samples. Since it is comparably big, it will very likely worsen the tightness of the bound on the estimated quantity.

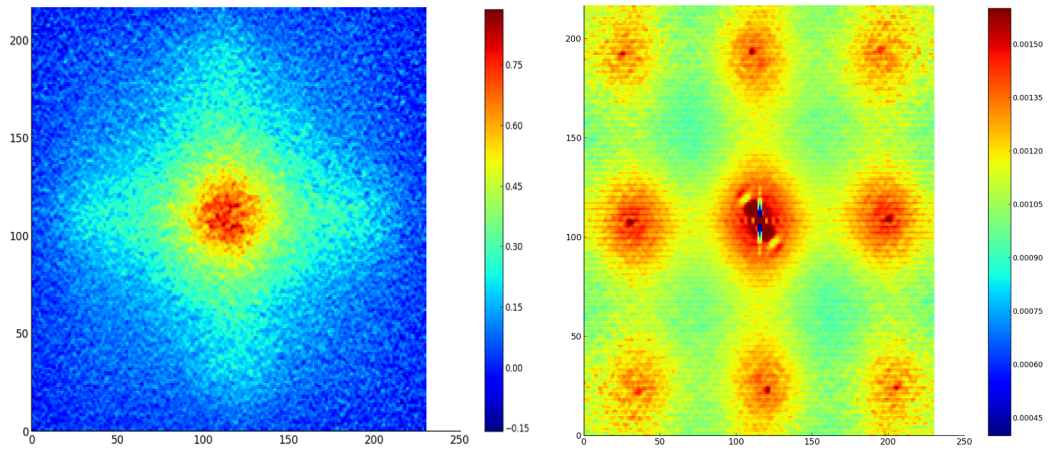

Figure 6.3.: On the left: density of the momentum distribution in a time-of-flight measurement. On the right: convolution of averages of momentum distributions showing correlation peaks typical for a Mott insulator.

Finite Detector Resolution

The finite resolution of the detector can be modelled by replacing the real density distribution of the momenta by a discretised version, where we integrate over small squares in the detector plane. This of course is also non-reversable. Whether it introduces correlations or not depends on "how discretised" the signal was before the measurement. For a continuous distribution, this would be bad, but since we detect the position of atoms in a cloud, a finite resolution, if small enough, should only introduce an error. The lattice spacing of the detection corresponds to $a_0 = 2.34\,\mu\text{m}$ in the cloud plane. All the atoms in a square of this size $a_0 \times a_0$ get mapped to one position in the cloud. This may have severe influence on the distribution of the noise, from which the density-density correlations are deduced. We believe that this is the main reason for the peaks being much lower than expected, which will influence the tightness of the bound. ($\mathcal{O}\left(10^{-4}\right)$ instead of $\mathcal{O}(1)$)

Point Spread Function

Every peak in the cloud plane gets spread during its way from the cloud plane to the detector plane. This effect can be modelled by a convolution of the original signal with a function F, which is close to Gaussian. The width of this function can be read of from the width of the correlation peaks, which should have a width of only one pixel. The influence of this process could be erased by neglecting everything but the centre of the peak for the numerical estimation, which would make the output highly dependent on

the respective choice. This does not seem like a reasonable approach. As discussed in the chapter on reconstructing positions of atoms in an optical lattice, the inversion of the point spread function is highly non-trivial. The low peak height in the measured pictures due to noise effects and information loss ($\mathcal{O}\left(10^{-4}\right)$ instead of the expected height of $\mathcal{O}(1)$) leads to a severe worsening of the bound. To obtain a better estimate one would have to integrate over a suitable area (the width of the peak), leading to closer estimate of the actual peak height and thus an improvement of the bound. However, this would lead to severe loss of information, as this integrated result could be produced by many different distributions and thus worsen the significance of the bound.

Integration via Tilted Axis

The detection is done via a photograph of the cloud which should be perpendicular to the detector plane. Since the cloud is three dimensional, there is a loss of information if the position of the particles in the cloud in this direction is not translationally invariant (which is not the case but might be a good assumption). Additionally, the axis is tilted, such that the integration also introduces a small average over different positions lying in the plane perpendicular to the optical axis. This error is similar to the discretisation introduced by the single detectors of the camera.

Finite Time-of-Flight

For infinite time-of-flight, the distribution in the cloud corresponds to the momentum distribution in the lattice. Since the time-of-flight is finite, there is an additional factor of e^{im/tr^2} in the relation between source and signal distribution.

This section concentrates on the errors introduced by the devices and setup used in the measurement process. There are also other difficulties concerning the intrinsic properties of time-of-flight measurements, which we will concentrate on in the next section.

6.3.2. Propagation of Information

In time-of-flight measurements of atoms initially confined in optical lattices, one measures the particle density $n(\vec{r})$ using the procedure described in the last section. For ideal measurements, the Quantity $n(\vec{r})$ is proportional to the momentum distribution in the trapped lattice state, since for infinite flying times (thus the starting point becomes irrelevant), the particle position only depends on the momentum of the particle at the time it got released. $n(\vec{r}) \approx m/(\hbar t) n_{\vec{Q}(\vec{r})}$ where $\vec{Q}(\vec{r}) = m/(\hbar t)\vec{r}$ is the relation between the position in the cloud and the momentum in the lattice [4]. $n_{\vec{Q}}$ thus gives us the number of particles in the lattice which had a certain momentum at the time of release. Now the

correlation function can be calculated from these measurement outcomes. By convolving the measured densities, one can see the fluctuations in the density-density-correlations which are expected to show up for Mott states. These are however suppressed by the errors, such as that it is important to carefully preselect the images to include in the calculations to get to see the fluctuations. This preselection is done by only allowing for a certain deviation in total particle number which is still large enough to guarantee for enough different pictures to get a reasonable expectation value. This preselection could have to be different for our purposes. In our method, we want to diagonalise the covariance matrix of the state, consisting of expectation values of the particle number operator at different lattice sites. In principle, we thus have to find a transformation from the number of particles having a certain momentum (thus position in the cloud) to their position in the lattice before the release.

The measurement of the atoms in the lattice projects the lattice state to good approximation into the lowest Bloch band via $A(\vec{r},t) = \sum_i w_i(\vec{r},t)a_i$ where a_i is the annihilation operator at site i and w_i is the free evolution of the Wannier function sitting in centred at lattice site i at position R_i [82]. For long times $w_i \approx \hbar t/(a_0 m)e^{-i\vec{Q}(\vec{r})\vec{R}_i}$ with a_0 being the width of the Wannier state on the lattice. This allows us to find a relation between the measured density and the second moments:

$$\langle n(\vec{r})\rangle = \frac{m}{\hbar t}\left\langle n_{\vec{Q}}\right\rangle = \frac{m}{\hbar t}\frac{(a_0 m)^2}{(\hbar t)^2}\sum_{i,j} e^{i(\vec{R}_i-\vec{R}_j)\vec{Q}(\vec{r})}\left\langle a_i^\dagger a_j\right\rangle \tag{6.61}$$

This looks like a kind of Fourier transform of the expectation value for the measured density. But, in this expression, $\langle n(\vec{r})\rangle$ depends only the distance between the lattice sites, $\vec{R}_i - \vec{R}_j$. The summation goes over all possible lattice sites, thus we effectively average over the second moments, where the vector between the lattice sites is the same. Trying to invert the relation leads to the condition, that

$$\left\langle a_i^\dagger a_j\right\rangle \propto \sum_{\alpha,\beta}\delta\Big((\vec{R}_\alpha - \vec{R}_\beta) - (\vec{R}_i - \vec{R}_j)\Big)\left\langle a_\alpha^\dagger a_\beta\right\rangle \tag{6.62}$$

which can be fulfilled, if $\left\langle a_i^\dagger a_j\right\rangle$ is translationally invariant. Then one could get rid of only a factor of the numbers of different realisations of this configuration in the grid. For a Mott state, this could be a good assumption, since $\left\langle a_i^\dagger a_j\right\rangle \approx \delta_{ij}$, but this is against the spirit of our method. Thus, inverting the relation is only possible for the simple case of the diagonal, where both expressions are the same. The same problem occurs for the fourth moments, where we obtain

$$\left\langle n(\vec{r})n(\vec{r}')\right\rangle = \frac{a_o^4 m^6}{(\hbar t)^6}\sum_{i,i',j,j'} e^{i(\vec{R}_i-\vec{R}_i')\vec{Q}(\vec{r})}e^{i(\vec{R}_j-\vec{R}_j')\vec{Q}(\vec{r})}\left\langle a_i^\dagger a_j^\dagger a_{j'}a_{i'}\right\rangle + \delta(\vec{r}-\vec{r}')\left\langle n(\vec{r})\right\rangle . \tag{6.63}$$

For our method, we need all expectation values of $\left\langle a_i^\dagger a_j \right\rangle$ as well as $\left\langle a_i^\dagger a_i a_j^\dagger a_j \right\rangle$ if we already fill in those values with a zero, which violate particle number conservation and use commutation relations to compute others. We need all these values, since we have to transform the measurement according to the symplectic transformation which allows us to reduce the problem to single mode problems. If we treat the lattice state as translationally invariant, it is cumbersome (because of the factor for the number of possibilities for the same distance), but possible to get the input necessary for the numerical estimation.

6.4. Outlook

In this chapter, we have introduced and elaborated on a method to directly detect local deviations from Gaussianity of quantum many-body systems. In this way, we have identified a way of witnessing the deviation from the ground or thermal state of a non-interacting model, hence directly observing strong correlations present in the state. We did so by estimating the relative entropy between a reference state and all states compatible with a certain measurement outcome. By making use of ideas of convex optimisation and Lagrange duality, we can efficiently derive reliable bounds on this relative entropy. It is desirable to be able to directly detect properties relevant for research questions at hand. This way, no prior assumptions can enter the reconstruction and could possibly change the interpretation of a measurement outcome. More conceptually speaking, the mindset that we advocated here constitutes a relevant further step to equip bounds and estimates with precise statistical confidence regions. This complements the more conventional approach in which a specific physical situation is modelled or classically simulated, and the predictions compared with those of data from measurements. It is our hope that the present work can contribute to the further development of such tools of certification for the study of many-body models, particularly relevant for the partial certification of the functioning and read out of analogue quantum simulators.

Part IV.

System Manipulation

7. Entanglement as a Resource

It has long been noted in the field of quantum information science that entanglement, a phenomenon unique to quantum mechanics, constitutes the key resource in various information processing and specifically communication tasks [10].

One of the most powerful applications of entanglement in quantum computing itself is the one-way quantum computer, which replaces unitary transformations by single-qubit measurements on a highly entangled state. The computation hereby is shifted to the initialisation of this state and explicitly uses entanglement as a resource for computation [77].

Apart from building a whole quantum computer, there is a vast amount of quantum communication protocols for many tasks such as secure quantum key distribution, which necessarily rely on entanglement, even in prepare and measure schemes [30, 49, 50]. A central goal in quantum information science has therefore been the development of techniques to transform less usable forms of entanglement into more suitable ones, and building our understanding of the laws governing the manipulation of entanglement.

7.1. Entanglement

A quantum state on a system that is described in a Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ is called entangled, whenever it is not separable, meaning that it cannot be written as a tensor product of two pure states on the subsystems $\mathcal{H}_A$ or $\mathcal{H}_B$. This implies, that the reduction of the state to either of the subsystems can only be described probabilistically. For the maximally entangled state, the reduction of the state on the subsystems will yield the maximally mixed state. This can easily be seen using the Schmidt decomposition:

$$|\xi\rangle_{AB} = \sum_{i=1}^{d} \sqrt{\lambda_i}\, |\eta_i\rangle_A \otimes |\zeta_i\rangle_B \,. \tag{7.1}$$

d is the minimum dimension of the subsystems and λ_i are the Schmidt coefficients, which satisfy $\sum_i \lambda_i = 1$. In this picture, the reduced state is given by

$$\rho_B = \mathrm{Tr}_A\, |\xi\rangle\langle\xi| = \sum_{i=1}^{d_A} \lambda_i\, |\zeta_i\rangle\langle\zeta_i| \,. \tag{7.2}$$

If the λ_i are evenly spread over the whole spectrum, we get the maximally mixed state. The amount of entanglement can be quantified more precisely. There are several measures for entanglement, that concentrate on different notions of entanglement. One of the oldest measures for entangled pure states is the Shannon entropy of the λ_i, also known as the von Neumann entropy of ρ_B.

$$S(\rho_B) = \mathrm{Tr}(\rho_B \log(\rho_B)) = \sum_i \lambda_i \log(\lambda_i) \tag{7.3}$$

For $\lambda_i = \frac{1}{2}$ the state ξ is maximally entangled. The von Neumann entropy is a convex function, meaning that mixing two states will increase the entropy. This corresponds two a loss in information.

Four special states that are maximally entangled are the so-called Bell states, which form a basis of the two-qubit Hilbert space. They form a complete orthonormal set and they can be transformed into each other by local unitary operations.

Entanglement can only be created through interactions between the subsystems. Without interaction, the Hamiltonian describing the evolution of the system would lead to a unitary evolution that can be written as a product, which will transform a separable state into another separable state.

$$H = H_A \otimes \mathbb{1}_B + \mathbb{1}_A \otimes H_B \rightarrow U(t) = U_A(t) \otimes U_B(t)$$

Operators that factorise into the product of the operators of the subsystems are called local operators. In many tasks in quantum information, it is a goal to use only those local operators as well as classical communication, as the qubits might be in different locations. Those LOCC operations can obviously not create or enhance entanglement, which is why so-called entanglement purification protocols are used to keep the amount of entanglement as high at is needed.

While we until now only considered bipartite entanglement of pure states, the whole picture is much more delicate. So-called multipartite entanglement is harder to define satisfactorily. Entanglement can for example be quantified regarding only certain partitions of the subsystems. On top of that, entanglement has to also be defined for mixed states of the quantum system, which will naturally appear in systems that interact with an environment. The entanglement measures have to be adapted accordingly.

A mixed state ρ is called separable if and only if it can be written as

$$\rho = \sum_k p_k \left|a_k, b_k\right\rangle\!\left\langle a_k, b_k\right|,$$

with $|a_k\rangle \in H_A$ and $|b_k\rangle \in H_B$. Following the former definition, it is called entangled whenever it is not separable. This definition agrees with the requirement, that entangled states cannot be generated from a product state using LOCC.

To probe whether a given state is entangled or not, the Peres-Horodecki criterion [58, 74] can be used. It provides a test using the partial transpose of a bipartite state ρ. If this partial transpose has a negative eigenvalue, the density operator is certainly entangled. This criterion is sufficient, but not necessary. For $\mathcal{H}_A$ and $\mathcal{H}_B$ of high dimension, there are density operators which are entangled without being detected by the criterion.

Another method of detecting entanglement in quantum states without having full tomographic knowledge is the use of an entanglement witnesses [38]. There, the detection of entanglement is connected to negative expectation values in the measurement of the Wigner function, which are a witness of entanglement.

7.2. Quantum Repeaters

In the last section we learnt that entangled states cannot be created without local interactions. For many quantum protocols, such as secure key distribution, we will however want to use entangled qubits, that are far away from each other. The quantum information has to be transmitted over large distances. In quantum mechanics, there are physical constraints that make this transmission completely different from the classical distribution of information. As a single measurement cannot give reliable information about an unknown quantum state, the whole state rather than its information has to be transmitted. Also, directly following from the same reasoning, a state cannot be cloned [105].

To get a good estimate of a state, a sufficient number of copies of the same state have to be created and measured to determine the coefficients in any chosen basis. The idea in quantum communication processes is thus to directly use the created state for further processing without measuring it. To achieve that, the quantum devices have to be connected through quantum channels. For photons, optical fibres provide the best candidate, where even losses can be included in the description of the quantum channel [15, 36].

To transport an intact quantum state from one place to another, quantum teleportation can be used. For that, the sender and the receiver of the state have to perform measurements on a maximally entangled pair they share. Thus, the problem of transferring quantum states can be brought to the problem of distributing maximally entangled pairs over large distances. As quantum channels such as optical fibres or other environments always introduce noise and decoherence, especially over very large distances, quantum repeaters and entanglement distillation and purification procedures come into play.

7.3. Entanglement Distillation

Entanglement distillation specifically captures the resource character of entanglement. as it aims at preparing maximally entangled states from noisy or less entangled ones [10]. Specifically, it refers to the task of transforming a collection of weakly entangled pairs into fewer highly entangled ones. There exist many different kinds of distillation protocols, even optimal ones as the hashing protocol. Sadly, those are most of the time hard to implement as they do not only require local operations that are easy to implement. The protocols we will discuss are far from optimal, but only use Clifford gates on two subsequent qubits and require only LOCC operations. This is important, if one aims for a realistic protocol to be used in a quantum repeater.

Distillation steps are part of *quantum repeater protocols* [17, 24, 34], necessary to distribute entanglement over arbitrary distances of a noisy quantum channel: In such a scheme, entanglement is being established between different repeater stations and transferred to the final designated nodes via suitable entanglement swapping steps. Distillation schemes thought of in this context are often iterative schemes, such as the recurrence protocol [8, 10, 31] and deterministic protocols based on error-correction codes [10, 56, 60, 65]. While iterative schemes do not achieve the maximum rates set by the distillable entanglement, they require less sophisticated and are more practically feasible operations.

We will explore those two kinds of protocols, the postselective one, where depending on the outcome of a measurement during one step of the protocol the outcome is either kept and more entangled or discarded, and a deterministic one, that uses error correction to arrive at a state that is more strongly entangled. The deterministic protocol only works for high initial entanglement, but has the important feature that it never fails to bring up entanglement. The postselective protocols allow for lower initial entanglement with the cost of the risk of losing the state completely. In usual realistic entanglement distillation setups a concatenation of the two protocols is used. First, the low entanglement is brought up using a postselective protocol to fit the requirements of a deterministic one. We will now introduce two standard protocols, the postselective Recurrence protocol and the 5-qubit error correction code—a deterministic protocol.

7.3.1. Recurrence Protocol

Similarly proposed by Deutsch and Bennett in 1996, the recurrence protocol is not very efficient in the speed of convergence to the maximally entangled state, but focuses on implementability. It is a $2 \to 1$ protocol, as it consumes 2 pairs in every step to arrive in one state that is closer to the maximally entangled one. By acting out this protocol in parallel, it consumes half the shared pairs in one step of the iteration. We will describe

the protocol in more detail, using the Pauli operators, the Hadamard gate, and the C_{NOT} gate, a controlled not gate. The C_{NOT} is a two-qubit gate.

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

While the Pauli operators can be interpreted as causing bit- and phase-flip operations, the C_{NOT} entangles the two qubits it is acted on, while the first qubit serves as the control bit.

$$C_{\mathrm{NOT}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad C_{\mathrm{NOT}}(|0\rangle + |1\rangle)\,|0\rangle = |00\rangle + |11\rangle$$

With this set of operations, the two communication partners A and B perform then an iterative procedure consisting of three steps. First, they both perform a C_{NOT} on two subsequent pairs. Second, they measure the target pair in the computational basis and third, compare their results. Whenever they measured the same outcome, they brought up the entanglement in the remaining state. If they however measured different results, they have to discard the remaining state, as they didn't succeed in bringing up the entanglement. It is this step of discarding pairs depending on the measurement outcome that all postselective protocols share. The recurrence protocol nevertheless is a useful protocol, as the probabilities of measuring the same outcome depend on the amount of entanglement that is present in the system. If we consider the distillation of ϕ_+ Bell states, we can formulate the protocol using probabilities. The index n denotes the nth iteration of the protocol.

$$p_n(\phi_+) = a_n, \quad p_n(\phi_-) = b_n, \quad p_n(\psi_+) = c_n, \quad p_n(\psi_-) = d_n$$

The probabilities are at the same time the overlap of the original state with the 4 orthogonal Bell states, that form a basis set. Every iteration step changes these probabilities. For one iteration step, the probabilities for bit flip errors (ψ instead of ϕ) are suppressed in the square of the strength of the noise ϵ, while phase flip ($-$ instead of $+$) is only suppressed in the order ϵ. One can use the Hadamard gate to interchanges these two errors and transform bit-flip errors to phase-flip errors and vice versa. (Explicitly, it interchanges ϕ_- with ψ_+.) The two communication partners thus both apply a Hadamard gate as a fourth step of their iterative procedure before performing another iteration step. This double step shrinks both phase-flip and bit-flip errors with at least the order ϵ^2 and for suitable initial conditions the protocol indeed converges to the maximally entangled

ϕ_+ state. The probabilities after two iterations are given by

$$\begin{aligned} a_{n+2} &= \left(a_n^2 + b_n^2\right)^2 + \left(c_n^2 + d_n^2\right)^2, \\ b_{n+2} &= 4a_n^2 b_n^2 + 4c_n^2 + d_n^2, \\ c_{n+2} &= 2\left(a_n^2 + b_n^2\right)\left(c_n^2 + d_n^2\right), \\ d_{n+2} &= 8a_n b_n c_n d_n. \end{aligned}$$

The protocol converges whenever the initial fidelity, so the size of a_n is higher then $\frac{1}{2}$, as Deutsch has shown [8, 31].

7.3.2. 5-qubit Error Correcting Code

In classical error correction, one often uses redundant encoding of the information to guarantee for successful transmission. The bit is replaced by a logical bit, that is more robust against single bit flips as multiple bit flips, occurring at the same time, would be needed to flip from the logical 0 to the logical 1. This number of flips is called the Hamming distance and often simply corresponds to the number of bits used in the encoding.

In the quantum setting, we have to obey the no cloning theorem and cannot just copy the state before the transmission. One can, however, follow a similar line of thought, and embed the qubit in a larger Hilbert space. Shor introduced a quantum error correcting code, a nine-qubit code, which is a generalisation of the ideas introduced in the Shor code [32].

Using a unitary transformation U, the state is encoded into a subspace of some large Hilbert space. This subspace is also called quantum error-correcting code. After the transmission over a noisy channel, a syndrome measurement is performed on the extra dimensions, also called ancillas. This measurement gives as output an error syndrome, which is used to determine the recovery operation to be applied to arrive at the original qubit. Each error syndrome belongs to an orthogonal subspace and corresponds to a particular error that can appear in the system. This way, the error can be recognised by the syndrome and can be reversed by the measurement.

Error correction can be used in entanglement purification to correct the errors that distinguish the state from the maximally entangled one. Different from the postselective protocol, only one way communication is needed in the protocols, as only the error syndrome has to be transmitted. Additionally, the state is kept whenever the error can be identified. The smallest error correcting code is the 5-qubit code. Every single qubit error is mapped to a unique error syndrome and it is set up solely through Clifford operations. The error syndrome could however also correspond to multi-qubit error and

the wrong error would be corrected, introducing a new additional error in the state. As long as the errors are uncorrelated and the error probability is small, the protocol still converges on average. This is why the error correcting protocol can only be used in much lower noise levels, compared to the postselective protocol. For the 5-qubit code the noise tolerance is $\epsilon \approx 0.046$.

Both protocols follow the assumption that the initial states are identically and independently distributed, which means that there is no correlation between subsequent pairs. This is an assumption that is very strong and does not agree with having sources of entangled pairs that are used at a high rate, as the source system could still be in an excited state after the measurement. Our work extends this picture and proves bounds for these protocols when the system shows not only wanted inter-pair entanglement but also unwanted correlations between subsequent pairs [18, 23].

8. MPO Purification and Renormalisation

The following work has been conducted together with Stephan Wäldchen, Earl Campbell and Jens Eisert and has been published in PRL in 2015 [100].

In entanglement distillation protocols, it is usually assumed that the initial entangled pairs are i.i.d. distributed and uncorrelated with each other, an assumption that may be very much inappropriate in any entanglement generating process involving memory channels. Here, we introduce a framework that captures entanglement distillation in the presence of natural correlations arising from memory channels. Conceptually, we bring together ideas of condensed-matter physics – that of renormalisation – with those of local entanglement manipulation. Formally, we introduce ideas of tensor networks and matrix product operators to the context of entanglement distillation, and rigorously prove convergence to maximally entangled states in several meaningful settings, introducing notions of renormalisation of matrix-product operators.

8.1. Considering Correlations

The silent assumption in almost all of the proposed distillation schemes, is that the initial resources have been identically prepared and show no correlations. Depending on the type of preparation, this may or may not be a reasonable assumption. Whenever memory effects or channels [1, 63, 76] are involved, one expects some correlations between the involved entangled pairs, going beyond an i.i.d. setting. These correlations are expected to decay rapidly over several pairs sent through a channel – reflecting the natural correlation structure arising from a *memory channel* (see Fig. 8.1). The mathematical definition of distillable entanglement in the presence of correlations has been developed in [15, 18]. The important practical problem of how to distil entanglement from correlated pairs arising from quantum memory channels is still wide open.

In this work, we propose a conceptually novel way forward to solve this problem, bringing together ideas from entanglement theory with those of condensed matter theory, specifically of renormalisation and tensor network states. We start by identifying the natural class of states arising from preparations and memory channels as bipartite *matrix-*

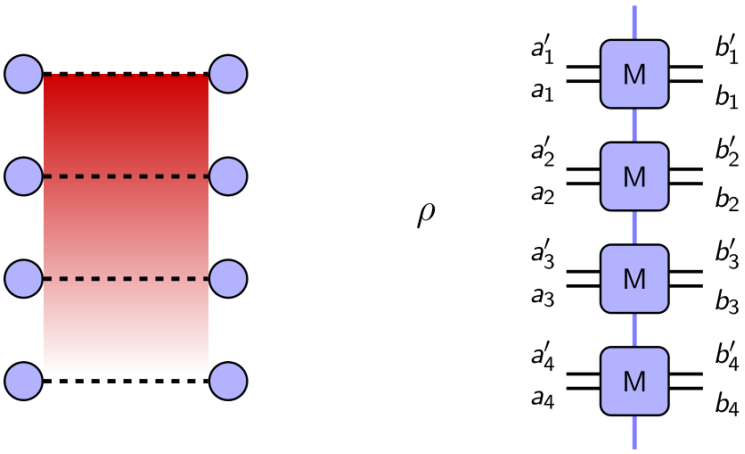

Figure 8.1.: On the left: Source with memory emitting weakly correlated photon pairs with a decaying correlation from one pair to the next pairs indicated in red. On the right: Same setup, naturally formulated as an MPO with the bond dimension indicated in blue.

product operators (MPO) [99, 107]. Such classes of states are usually considered in the condensed matter context to capture thermal many-body states or those arising from open systems dynamics. Here, we encounter natural bipartite – and still multipartite – instances thereof. Entanglement distillation is then identified as a *renormalisation of bipartite matrix-product operators*. Viewed in this mindset, the methods are inspired and derived from renormalisation [98] of *matrix-product states* [36, 40, 73, 75, 84, 86], again from many-body theory.

Specifically, what we show is that both the recurrence protocol and the error correction protocol converge to pairs of maximally entangled pure states for subsequently correlated pairs naturally described by an MPO. This leads to entanglement distillation very similar to the i.i.d. case and numerical evidence suggests that allowing for principally unwanted correlations between subsequent pairs can even speed up the convergence to maximally entangled pairs compared to the uncorrelated i.i.d. case. We introduce a simple physical example, where this is the case for a large parameter region.

8.2. Setting and Formalism

We consider a sequence of L pairs of qubits, where two parties (say Alice and Bob) each hold one qubit from each pair. The focus on qubits is set for simplicity of notation only, it is clear that the same framework can be applied also to systems of other physical dimension. These pairs are entangled, as well as correlated with each other, as a con-

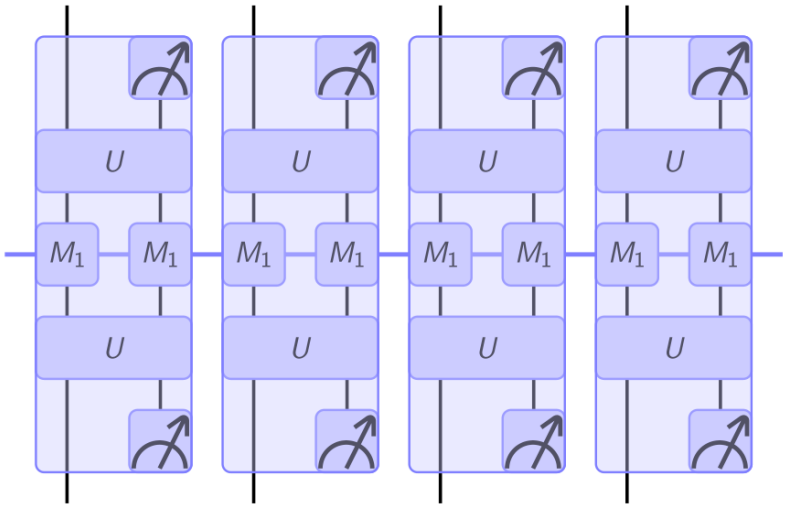

Figure 8.2.: Renormalisation schemes of MPO mapping N pairs to M in an $N \to M$ scheme. In the $2 \to 1$ recurrence protocol two neighbouring MPOs are conjugated with a local unitary and subjected to a measurement. Contraction of the tensor network leads to the MPO at the subsequent scale.

sequence of the preparation procedure involving stationary quantum memory effects. A natural preparation exhibiting such a memory involves an auxiliary quantum system C of some dimension d that embodies all the degrees of freedom of the memory. The state is then prepared in a sequential fashion, with the memory unitarily interacting with the first entangled pair, then the second, and so on [75, 76, 83]. A state generated in this way is a matrix-product state, if it is pure, or a *matrix-product operator* in case of noisy mixed states [99, 107], as they are considered here, with d taking the role of the *bond dimension*. The decay of memory effects in the distance between the entangled pairs naturally emerges in this construction. It is the setting of naturally correlated bipartite MPO arising that is being introduced here.

More specifically, we work in a numerically indexed Bell basis $\{|\phi_1\rangle, |\phi_2\rangle, |\phi_3\rangle, |\phi_4\rangle\}$, more commonly labelled as $\left\{|\phi^+\rangle, |\phi^-\rangle, |\psi^+\rangle, |\psi^-\rangle\right\}$. We consider a sequence of L pairs of qubits, with basis vectors $|\Phi_{\mathbf{x}}\rangle = |\phi_{x_1}\rangle |\phi_{x_2}\rangle \dots |\phi_{x_L}\rangle$, where Alice holds the first qubit of each pair and Bob holds its partner. Translationally invariant mixed states reflecting stationariness of the source are described in the MPO language as

$$\langle \Phi_{\mathbf{x}}|\rho|\Phi_{\mathbf{y}}\rangle = \mathrm{Tr}[M^{x_1,y_1} M^{x_2,y_2} \dots M^{x_L,y_L}]. \tag{8.1}$$

The dimension of the matrices $M^{x,y} \in \mathbb{C}^{d\times d}$, $x, y \in \{1, \dots, 4\}$ limits the correlations between pairs, and by increasing this bond dimension d arbitrary quantum states can be

described in this formalism. There is a gauge freedom in our choice of MPO matrices as for any invertible X, mapping $M^{x,y} \mapsto X M^{x,y} X^{-1}$ will give an alternative description of the same physical state. Generally, 16 matrices are needed for the description of each pair, since a two-particle density matrix has 16 entries. However, without loss of generality we can assume the state is Bell diagonal, so that $M^{x,y} = 0$ whenever $x \neq y$. If the state was not Bell diagonal originally, it can be brought into this form by a suitable local group twirl over the Pauli group [10]. Since the employed protocols make use of Clifford operations, the group twirl will conjugate with these operations, so that it can be implemented at the very end or merely at the level of classical data processing. For this reason we herein use the shorthand $A = M^{1,1}$, $B = M^{2,2}$, $C = M^{3,3}$ and $D = M^{4,4}$. Without loss of generality, we consider the distillation of maximally entangled ϕ^+ pairs. The "A" matrix will be the dominant matrix and the others we will call noise matrices. We introduce the noise contribution of the coefficient matrices B, C and D as $\epsilon_n = \max\left(\|\mathcal{B}_n\|_{1\to1}, \|\mathcal{C}_n\|_{1\to1}, \|\mathcal{D}_n\|_{1\to1}\right)$.

8.3. Protocols and Renormalisation

An $N \to M$ iterative protocol for entanglement distillation of i.i.d. states will act on N pairs at a time and output M (where $M < N$) pairs. For more than N i.i.d. pairs, the protocol is performed in parallel on blocks of N pairs. In the MPO setting, pairs are not i.i.d. and so we must specify which pairs are involved in each block of a protocol. We choose neighbouring pairs so the first N pairs are distilled into M pairs, while simultaneously the next N pairs are distilled, and so on. This natural choice has the practical merit of respecting locality, and has the additional advantage that the output state is easily shown to again be a MPO of the same bond dimension (see Fig. 8.2). Every iteration of the distillation protocol now acts as a map from a MPO on one scale to the subsequent one and reducing the chain length from L to LM/N. After each step, a positive MPO is retained [62]. Indeed, it can be naturally seen as a process of *MPO renormalisation*, this being a mixed-state and bipartite analogue of the renormalisation of matrix product states discussed in Ref. [98]. After the nth step, we label the MPO operators $\{A_n, B_n, C_n, D_n\}$, where the initial raw state provides the $n = 0$ matrices. We further introduce the transition matrix $E_n = A_n + B_n + C_n + D_n$ as it has useful properties. We prove several results on convergence to entangled states which show the functioning of the schemes; proofs that can also be interpreted as convergence proofs for renormalisation flow of the MPOs. Specifically we consider the recurrence protocol and a distillation through error correction using the 5-qubit code. Both protocols rely solely on Clifford operations that preserve the Bell diagonality, and use local operations and classical communication with respect to the partition between Alice and Bob. Intuitively,

one can say that in many practically relevant settings, the entanglement and correlations between pairs are being "renormalised" into more useful entanglement shared between Alice and Bob, to be employed in subsequent key generation.

8.4. 5-qubit Protocol

For every error correcting code that encodes k qubits into N physical qubits, there exists an iterative $N \to k$ one-way entanglement distillation protocol [10]. In these protocols, noise information is extracted by local measurements, but instead of post-selecting when errors are detected we attempt to correct them by determining the smallest weight error consistent with the measurement data. The advantage over the recurrence protocol is that this protocol is deterministic, and that one-way distillation schemes require much less classical communication. In particular, we consider one-way entanglement distillation using the 5-qubit code (so a $5 \to 1$ protocol), which is the smallest code capable of correcting any single qubit error, and we state the following theorem

Theorem 3 (Error correction)**.** *Given a translationally invariant Bell diagonal MPO ρ_0 with coefficients A_0, B_0, C_0, and D_0, the iterative application of the 5-qubit error correcting code leads to convergence of ρ_n to uncorrelated pairs in the pure and maximally entangled state ϕ_+ for $\epsilon_0 \leq 1/33$. ϵ_0 is defined as above in the gauge where $\mathcal{E}_0$, the quantum channel that is isomorphic to E_0, is trace-preserving.*

Again, the full proof will be appear in the appendix. Similar to the post-selective case, we show that with a growing number of iterations n, the contribution of the dominant matrix A_n to the state ρ_n grows exponentially faster than the contribution of the noise matrices. In the deterministic case, we can use our gauge freedom to make the map corresponding to the transition matrix E trace-preserving. Since we do not post-select, we do not need to renormalise ρ_n in every round. In the 5-qubit error correction code, the transition matrix is always mapped to its fifth power, $E_{n+1} = E_n^5$. Thus, if we initially choose a gauge where the corresponding map $\mathcal{E}_0$ is trace-preserving, the transition matrix will keep this property over the course of the iteration. Using combinatorical arguments, we then prove that for suitably small ϵ_0, ϵ_n converges to zero, entailing convergence in fidelity of the physical state.

Instead of giving a lower bound on the fidelity, we give an upper bound on the infidelity, or the probability of measuring ϕ_2, ϕ_3 or ϕ_4 for a pair of qubits. For ϕ_2 we have

$$\mathrm{Tr}[\rho(|\phi_2\rangle\langle\phi_2| \otimes \mathbb{1})] = \frac{\mathrm{Tr}\left[BE^{L-r}\right]}{\mathrm{Tr}[E^L]}, \tag{8.2}$$

where the $\mathbb{1}$ acts on all other pairs of qubits. Similar expressions hold for ϕ_3 and ϕ_4 by replacing B with C or D, respectively. We will find an upper bound of this expression in

terms of the channel norm of the associated observable transfer matrices B, C, and D. We assume locally purifiable MPOs which allows us to make use of an isomorphism between the MPO matrices and completely positive maps. Specifically, we define a norm $\|M\|$ in terms of a channel $\mathcal{M}$ Choi-isomorphic to M, so that $\|M\| = \|\mathcal{M}\|_{1\to 1}$ and we use the induced "1-to-1" Schatten norm. See Ref. [11] and appendix A.2.1 for more on norms. We call $\mathcal{E}$ the channel Choi-isomorphic to the transfer matrix E. Our lemma 12 of the appendices proves that, assuming $\mathcal{E}$ is trace-preserving, we have

$$\mathrm{Tr}[\rho\left(|\phi_2\rangle\langle\phi_2| \otimes \mathbb{1}\right)] \leq \|B\|_{1\to 1} \frac{1 + d^{5/2}\tau(\mathcal{E})^{L-r}}{1 - d^{5/2}\tau(\mathcal{E})^{L}}, \tag{8.3}$$

where τ is the ergodicity coefficient of a channel, defined as

$$\tau(\mathcal{M}) = \max_{\mathrm{Tr}[\sigma]=0} \frac{\|\mathcal{M}(\sigma)\|_1}{\|\sigma\|_1}, \tag{8.4}$$

which allows quantification of how rapidly a channel mixes input states into the channel's stationary state. Proving convergence of the initial MPO in the state ρ to the maximally entangled state ϕ^+ is achieved by showing that the noise matrices vanish exponentially faster than the A_n matrices.

One step of the protocol takes 5 pairs as input and returns one pair as output, correcting all zero and one qubit errors and thus reducing the error probability to at least quadratic order in ϵ_n. As discussed before, the code applies without postselection, which means that the state does not have to be renormalised. As the length of the chain L_n is divided by five in every step, the new transfer matrix E_{n+1} is simply the fifth power of the previous transfer matrix E_n. So if we start with a trace-preserving transfer matrix, it remains trace-preserving

$$E_{n+1} = E_n^5 \Longrightarrow E_n = E_0^{5^n} \tag{8.5}$$

and

$$L_{n+1} = \frac{L_n}{5} \Longrightarrow L_n = \frac{L_0}{5^n}.$$

By submultiplicativity of the ergodicity, we have

$$\tau(\mathcal{E}_n)_n^L \leq \tau(\mathcal{E}_0)^{5^n \frac{L_0}{5^n}} = \tau(\mathcal{E}_0)^{L_0}. \tag{8.6}$$

This already gives us an upper bound for the normalisation for every step in the iteration just from the initial length of the chain and ergodicity. The second step is to upper bound $\|\mathcal{B}_n\|_1$, $\|\mathcal{C}_n\|_1$ and $\|\mathcal{D}_n\|_1$. We introduced, as a measure of the noise, the quantity

$$\epsilon_n = \max\{\|B_n\|_1, \|C_n\|_1, \|D_n\|_1\}.$$

defined in a gauge and scaling, where $\mathcal{E}_n$ is trace-preserving. This gauge and scaling stays constant over the iteration. The complete update rules can be derived following

the methodology of section A.1.1. We obtain a closed form for the map acting on the coefficient matrices in each iteration, though we omit it here as each expression contains 4^4 terms and is not very insightful. Rather, we present just the four leading terms here

$$\begin{aligned}
A_{n+2} &= A_n^5 + A_n^4 B_n + A_n^4 C_n + A_n^4 D_n + \dots, \\
B_{n+2} &= A_n^3 B_n^2 + A_n^3 C_n D_n + A_n^3 D_n C_n + A_n^2 B_n A_n C_n + \dots, \\
C_{n+2} &= A_n^3 B_n C_n + A_n^3 C_n B_n + A_n^3 D_n^2 + A_n^2 B_n A_n D_n + \dots, \\
D_{n+2} &= A_n^3 B_n D_n + A_n^3 C_n^2 + A_n^3 D_n B_n + A_n^2 B_n A_n B_n + \dots .
\end{aligned}$$

Furthermore, we do not need the explicit iteration rule but only its norm. Each term is a product consisting of A_n and some noise matrices. We can upper bound the channel norm by using submultiplicativity and subadditivity and the definition of our noise measure.

$$\begin{aligned}
\|B_{n+1}\|_{1\to 1} &\leq 30\epsilon_n^2 \|A_n\|_{1\to 1}^3 + 70\epsilon_n^3 \|A_n\|_{1\to 1}^2 + 90\epsilon_n^4 \|A_n\|_{1\to 1}^2 + 66\epsilon_n^5, \\
\|C_{n+1}\|_{1\to 1} &\leq 30\epsilon_n^2 \|A_n\|_{1\to 1}^3 + 70\epsilon_n^3 \|A_n\|_{1\to 1}^2 + 90\epsilon_n^4 \|A_n\|_{1\to 1}^2 + 66\epsilon_n^5, \\
\|D_{n+1}\|_{1\to 1} &\leq 30\epsilon_n^2 \|A_n\|_{1\to 1}^3 + 70\epsilon_n^3 \|A_n\|_{1\to 1}^2 + 90\epsilon_n^4 \|A_n\|_{1\to 1}^2 + 66\epsilon_n^5.
\end{aligned}$$

As we see the resulting iteration rule is symmetric for all three noise terms and at least of quadratic order. The norm of A_n can be easily upper bounded using the properties of channels and channel norms $\|\mathcal{A}_n\|_{1\to 1} \leq \|\mathcal{E}_n\|_{1\to 1} = 1$. The ensuing iteration for the noise measure is thus

$$\epsilon_{n+1} \leq 30\epsilon_n^2 + 70\epsilon_n^3 + 90\epsilon_n^4 + 66\epsilon_n^5. \tag{8.7}$$

Clearly, for errors to reduce ϵ_n must be at least smaller than $1/30$. In this regime we have

$$\begin{aligned}
\epsilon_{n+1} &\leq 30\epsilon_n^2 + \frac{7}{3}\epsilon_n^2 + \frac{1}{10}\epsilon_n^2 + \frac{11}{4500}\epsilon_n^2 \\
&= \left(30 + \frac{7}{3} + \frac{1}{10} + \frac{11}{4500}\right)\epsilon_n^2 \leq 33\epsilon_n^2
\end{aligned}$$

So we can be sure the iteration converges if $\epsilon_0 \leq \frac{1}{33} \approx 0.0303$. Numerically we find that $\epsilon_0 \leq 0.031$. The speed of the convergence is doubly exponential in the number of rounds,

$$\epsilon_n \leq \frac{1}{33}(33\epsilon_0)^{2^n}.$$

Making use of Eq. (8.3) and Eq. (8.6), we find that infidelity is bounded by

$$p(\phi^-) \leq \frac{1}{33}(33\epsilon_0)^{2^n} \frac{1 + d^{5/2}\tau(\mathcal{E}_0)^{L_0 - 1}}{1 - d^{5/2}\tau(\mathcal{E}_0)^{L_0}}.$$

Since $|\tau(\mathcal{E}_0)| \leq 1$, we have in generic cases that in the limit of an infinite chain

$$p(\phi^-) \leq \frac{1}{33}(33\epsilon_0)^{2^n},$$

and it converges whenever $\|B_0\|_{1\to1}$, $\|C_0\|_{1\to1}$, $\|D_0\|_{1\to1} \leq \frac{1}{33}$.

It was shown that we can formulate a threshold depending on the 1-to-1-norm of the three noise channels. This concludes our proof of a threshold for the five-qubit error correcting code. This is the most important part of this work, since a real world repeater scheme would most likely employ a deterministic protocol that requires only one-way communication. However, finding a threshold for the postselective protocol is also of interest. It turns out that the proof is more sophisticated since we cannot rely on the trace-preserving property of the transfer channel. This requires a refined approach that involves deriving a perturbation bound for the left Perron vector of a quantum channel.

8.5. Recurrence Protocol

The recurrence protocol is a $2 \to 1$ iterative protocol, which uses post-selection. At every round measurement outcomes are being produced and we only proceed if certain outcomes are obtained. Here we use the recurrence scheme proposed in Ref. [31]. Cast into the MPO language, the iteration formula is found (see appendix A.1.1) to be

$$\begin{aligned} A'_{n+1} &= A_n^2 + B_n^2, \qquad & C'_{n+1} &= A_nB_n + B_nA_n \\ B'_{n+1} &= C_n^2 + D_n^2, \qquad & D'_{n+1} &= C_nD_n + D_nC_n. \end{aligned} \tag{8.8}$$

where the role of the prime is explained shortly. Replacing matrices by commutative scalars recovers the original i.i.d. result. These update rules are considerably simpler than those of the error correcting protocol, but in every other regard analysis of the recurrence protocol is more involved.

We introduce the noise contribution of the coefficient matrices B', C' and D' as $\epsilon'_n = \max\left(\|\mathcal{B}'_n\|_{1\to1}, \|\mathcal{C}'_n\|_{1\to1}, \|\mathcal{D}'_n\|_{1\to1}\right)$. Due to norm submultiplicativity, we expect that the norm of small noise matrices will shrink in size. However, ensuring A'_{n+1} stays large is difficult. To do so, we shall adjust the MPO gauge in every step

$$\begin{aligned} A_n &\to \alpha_n X_n A'_n X_n^{-1}, \qquad & B_n &\to \alpha_n X_n B'_n X_n^{-1} \\ C_n &\to \alpha_n X_n C'_n X_n^{-1}, \qquad & D_n &\to \alpha_n X_n D'_n X_n^{-1} \end{aligned} \tag{8.9}$$

where α_n is a constant, so that A_n trace-preserving (we also choose this gauge initially) and hence $\|A_{n+1}\|_{1\to1} = 1$. The matrix norms are not gauge invariant and we denote ϵ_n the noise parameter in this gauge. We will find that the strength of the gauge change depends on the ergodicity of $\mathcal{A}_n$, for which we use the shorthand $\tau_n := \tau(\mathcal{A}_n)$. We are now ready to state our main result.

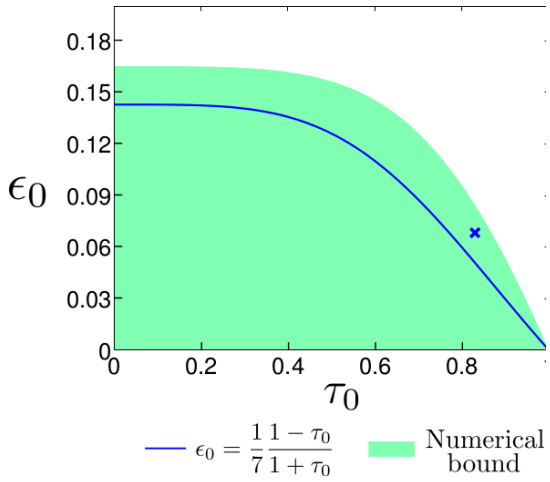

Figure 8.3.: Region of convergence for the recurrence protocol. The area under the blue line is the region fulfilling the conditions given in theorem 4. The green region is a slightly improved bound that can be obtained with computer assistance, but for which we have no closed form expression. The blue dot corresponds to our physical example discussed in section 8.6.

Theorem 4 (Convergence in the recurrence protocol)**.** *Given a translationally invariant Bell diagonal MPO ρ_0 with coefficients A_0, B_0, C_0, and D_0, the iterative application of the recurrence protocol leads to convergence of ρ_n to uncorrelated pairs in the pure and maximally entangled state ϕ_+ for $\epsilon_0 \leq \frac{1}{7}\frac{1-\tau_0^4}{1+\tau_0^4}$.*

The conditions for convergence are illustrated in Fig. 8.3. The proof is technically involved and is presented in full length in the appendix. We will sketch it here, however, and provide significant intuition. To show convergence, we need to show that the noise matrices go down exponentially fast, while A_n stays large. The first part can be shown by taking into account a double step of the protocol after which all norms of the noise matrices are at least of order ϵ_n^2. This can be shown via subadditivity and submultiplicativity of the 1-to-1 norm.

$$\epsilon'_{n+2} \leq 4\left(1+\epsilon_n^2\right)\epsilon_n^2 \tag{8.10}$$

However, to ensure the contribution of the dominant matrix stays large, our approach is to regauge so that A_{n+2} is trace-preserving. A channel $\mathcal{A}_n$ is trace-preserving if and only if the dual channel $\mathcal{A}^\dagger$ (the Heisenberg representation of $\mathcal{A}_n$) satisfies $\mathcal{A}_n^\dagger(\mathbb{1}) = \mathbb{1}$. When instead $\mathcal{A}_n^\dagger(\xi_n) = \lambda_n \xi_n$ (where λ_n is the largest such eigenvalue), then the Perron-Frobenius theorem ensures ξ_n is Hermitian and positive semi-definite. The gauge transformation will recover trace-preservingness by setting $\alpha_n = \lambda_n^{-1}$ and $X_n = \sqrt{\xi} \otimes \sqrt{\xi}$, provided ξ is invertible. This transformation potentially increases the norm of the

noise matrices. Using submultiplicativity we find that $\epsilon_{n+2} = \lambda_n^{-1}\kappa(\mathcal{X}_n)\epsilon'_{n+2}$, where $\kappa(\mathcal{X}_n) = \|\mathcal{X}_n\|_{1\to 1}\left\|\mathcal{X}_n^{-1}\right\|_{1\to 1}$ is the condition number of $\mathcal{X}_n$. The factor λ_n^{-1} can be dismissed as one easily finds $\lambda_n^{-1} \leq 1$. We wish to show that $\mathcal{A}_{n+1}$ stays close to trace-preserving to keep the condition number small, and we demonstrate

$$\kappa(\mathcal{X}_n) \leq (1-2k_n)^{-1} \quad \text{with} \quad k_n = \frac{1+\tau_n^4}{1-\tau_n^4}\left(4\epsilon_n^2 + 10\epsilon_n^4\right). \tag{8.11}$$

The proof bears similarities to the perturbation of the steady state of a trace-preserving quantum Markov chain, which also depends on the ergodicity coefficient of the transition matrix. The basic intuition is that if $\mathcal{A}_n$ is a rapidly mixing channel, with small τ_n, then $\mathcal{A}_n^4$ is also rapidly mixing. Before we apply our gauge transformation, $\mathcal{A}_{n+1}$ is a sum of $\mathcal{A}_n^4$ and some small noise matrices. The more rapidly mixing a channel, the more its eigenstates are robust against the perturbative addition of noise matrices. We desire the dual eigenstate state ξ_n to stay close to $\mathbb{1}$, which we expect for rapid mixing channels (small τ_n) and low noise (small ϵ_n) as we show rigorously in a spirit similar to the ideas of [92].

Although ergodicity is not a matrix norm, it has similar properties such as subadditivity and submultiplicativity, from which one can derive an upper bound on τ_{n+2} in terms of τ_n and ϵ_n. A double step of the protocol raises the ergodicity coefficient of A to the fourth power, adds the perturbation and sets the new gauge. Since the perturbation is small and exponentially decaying, we can bound $\tau_{n+2} \lessapprox \tau_n^4 f(\epsilon_n)$ where $f(\epsilon_n) \to 1$ as $\epsilon_n \to 0$. The full and exact bound is given in the appendices as lemma 3. The essential point is that we have bounds on the pair $(\epsilon_{n+2}, \tau_{n+2})$ in terms of (ϵ_n, τ_n). It is now straightforward to numerically determine the region of convergence to the fixed point $(0,0)$, which we show in Fig. (8.3). Also shown in this figure is an analytic curve, for which we show $(\epsilon_n, \tau_n) \to (0,0)$ without numerical aid (see appendix A.2.6). Finally, convergence in MPO operators again entails convergence of the density matrix. This whole argument is presented in more rigorously in appendix A.1.1.

8.6. Numerical Studies and Physical Hamiltonians

To complement the rigorous and analytical results presented above, we have also performed numerical studies on randomly drawn matrices. These results confirm that a state within the distillable region is guaranteed to converge, but we also observe convergence for many states outside the analytically demonstrated region of distillability. Our analysis has assumed the worst case scenario, where correlations are always pernicious. However, our numerics indicate that in many strongly correlated chains the correlations can also be beneficial and enable distillation at noise levels well above the rigorous

threshold as well as in cases where the protocols do not converge for uncorrelated i.i.d. states.

To present a specific example we introduce a simple memory based on a Heisenberg interaction. Initially we consider uncorrelated states (Werner State with F_0)

$$\rho_0 = F_0 \left|\Phi^+\right\rangle\!\left\langle\Phi^+\right| + \frac{1-F_0}{3}\Big(\left|\Phi^-\right\rangle\!\left\langle\Phi^-\right| + \left|\Psi^+\right\rangle\!\left\langle\Psi^+\right| + \left|\Psi^-\right\rangle\!\left\langle\Psi^-\right|\Big),$$

which subsequently undergo a unitary interaction with a memory bit on Bob's side.

$$U(t,J) = \exp(itH) \quad \text{with} \quad H = J(X\otimes X + Y\otimes Y + Z\otimes Z) + Z\otimes\mathbb{1} + \mathbb{1}\otimes Z.$$

We further implement a dephasing channel for forgetful memory, which is applied to the memory in-between two interactions with Bob's qubit,

$$D_c(\sigma) = (1-c_D)\sigma + c_D\mathbb{1}.$$

This procedure is depicted in Fig. 8.4. We compare the performance of the recurrence

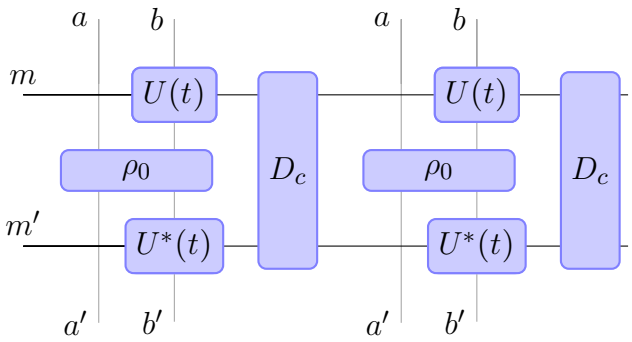

Figure 8.4.: MPO diagram of the process generating our physical example.

protocol on (i) sequentially prepared states with implemented memory with (ii) perfect memoryless i.i.d. distributions of the same local fidelity. As a measure of how a specific memory setting performs we introduce the notion of relative noise

$$\gamma_n = \frac{1-F_n^{\text{MPO}}}{1-F_n^{\text{i.i.d.}}}.$$

after n rounds of iteration.

For certain parameters, e.g. $F_0 = 0.9, J = 1, c_D = 0.04$ and $t = 0.1$, the MPO setting converges significantly faster than the i.i.d. setting of the same local fidelity. After one round we have $\gamma_1 = 0.9$, meaning that after one round we have only 90 % of the noise compared to the i.i.d. case. We included this specific MPO in Fig. 8.3. For longer

interactions ($t = 0.47$) we get a local fidelity of ≤ 0.4. An i.i.d. setting with this fidelity does not succeed, but the MPO setting does. This shows that we transport the unwanted inter pair correlations introduced by the memory into the wanted correlations between the pairs. We tested our Heisenberg memory model for different parameters and the distillation of the correlated states performs better than the distillation in the i.i.d. case for a large range of interaction times, see also Fig. 8.5.

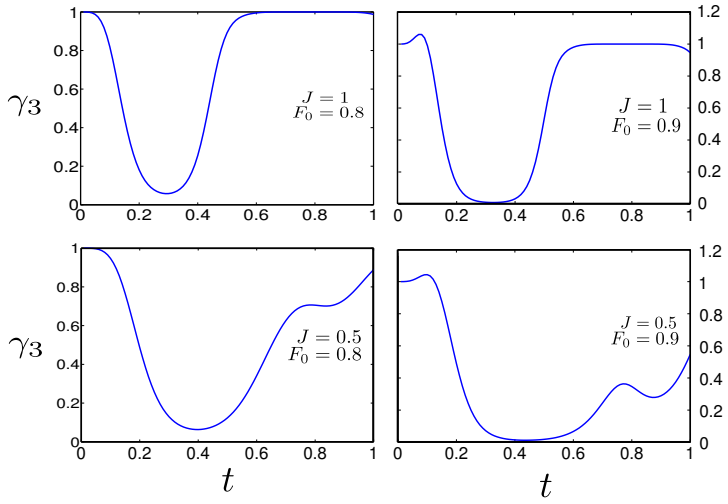

Figure 8.5.: Relative noise after three steps of iteration for different initial fidelities F_0 and interaction parameter J with $c_D = 0.04$. Whenever the relative noise is smaller than one, we see a comparable advantage over the i.i.d. case.

8.7. Perspectives.

In this work, we have introduced a framework of renormalising entanglement in order to achieve entanglement distillation in the presence of natural correlations. We have proven that protocols known to work for i.i.d. pairs above a threshold fidelity also give rise to feasible entanglement distillation. We have identified criteria to ensure convergence of correlated pairs described by a MPO to a number of independent maximally entangled pure states. On intuitive grounds, one might expect that if the MPO is only weakly correlated between the pairs and the reduced density matrix of a single pair is sufficiently close to a Bell pair, the distillation protocols should behave similarly to the i.i.d. case. Indeed, convergence can be proven for threshold fidelities and conditions on the correlation between the pairs. The programme initiated here shows that correlations are not necessarily a disadvantage, and one does not have to aim at decorrelating pairs or resetting preparation procedures, steps that will take time and will in practice lead to further entanglement deterioration. This work shows that such correlations can be

largely renormalised away, with no modification to the schemes applied. We hope that this work triggers further studies on entanglement distillation and repeater protocols in the presence of realistic memory effects.

Part V.

Outlook

9. Perspectives and Open Questions

This work aims to advertise an information theoretic mindset in the approach to quantum mechanics. A clear description of the wanted information contained in the system in question not only allows to investigate the transformations and the losses this particular information undergoes, but also helps determining which part of the wanted information is theoretically accessible for different kinds of measurement. As system sizes grow and new technologies like the internet of things introduce the need for faster and faster processing, principles from quantum information theory can offer important leads to new ideas on system identification and transformation.

In this work, we develop some ideas following this mindset. We offer new perspectives on measurement techniques as well as the interpretation of results, we introduce meaningful measures to gain insight on special properties of quantum states from only partial measurement data and we transform the entanglement content in a highly correlated state to be able to use this resource in the most sustainable way.

In chapter five, we used methods related to the compressed sensing technique to perform non-linear deconvolution of measurement data. This was done to invert the influence of the line spread function to recover the source function and with it the amplitudes and positions of atoms in an optical lattice. We applied several reconstruction methods to experimental data from measurements performed in Bonn in the group of Prof. Meschede as well as simulated data using the original line spread function. Our goal was to gain reliable knowledge of the atom positions by introducing honest error bars on the reconstruction. Furthermore, we wanted to gain insight on possible future improvements of such measurement systems to boost the distinguishability of close-by atoms by reducing information loss during the measurement process.

We present the limits and the merits of using a compressed sensing approach in this slightly non-linear measurement process. On one hand, we are able to deduce reliable error bars and benchmark our new reconstruction techniques using the positions reconstructed by the experimental group in their two stage reconstruction algorithm. We show great accordance of both reconstruction techniques, which makes the non-linear compressed sensing technique a reasonable tool to apply even in the traditional measurement setup, whose map is very coherent. On the other hand, we experiment with different slight adjustments to the measurement setup using the simulation that mimics

the measurement setup as closely as possible. We present that without introducing further incoherence, the compressed sensing techniques do not offer an improvement on the reconstructability without further reducing the noise. For reduced noise, however, we present the idea of a moving camera, where the detector plane is moved in a controlled way while the picture is taken. This allows to achieve reliable subpixel resolution in the reconstruction, where the location of an atom can be located within the detector induced pixel. Other ideas involve a shift of the measurement plane to the optical Fourier plane of the microscope. This way, the atomic norm reconstruction methods closely related to the compressed sensing method could be applied. However, the current linespread function is not suited to this approach, as it's Fourier transform is also close to a Gaussian distribution and does not cover much of the spectrum. This results in high information losses. To reduce these losses, another approach seems promising. Rather than reducing the noise in the experimental setup, which is a very time-consuming as well as costly procedure, one could introduce incoherence in the measurement map, thereby making it more suitable to the compressed sensing approach. Similar to the Fourier approach, the incoherence of the measurement map makes it harder to interpret for the human brain, but distributes the information over the whole image. But as long as two neighbouring atoms are mapped to very distinct pictures, a computer aided reconstruction would result in a significant improvement of the resolution of a reconstruction. An interesting way of introducing incoherence in measurement maps is to place randomly selected diffraction filters in the optical setup before the signal is disturbed by any linespread function. However complicated this might seem to implement in the lab, it might be worth a thought as the possibility to reliably distinguish states of neighbouring atoms in optical lattices is needed the development of optical quantum technology like repeaters or simulators.

The most intriguing extension of the reconstruction techniques points in the direction of wireless quantum technology. Rather than using the measurement map to reconstruct a signal, the reconstruction of the measurement map is included in the so-called blind deconvolution approach. The measurement map in wireless transmission is changed by obstacles and the individual path of each signal. To gain the high precision reconstruction needed in communication protocols, the transmission map must be known with high confidence. Blind deconvolution techniques could serve as a catalyst to unravel the information hidden in noisy pictures. The fuel for both techniques is the amount of incoherence in the measurement map. Using seemingly random, but known prefactors, protects the information from the influence of noise. This way, the information can be unveiled by inverting the influence of the measurement process.

If this measurement process cannot be inverted completely, one has to rely on quantities that are immune to the loss of information introduced by the measurement. In chapter six,

we introduced such a quantity—the local deviation from Gaussianity—in the context of time-of-flight measurements in optical lattices with cold atoms. This kind of measurement results in an average over the single particle distances, and thus the individual distances between particles cannot be reconstructed without further assumptions. We introduce a measure for Gaussianity and derive bounds that can be estimated using only partial data without introducing assumptions. With this, we can directly detect local deviation from Gaussianity in quantum many-body systems. This deviation can witness that a state is neither a ground nor thermal state of a non-interacting model, which means that strong correlations are present in the state. The bound is efficiently derived by using convex optimisation of the Lagrange dual function and optimises over all possible states compatible with a certain measurement outcome. This way, we avoid introducing information through prior assumptions and possibly altering the interpretation of a measurement outcome. This mindset is a relevant further step to introduce precise statistical confidence regions in derived bounds or estimates. Different from the more conventional approach of comparing the outcomes of a simulation of a specific physical situation with the actual data from real experiments, our approach allows to certify to what extent a specific interpretation of a measurement outcome is correct. Further tools for this kind of certification will be needed on the road to secure quantum communication, quantum simulators, but also in general for partial certification of measurement processes in quantum physics.

As quantum communication develops, a certain device plays an important and fascinating role: The quantum repeater. As in normal data transfer, the quality of the signal drops with the distance of transmission. In classical communication, this loss is removed by using error correction techniques and most of all by sending multiple copies of the same state through the transmission channel. In quantum communication, one generally uses shared entangled states to generate reliably secret keys, that can then be used to encrypt the message. To prove security, these key distribution protocols rely on a high amount of entanglement between the pairs. Due to incoherence effects, this would restrict communication to a very limited range. It is at this point, that the quantum repeater is an essential tool to allow for long-range quantum communication. Due to the no-cloning theorem, the repeater cannot just clone the state. In order to retrieve the original amount of entanglement, so-called entanglement distillation protocols are employed, that map many weakly entangled pairs to less strongly entangled pairs in an iterative procedure. As entanglement cannot be produced in local operations and classical communication, the efficiency of these protocols relies on carefully using the correlations already present in the system to bring up the entanglement in the preferred subspace. In chapter seven and eight, we develop a way of using unwanted inter-pair correlations introduced by noisy sources to promote the wanted correlations between the

two entangled partners. We do so by making use of the language of Matrix Product Operators and translate protocols from quantum optics to the language of tensor networks. Entangled pairs from noisy sources can naturally be described in the MPO setting and we investigate the steps of entanglement distillation protocols using renormalisation techniques. As we show, in certain physical models for noisy sources, the approach can lead to a serious speed-up in the convergence to highly entangled pairs without correlations between subsequent pairs (so-called i.i.d. pairs). We identify the parameter spaces and threshold fidelities for the convergence in two different protocols, the post-selective Recurrence protocol as well as the deterministic 5-qubit error correcting code, showing the interesting feature that the deterministic protocol is seemingly immune to a high amount of correlations. In the Recurrence protocol, the amount as well as the nature of the correlations drastically change the convergence behaviour. Indeed, the correlations are not necessarily a disadvantage. This allows for faster procedures, as time-consuming procedures to decorrelate pairs or reset preparation procedures can be omitted without modifications to the entanglement distillation schemes applied.
It is this last part in particular, that contributes to the vision of realising an unconditionally secure communication network. In quantum communication, information is transmitted encoded in quantum systems. So-called prepare-and-measure schemes, such as the BB84 protocol [7], a key distribution protocol, rely on sending perfectly distinguishable orthogonal states. Modern security proofs even allow for noisy communication channels [52, 79, 89] and provide unconditionally secure ways of key distribution. One future task is to concentrate on another tool from classical information theory—the channel capacity, or the classical information capacity. To introduce quantum communication into secure communication networks as needed in the internet of things, the capacities of quantum communication channels have to continue to grow to offer key rates that are large enough to enter the race of modern communication.

A. MPO Renormalisation

A.1. Protocols

A.1.1. Recurrence Protocol

Computational vs. Bell Basis

The recurrence protocol can be broken up into two steps performed locally by Alice and also Bob. So Alice (and also Bob) divide their qubits up into adjacent pairs within the MPO chain. For each pair the isometry $K = |0\rangle\langle 0,0| + |1\rangle\langle 1,1|$ is applied. This can be implemented using a local CNOT, measurement, postselection and disposing of the measured qubit. The second step consists of a Hadamard rotation, again by both Alice and Bob. It is easy to verify that after the first step we again get an MPO with the same bond dimension (see Fig. 8.1) such that

$$M_C^{x,y} \to (M_B^{x,y})^2. \tag{A.1}$$

where the C subscript denotes that these are matrix product operators for the computational basis of 2 qubits, and so $x, y \in \{(0,0),(1,0),(0,1),(1,1)\}$. Indeed, this is direct analogous to the i.i.d case where the density matrix elements map as $\rho_{x,y} \to \rho_{x,y}^2$, though of course only in the computational basis. The phase noise is dealt with by the second step. The bilateral Hadamard operation effectively swaps bit and phase flip noise so that both are dealt with. In the computational basis the bilateral Hadamard operation is unwieldy so we switch to the Bell basis. For the Bell basis we use the following shorthand:

$$\begin{aligned}
A = M^{1,1} &= M_C^{(0,0),(0,0)} + M_C^{(1,1),(0,0)} + M_C^{(0,0),(1,1)} + M_C^{(0,0),(1,1)};\\
B = M^{2,2} &= M_C^{(0,0),(0,0)} - M_C^{(1,1),(0,0)} - M_C^{(0,0),(1,1)} + M_C^{(0,0),(1,1)};\\
C = M^{3,3} &= M_C^{(0,1),(0,1)} + M_C^{(0,1),(1,0)} + M_C^{(1,0),(0,1)} + M_C^{(1,0),(1,0)};\\
D = M^{4,4} &= M_C^{(0,1),(0,1)} - M_C^{(0,1),(1,0)} - M_C^{(1,0),(0,1)} + M_C^{(1,0),(1,0)};
\end{aligned}$$

which only defines the MPO operators for the diagonal elements in the Bell-basis, but we will see that they are decoupled from other elements and so it is sufficient to consider

these alone. In this basis the recurrence procedure implements

$$\begin{aligned}
A_n &\to A_n^2 + B_n^2 \to A_n^2 + B_n^2 = A_{n+1}, \\
B_n &\to \{A_n, B_n\} \to C_n^2 + D_n^2 = B_{n+1}, \\
C_n &\to C_n^2 + D_n^2 \to \{A_n, B_n\} = C_{n+1}, \\
D_n &\to \{C_n, D_n\} \to \{C_n, D_n\} = D_{n+1},
\end{aligned}$$

where each of the two steps are shown, and a subscript n is introduced to denote the MPO after n iterations. The brackets $\{\cdot,\cdot\}$ denote the anti-commutator. For matrices that are simply scalars, where $B_n, C_n, D_n \sim \epsilon_n$ and $A_n \sim 1$, we see $B_{n+1}, D_{n+1} \sim 2\epsilon^2$ but $C_{n+1} \sim \mathcal{O}(\epsilon)$. This occurs because in a single round only one type of noise is decreased, and so to see an overall ϵ^2 error reduction we must consider two rounds of iteration

$$\begin{aligned}
A_{n+2} &= (A_n^2 + B_n^2)^2 + (C_n^2 + D_n^2)^2, \\
B_{n+2} &= \{A_n, B_n\}^2 + \{C_n, D_n\}^2, \\
C_{n+2} &= \{A_n^2 + B_n^2, C_n^2 + D_n^2\}, \\
D_{n+2} &= \{\{A_n, B_n\}, \{C_n, D_n\}\}.
\end{aligned} \tag{A.2}$$

Now treating the matrices as scalars, we see B, C, D all go from size $\mathcal{O}(\epsilon)$ to $\mathcal{O}\left(\epsilon^2\right)$ or smaller. This is the intuition from the i.i.d. case and we next turn to making this rigorous by quantifying this size with appropriate matrix norms.

Detailed Outline of Convergence Proof for Recurrence Protocol

In the main text we presented a sketch of the proof of Thm. 4. We first recap this sketch, filling in some details, and clearly stating required lemmas. Numerous technical tools relating to norms and the ergodicity coefficient are covered in Sec. A.2.1. Many readers may find it useful to first familiarise themselves with the simpler proof for deterministic protocols. The first step is showing the iterative formulae relating MPO operators after $n+1$ distillation rounds as a polynomial of MPO operators after n rounds, as introduced in Eq. A.2. Much of the proof centres around obtaining an upper bound on

$$\epsilon_n = \max\left(\|\mathcal{B}_n\|_{1\to1}, \|\mathcal{C}_n\|_{1\to1}, \|\mathcal{D}_n\|_{1\to1}\right), \tag{A.3}$$

where a gauge is used so that $\mathcal{A}_n$ is trace-preserving. As argued in the main text we make use of the following lemma.

Lemma 1. *Let $\mathcal{M}$ be a completely-positive channel, with largest eigenvalue λ and $\mathcal{M}^\dagger(\xi) = \lambda\xi$. Let $\mathcal{X}$ be a channel mapping $\mathcal{X}(\rho) = \sqrt{\xi}\rho\sqrt{\xi}$. If ξ is invertible, $\mathcal{X}^{-1}$ exists and $\mathcal{X} \circ \mathcal{M} \circ \mathcal{X}^{-1}$ is trace-preserving.*

We give further steps of the proof in Sec. A.2.2. Applying the lemma to $\mathcal{A}'_{n+2}$, gives a gauge transform we herein label $\mathcal{X}_n$. As argued in the main text, we have a recursive relation

$$\epsilon_{n+2} \leq 4\kappa(\mathcal{X}_n)\Big(1+\epsilon_n^2\Big)\epsilon_n^2, \tag{A.4}$$

where $\kappa(\mathcal{X}_n) = ||\mathcal{X}_n|| \cdot ||\mathcal{X}_n^{-1}||$ is the condition number of $\mathcal{X}_n$. A substantial amount of our technical work goes into proving

Lemma 2. *The condition number of $\kappa(\mathcal{X}_n)$ where $\mathcal{X}_n$ is our gauge change, is upper bounded by*

$$\kappa(\mathcal{X}_n) \leq (1-2k_n)^{-1} \quad \textit{with} \quad k_n = \frac{1+\tau_n^4}{1-\tau_n^4}\left(4\epsilon_n^2 + 10\epsilon_n^4\right), \tag{A.5}$$

which we prove in Sec. A.2.3. Notice that the condition number depends on both τ_n and ϵ_n, and in turn ϵ_{n+2} is now upper bounded by a function of only ϵ_n and τ_n. In Sec. A.2.3 we also introduce Thm. 8, an eigenvalue perturbation theorem, which we prove in Sec. A.2.4.

We already understand the iterative behaviour of ϵ_n, but not of τ_n. In Sec. A.2.5 we show that

Lemma 3. *The ergodicity coefficient $\mathcal{A}_n$ obeys*

$$\tau_{n+2} \leq \tau_n^4(1+\Delta_n) + \frac{1+\Delta_n}{1-\Delta_n}\Big(\Delta_n + 2\epsilon_n^2 + 5\epsilon_n^4\Big) \tag{A.6}$$

where

$$\Delta_n = \frac{k_n}{1-k_n} \tag{A.7}$$

with k_n as in Lemma A.5.

Notice that τ_{n+2} depends only on τ_n and ϵ_n. Therefore, these two lemmas provide a pair of coupled equations that provide $(\epsilon_{n+2}, \tau_{n+2})$ as a function of (ϵ_n, τ_n). Therefore, it is straightforward to numerically study the initial conditions (ϵ_0, τ_0) flow towards the desired point $(0,0)$, and this is presented in Fig. 8.3. These results show that a state within this distillable region is guaranteed to converge, but many steps in our argument make worst case assumptions and so we may also find convergence for many actual MPO states with (ϵ_0, τ_0) outside this region.

The distillable region is difficult to analytically characterise. However, we can analytically prove convergence on a slightly smaller region.

Lemma 4. *Given the iterative formulae for upper bounds on ϵ_n and τ_n, we know that $(\epsilon_n, \tau_n) \to (0,0)$ whenever*

$$\epsilon_0 \leq \min\left(\frac{1}{7}\frac{1-\tau_0^4}{1+\tau_0^4}\right). \tag{A.8}$$

We give a full proof in Sec. A.2.6. We have shown that the MPO noise matrices vanish, while the $\mathcal{A}_n$ remains trace-preserving and so of constant norm. As $\epsilon_n \to 0$ we have $\mathcal{A}_n \to \mathcal{E}_n$ and we can conclude convergence in fidelity as follows from lemma 12.

A.2. Mathematical Concepts and Proofs

A.2.1. Definitions and Properties of Norms

Properties of Norms

As norms for the states we use the Schatten p-norm, $\|\rho\|_p = \left(\sum_i \sigma_i(\rho)^p\right)^{1/p}$, where $\sigma_i(\rho)$ is the ith singular value of ρ. For the c.p. maps we will use norms defined over a variation with Schatten norms $\|\mathcal{P}\|_{a\to b} = \max_\sigma \frac{\|(\mathcal{P}\sigma)\|_a}{\|\sigma\|_b}$.

We can now use, that

$$\|\sigma_1\|_p = \max_{\sigma_2} \frac{\mathrm{Tr}[\sigma_2\sigma_1]}{\|\sigma_2\|_q} \quad \text{with} \quad \frac{1}{p} + \frac{1}{q} = 1 \text{ or } q = \frac{p}{p-1},$$

and deduce the following lemma.

Lemma 5. *Let $\mathcal{P}$ be a map. Then*

$$\|\mathcal{P}\|_{r\to s} = \left\|\mathcal{P}^\dagger\right\|_{\frac{s}{s-1}\to\frac{r}{r-1}}$$

where we use $\mathcal{M}^\dagger$ to denote the unique channel such that for all A, B we have $\mathrm{Tr}(A\mathcal{M}(B)) = \mathrm{Tr}\big(\mathcal{M}^\dagger(A)B\big)$.

$$\begin{aligned}
\|\mathcal{P}\|_{r\to s} &= \max_{\sigma_1} \frac{\|\mathcal{P}(\sigma_1)\|_r}{\|\sigma\|_s} && \text{(A.9)}\\
&= \max_{\sigma_1}\max_{\sigma_2} \frac{\mathrm{Tr}[\sigma_2\mathcal{P}(\sigma_1)]}{\|\sigma_2\|_{\frac{r}{r-1}}\|\sigma_1\|_s} && \text{(A.10)}\\
&= \max_{\sigma_1}\max_{\sigma_2} \frac{\mathrm{Tr}\left[\sigma_1\mathcal{P}^\dagger(\sigma_2)\right]}{\|\sigma_2\|_{\frac{r}{r-1}}\|\sigma_1\|_s} && \text{(A.11)}\\
&= \max_{\sigma_2} \frac{\left\|\mathcal{P}^\dagger(\sigma_2)\right\|_{\frac{s}{s-1}}}{\|\sigma_2\|_{\frac{r}{r-1}}} && \text{(A.12)}\\
&= \left\|\mathcal{P}^\dagger\right\|_{\frac{s}{s-1}\to\frac{r}{r-1}}. && \text{(A.13)}
\end{aligned}$$

Using Lemma 5, we state that $\|\mathcal{P}\|_{1\to 1} = \left\|\mathcal{P}^\dagger\right\|_{\infty\to\infty}$.

Lemma 6. *Let $\mathcal{P}$ be a positive map. Then the following holds:*

$$\max_{\sigma} \frac{\|\mathcal{P}(\sigma)\|_1}{\|\sigma\|_1} = \max_{\sigma \geq 0} \frac{\|\mathcal{P}(\sigma)\|_1}{\|\sigma\|_1}.$$

Assume that the maximum for $\mathcal{P}$ is reached by $\sigma_{\max} = \sigma^+ - \sigma^-$ where $\sigma^+, \sigma^- \geq 0$ and $\mathrm{Tr}\Big((\sigma^+)^\dagger \sigma^-\Big) = 0$. Though this matrix is potentially non-positive ($\sigma \neq 0$), we show that there always exists a non-negative matrix $\sigma_{\max}$ that also achieves the maximum. A completely positive map $\mathcal{P}$ preserves positivity, and so

$$\|\mathcal{P}(\sigma_{\max})\|_1 = \Big\|\mathcal{P}\Big(\sigma^+\Big) - \mathcal{P}\Big(\sigma^-\Big)\Big\|_1, \tag{A.14}$$

$$\leq \Big\|\mathcal{P}\Big(\sigma^+\Big)\Big\|_1 + \Big\|\mathcal{P}\Big(\sigma^-\Big)\Big\|_1. \tag{A.15}$$

where we again use the triangle inequality. For a positive Hermitian matrix $||M||_1 = \mathrm{Tr}(M)$, and since $\sigma^\pm$ are positive and $\mathcal{P}$ preserves positivity (it is a cp map) we infer

$$\|\mathcal{P}(\sigma_{\max})\|_1 \leq \mathrm{Tr}\Big[\mathcal{P}\Big(\sigma^+\Big)\Big] + \mathrm{Tr}\Big[\mathcal{P}\Big(\sigma^-\Big)\Big] \tag{A.16}$$

$$= \mathrm{Tr}\Big[\mathcal{P}\Big(\sigma^+\Big) + \mathcal{P}\Big(\sigma^-\Big)\Big] \tag{A.17}$$

$$= \mathrm{Tr}\Big[\mathcal{P}\Big(\sigma^+ + \sigma^-\Big)\Big] \tag{A.18}$$

$$= \Big\|\mathcal{P}\Big(\sigma^+ + \sigma^-\Big)\Big\|_1. \tag{A.19}$$

Additionally we have

$$\|\sigma_{\max}\|_1 = \sum_i |\lambda_i(\sigma_{\max})| = \Big\|\sigma^+ + \sigma^-\Big\|_1. \tag{A.20}$$

So we can conclude that

$$\frac{\|M(\sigma_{\max})\|_1}{\|\sigma_{\max}\|_1} \leq \frac{\Big\|M(\sigma^+ + \sigma^-)\Big\|_1}{\|\sigma^+ + \sigma^-\|_1}. \tag{A.21}$$

Therefore, there exists a strictly positive matrix $\sigma'_{\max} = \sigma^+ + \sigma^-$ that also achieved the maximum value.

Corollary 5. *If $\mathcal{P}$ is a positive map then*

$$\|\mathcal{P}\|_{1\to 1} = \Big\|\mathcal{P}^\dagger(\mathbb{1})\Big\|_\infty. \tag{A.22}$$

Since in the variational definition of the $1 \to 1$-norm stated above, the maximum is always reached on a positive state, we can use the trace properties and the variational relations of the Schatten norms.

$$\max_{\sigma \geq 0} \frac{\|\mathcal{P}(\sigma)\|_1}{\|\sigma\|_1} = \max_{\sigma \geq 0} \frac{\mathrm{Tr}[\mathcal{P}(\sigma)]}{\|\sigma\|_1} = \max_{\sigma \geq 0} \frac{\mathrm{Tr}[\mathbb{1}\mathcal{P}(\sigma)]}{\|\sigma\|_1} \tag{A.23}$$

$$= \max_{\sigma \geq 0} \frac{\mathrm{Tr}\Big[\mathcal{P}^\dagger(\mathbb{1})\sigma\Big]}{\|\sigma\|_1} = \Big\|\mathcal{P}^\dagger(\mathbb{1})\Big\|_\infty. \tag{A.24}$$

From this corollary, two further corollaries follow.

Corollary 6. *If $\mathcal{P}$ is a positive and trace-preserving map then $\|\mathcal{P}\|_{1\to 1} = 1$.*

Corollary 7. *For a channel of the form $\mathcal{X}(\rho) = \sqrt{\xi}\rho\sqrt{\xi}$ we have that*

$$\|\mathcal{X}\|_{1\to 1} = \left\|\sqrt{\xi}\mathbb{1}\sqrt{\xi}\right\|_\infty = \|\xi\|_\infty.$$

Ergodicity Coefficient and Fundamental Channel

We make frequent use of another concept originating from Markov chains, which is the ergodicity coefficient. It is a measure for how close a quantum channel is to a projection onto its steady state. The ergodicity is defined as follows. If ρ is the steady state of a completely positive map and $\|\sigma\|_1 = \mathrm{Tr}\left[\sqrt{\sigma^\dagger\sigma}\right]$ is the trace norm of σ, then

$$\tau(\mathcal{F}) = \max_{\mathrm{Tr}[\sigma\perp\rho]=0} \frac{\|\mathcal{F}(\sigma)\|_1}{\|\sigma\|_1}. \tag{A.25}$$

The ergodicity coefficient is similar to the second eigenvalue of the map, and in fact it is straightforward to see that it always upper bounds the second eigenvalue. The ergodicity coefficient is submultiplicative. If two maps $\mathcal{F}_1$ and $\mathcal{F}_2$ are trace-preserving then

$$\tau(\mathcal{F}_1\mathcal{F}_2) \leq \tau(\mathcal{F}_1)\tau(\mathcal{F}_2). \tag{A.26}$$

So the product of the two maps is at least as close to being a projector as both maps individually. Multiplying a lot of quantum channels, each with an ergodicity smaller than one, eventually leads to a projection.

A.2.2. Preserving the Trace

The observation that $A_iA_{i+1} = (A_iX)(X^{-1}A_{i+1})$ means that an MPO is not uniquely defined and offers a gauge freedom. Since we deal with unnormalised MPO, we also have the freedom of rescaling. There is a canonical gauge and scale corresponding to the transfer channel $\mathcal{E}$ being trace-preserving. We apply the gauge transformation $A_i \to A' = XAX^{-1}$ with $X = \sqrt{\xi}$. Consequently,

$$E \to E' = \left(\sqrt{\xi}\otimes\sqrt{\xi}\right)E\left(\sqrt{\xi^{-1}}\otimes\sqrt{\xi^{-1}}\right) \quad \text{and} \tag{A.27}$$

$$\mathcal{E}(\rho) \to \sqrt{\xi}\mathcal{E}\left(\sqrt{\xi^{-1}}\rho\sqrt{\xi^{-1}}\right)\sqrt{\xi}. \tag{A.28}$$

Now we are in a gauge where $\mathbb{1}$ is the left Perron state.

$$\mathcal{E}^\dagger(\mathbb{1}) \to \sqrt{\xi^{-1}}\mathcal{E}\left(\sqrt{\xi}\mathbb{1}\sqrt{\xi}\right)\sqrt{\xi^{-1}} = \sqrt{\xi^{-1}}\mathcal{E}(\xi)\sqrt{\xi^{-1}}, \tag{A.29}$$

$$= \sqrt{\xi^{-1}}\xi\sqrt{\xi^{-1}} = \mathbb{1}. \tag{A.30}$$

A.2.3. Condition Number

Here we prove the condition number bound presented as Lem. 2. The condition number is small provided ξ_n is close to $\mathbb{1}$. Specifically, following Cor. (7) we can upper bound

$$\|\mathcal{X}\|_{1\to 1} \leq \|\xi\|_\infty, \tag{A.31}$$
$$\left\|\mathcal{X}^{-1}\right\|_{1\to 1} \leq \left\|\xi^{-1}\right\|_\infty, \tag{A.32}$$

and so $\kappa(\mathcal{X}_n) \leq \|\xi\|_\infty \left\|\xi^{-1}\right\|_\infty$. Observe that if ξ_n is close to the identity then $\delta_n = \xi_n - \mathbb{1}$ is small, and it is helpful to reformulate in terms of $\Delta_n := \|\delta_n\|_\infty$. We find

$$\|\xi\|_\infty = \|\mathbb{1} - (\mathbb{1} - \xi)\|_\infty \leq 1 + \|\mathbb{1} - \xi\|_\infty = 1 + \Delta_n. \tag{A.33}$$

Similarly we can deduce

$$\mathbb{1} = \xi^{-1} - \xi^{-1}(\mathbb{1} - \xi), \tag{A.34}$$
$$\|\mathbb{1}\|_\infty = \left\|\xi^{-1} - \xi^{-1}(\mathbb{1} - \xi)\right\|_\infty, \tag{A.35}$$
$$1 \geq \left\|\xi^{-1}\right\|_\infty - \left\|\xi^{-1}\right\|_\infty \|\mathbb{1} - \xi\|_\infty, \tag{A.36}$$
$$\Rightarrow \left\|\xi^{-1}\right\|_\infty \leq \frac{1}{1-\Delta}. \tag{A.37}$$

This give a new expression for the condition number

$$\kappa(\mathcal{X}_n) = \|\xi_n\|_\infty \|\xi_n^{-1}\|_\infty \leq \frac{1+\Delta_n}{1-\Delta_n}. \tag{A.38}$$

To proceed we need to upper bound Δ_n, which we achieve using the following powerful result.

Theorem 8. *(Eigenvalue Perturbation Theorem) Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be completely positive maps with spectral radius 1, and $\mathcal{P}_1$ is trace-preserving. If ξ is the left Perron state for $\mathcal{P}_2$, so that $\mathcal{P}_2^\dagger(\xi) = \xi$, then*

$$\|\mathbb{1} - \xi\|_\infty \leq \frac{k}{1-k} \tag{A.39}$$

where

$$k = \left(\frac{1+\tau(\mathcal{M}_1)}{1-\tau(\mathcal{M}_1)}\right) \|\mathcal{P}_1 - \mathcal{P}_2\|_{1\to 1}. \tag{A.40}$$

We prove this result in the following subsections, but here make direct use of it. We set $\mathcal{M}_1 = \mathcal{A}_n^4$, which inherits the required properties from $\mathcal{A}_n$. Likewise, we set $\mathcal{M}_2 = \lambda^{-1}(\mathcal{A}_n^4 + \mathcal{P}_n)$ where $\mathcal{P}_n := \mathcal{A}'_{n+2} - \mathcal{A}_n^4$ is a perturbation composed of noise matrices and λ is a normalisation constant ensuring that $\mathcal{M}_2$ has spectral radius 1. Note

that $\mathcal{M}_2$ differs from $\mathcal{A}'_{n+2}$ by only a constant so they have the same eigenvectors, e.g. ξ. Therefore, the theorem tells us that $\Delta_n \leq k_n/(1-k_n)$ where

$$k_n = \left(\frac{1+\tau(\mathcal{A}_n^4)}{1-\tau(\mathcal{A}_n^4)}\right) \left\|\mathcal{A}_n^4 - \lambda^{-1}(\mathcal{A}_n^4 + \mathcal{P}_n)\right\|_{1\to 1} \tag{A.41}$$

Since $\mathcal{A}_n$ is trace-preserving $\tau(\mathcal{A}_n^4) = \tau(\mathcal{A}_n)^4 = \tau_n$. Looking at the second factor, we collect the $\mathcal{A}_n^4$ terms and use the triangle inequality

$$\begin{aligned}\|\mathcal{P}_1 - \mathcal{P}_2\|_{1\to 1} &= \left\|(1-\lambda^{-1})\mathcal{A}_n^4 - \lambda^{-1}\mathcal{P}_2\right\|_{1\to 1} \\ &\leq \left|1-\lambda^{-1}\right| \left\|\mathcal{A}_n^4\right\|_{1\to 1} + \lambda^{-1}\left\|\mathcal{P}_n^4\right\|_{1\to 1}.\end{aligned}$$

To proceed we need information about λ, which is the spectral radius of $\mathcal{A}_n^4 + \mathcal{P}_n$ and so $\lambda \leq 1 + \|\mathcal{P}_n\|_{1\to 1}$. Furthermore, because $\mathcal{A}_n^4$ and $\mathcal{P}_n$ are both positive channels, we know that λ must exceed the spectral radius of $\mathcal{A}_n^4$ and so $1 \leq \lambda$. Therefore, $\lambda^{-1} \leq 1$ and $|1-\lambda^{-1}| \leq \|\mathcal{P}_n\|_{1\to 1}$. Combining these observations we have

$$\|\mathcal{P}_1 - \mathcal{P}_2\|_{1\to 1} \leq 2\|\mathcal{P}_n\|_{1\to 1}.$$

and so

$$k_n = 2\left(\frac{1+\tau_n^4}{1-\tau_n^4}\right)\|\mathcal{P}_n\|_{1\to 1}. \tag{A.42}$$

To upper bound $\left\|\mathcal{P}_n^4\right\|_{1\to 1}$ we have to refer back to the iterative formulae and use norm submultiplicatively to show $\left\|\mathcal{P}_n^4\right\|_{1\to 1} \leq 2\epsilon_n^2 + 5\epsilon_n^4$. Substituting this into k_n we get

$$k_n = \left(\frac{1+\tau_n^4}{1-\tau_n^4}\right)(4\epsilon_n^2 + 10\epsilon_n^4), \tag{A.43}$$

which proves Lem. 2. We give a quick overview of the bounds we have derived.

$$\|\mathbb{1} - \mathcal{X}\|_{1\to 1} \leq 3\Delta, \tag{A.44}$$

$$\left\|\mathbb{1} - \mathcal{X}^{-1}\right\|_{1\to 1} \leq \frac{3\Delta}{1-\Delta}, \tag{A.45}$$

$$\|\mathcal{X}\|_{1\to 1} \leq 1 + \Delta, \tag{A.46}$$

$$\left\|\mathcal{X}^{-1}\right\|_{1\to 1} \leq \frac{1}{1-\Delta}, \tag{A.47}$$

$$\kappa(X) \leq \frac{1+\Delta}{1-\Delta}. \tag{A.48}$$

However, the proof rests upon Thm.8, which we turn to in the next section.

A.2.4. Proof of the Eigenvalue Perturbation Theorem

Our methodology for proving Thm.8 is in the spirit of [92], but generalised so that one of the channels need not be trace-preserving. The proof requires some new concepts we have not yet introduced, including the fundamental channel.

Definition 7. *Let $\mathcal{P}$ be a channel, then the fundamental channel of $\mathcal{P}$ is*

$$\mathcal{Z} = (\mathbb{1} - \mathcal{P} + \mathcal{P}^{\infty})^{-1} \tag{A.49}$$

This definition is central to the following two lemmas

Lemma 8. *$\mathcal{P}_1$ and $\mathcal{P}_2$ are completely positive maps with spectral radius 1, and $\mathcal{P}_1$ is trace-preserving. $\mathcal{Z}_1$ is the fundamental channel of $\mathcal{P}_1$. The left Perron state for $\mathcal{P}_2$ is ξ. Then we can say*

$$\|\mathbb{1} - \xi\|_{\infty} \leq \frac{\|\mathcal{Z}_1\|_{1\to 1}\|\mathcal{P}_1 - \mathcal{P}_2\|_{1\to 1}}{1 - \|\mathcal{Z}_1\|_{1\to 1}\|\mathcal{P}_1 - \mathcal{P}_2\|_{1\to 1}}$$

and

Lemma 9. *Let $\mathcal{P}$ be a CPT-map, and denote $\mathcal{Z}$ to be the fundamental channel of $\mathcal{P}$. It follows that*

$$\|\mathcal{Z}\|_{1\to 1} \leq \frac{1 + \tau(\mathcal{P})}{1 - \tau(\mathcal{P})}.$$

Combining these results straightforwardly leads to Thm.8, and so the remainder of this section will prove these lemmas. We being by bounding how much the left Perron state is perturbed from the identity. First we look at the projectors of our maps in the matrix picture.

$$P_1^{\infty} = |\rho_1\rangle\langle\mathbb{1}| \quad \text{with} \quad \langle\mathbb{1}|\rho_1\rangle = \mathrm{Tr}[\rho_1] = 1, \tag{A.50}$$

$$P_2^{\infty} = |\rho_2\rangle\langle\xi| \quad \text{with} \quad \langle\xi|\rho_2\rangle = \mathrm{Tr}\left[\xi^{\dagger}\rho_2\right] = 1. \tag{A.51}$$

Since the eigenvector matters only up to a constant we can rescale $|\xi\rangle$ and $|\rho_2\rangle$ so they still satisfy $\mathrm{Tr}\left[\xi^{\dagger}\rho_2\right] = 1$, but also satisfy $\langle\rho_1|\xi\rangle = \mathrm{Tr}\left[\rho_1^{\dagger}\xi\right] = 1$.

Since we are dealing with the left eigenvectors we have to transpose our maps for easier notation. Thus the transposed projectors are

$$P_1^{\infty\dagger} = |\mathbb{1}\rangle\langle\rho_1| \quad \text{with} \quad \langle\mathbb{1}|\rho_1\rangle = 1, \tag{A.52}$$

$$P_2^{\infty\dagger} = |\xi\rangle\langle\rho_2| \quad \text{with} \quad \langle\xi|\rho_1\rangle = 1. \tag{A.53}$$

We start by applying $\left(\mathcal{Z}_1^\dagger\right)^{-1}$, the inverse fundamental matrix of transposed $\mathcal{P}_1$ to the difference between ξ and the identity matrix $\mathbb{1}$.

$$\begin{aligned}\left(\mathcal{Z}_1^\dagger\right)^{-1}(\mathbb{1}-\xi) &= \left(\mathbb{1}-\mathcal{P}_1^\dagger+\mathcal{P}_1^{\infty\dagger}\right)(\mathbb{1}-\xi) && \text{(A.54)}\\ &= \mathbb{1}-\mathbb{1}+\mathbb{1}-\xi+\mathcal{P}_1^\dagger(\xi)-\mathbb{1}\,\mathrm{Tr}\left[\rho_1^\dagger\xi\right] \\ &= \xi+\mathcal{P}_1^\dagger(\xi) && \text{(A.55)}\end{aligned}$$

Going from first to second lines, we have used $\mathcal{P}_1^\dagger(\mathbb{1})=\mathbb{1}$ and $\mathcal{P}_1^{\infty\dagger}(\mathbb{1})=\mathbb{1}$. Going from second to third we use the normalisation condition $\mathrm{Tr}\left[\rho_1^\dagger\xi\right]=1$ and cancel the identities. It is a condition of the lemma that $\xi=\mathcal{P}_2^\dagger(\xi)$ and so

$$\left(\mathcal{Z}_1^\dagger\right)^{-1}(\mathbb{1}-\xi)=\left(\mathcal{P}_1^\dagger-\mathcal{P}_2^\dagger\right)(\xi) \tag{A.56}$$

Now we multiply by $\mathcal{Z}_1^\dagger$ on both sides and take the ∞-norm.

$$\begin{aligned}\|(\mathbb{1}-\xi)\|_\infty &= \left\|\mathcal{Z}_1^\dagger\circ(\mathcal{P}_1^\dagger-\mathcal{P}_2^\dagger)(\xi)\right\|_\infty && \text{(A.57)}\\ &\leq \frac{\left\|\mathcal{Z}_1^\dagger\circ(\mathcal{P}_1^\dagger-\mathcal{P}_2^\dagger)(\xi)\right\|_\infty}{\|\xi\|_\infty}\|\xi\|_\infty && \text{(A.58)}\\ &\leq \left\|\mathcal{Z}_1^\dagger\circ(\mathcal{P}_1^\dagger-\mathcal{P}_2^\dagger)\right\|_{\infty\to\infty}\|\xi\|_\infty && \text{(A.59)}\\ &\leq \left\|\mathcal{Z}_1^\dagger\circ(\mathcal{P}_1^\dagger-\mathcal{P}_2^\dagger)\right\|_{\infty\to\infty}(\|\mathbb{1}\|_\infty+\|\mathbb{1}-\xi\|_\infty)\end{aligned}$$

Now we use Lem. 5 to deal with the adjoints and use that the identity has ∞-norm 1.

$$\|(\mathbb{1}-\xi)\|_\infty\leq\|(\mathcal{P}_1-\mathcal{P}_2)\circ\mathcal{Z}_1\|_{1\to1}(1+\|\mathbb{1}-\xi\|_\infty)$$

Now we subtract the term $\|(\mathcal{P}_1-\mathcal{P}_2)\circ\mathcal{Z}_1\|_{1\to1}\|\mathbb{1}-\xi\|_\infty$ from both sides.

$$\begin{aligned}\|(\mathbb{1}-\xi)\|_\infty(1-\|(\mathcal{P}_1-\mathcal{P}_2)\circ\mathcal{Z}_1\|_{1\to1}) \\ \leq\|(\mathcal{P}_1-\mathcal{P}_2)\circ\mathcal{Z}_1\|_{1\to1} && \text{(A.60)}\end{aligned}$$

Assuming that $1-\|(\mathcal{P}_1-\mathcal{P}_2)\circ\mathcal{Z}_1\|_{1\to1}\geq0$, which is true for $\mathcal{P}_1$ and $\mathcal{P}_2$ being close enough we divide by $1-\|(\mathcal{P}_1-\mathcal{P}_2)\circ\mathcal{Z}_1\|_{1\to1}$.

$$\|\mathbb{1}-\xi\|_\infty\leq\frac{\|\mathcal{Z}_1\|_{1\to1}\|\mathcal{P}_1-\mathcal{P}_2\|_{1\to1}}{1-\|\mathcal{Z}_1\|_{1\to1}\|\mathcal{P}_1-\mathcal{P}_2\|_{1\to1}} \tag{A.61}$$

This completes the proof of Lem. 8.

Next we turn our attention to showing Lem. 9. First we reformulate the definition of the fundamental channel.

$$\mathcal{Z}(\mathcal{P}) = (\mathbb{1} - \mathcal{P} + \mathcal{P}^\infty)^{-1} = \sum_{k=0}^{\infty} (\mathcal{P} - \mathcal{P}^\infty)^k \tag{A.62}$$

$$= \mathbb{1} + \sum_{k=1}^{\infty} \left(\mathcal{P}^k - \mathcal{P}^\infty\right) \tag{A.63}$$

$$= \mathbb{1} + \sum_{k=0}^{\infty} (\mathcal{P} - \mathcal{P}^\infty)^k (\mathcal{P} - \mathcal{P}^\infty) \tag{A.64}$$

$$= \mathbb{1} + \sum_{k=0}^{\infty} \mathcal{P}^k (\mathcal{P} - \mathcal{P}^\infty) \tag{A.65}$$

In those steps we used that $(\mathcal{P}_1 - \mathcal{P}^\infty)^k = P_1^k - \mathcal{P}^\infty$ and that $\mathcal{P}^\infty(\mathcal{P} - \mathcal{P}^\infty) = 0$. It is important to note that for any σ, the expression $(\mathcal{P} - \mathcal{P}^\infty)(\sigma)$ is traceless.

$$\|\mathcal{Z}\|_{1\to 1} = \max_{\sigma_1} \frac{\left\|\left(\mathbb{1} + \sum_{k=0}^{\infty} \mathcal{P}^k \circ (\mathcal{P} - \mathcal{P}^\infty)\right)(\sigma_1)\right\|_1}{\|\sigma_1\|_1} \tag{A.66}$$

$$\leq \max_{\sigma_1} \frac{\|\mathbb{1}(\sigma_1)\|_1}{\|\sigma_1\|_1} + \frac{\left\|\sum_{k=0}^{\infty} \mathcal{P}^k \circ (\mathcal{P} - \mathcal{P}^\infty)(\sigma_1)\right\|_1}{\|\sigma_1\|_1} \tag{A.67}$$

$$\leq 1 + \max_{\sigma_1} \frac{\left\|\sum_{k=0}^{\infty} \mathcal{P}^k(\mathcal{P} - \mathcal{P}^\infty(\sigma_1))\right\|_1}{\|\mathcal{P} - \mathcal{P}^\infty(\sigma_1)\|_1} \frac{\|\mathcal{P} - \mathcal{P}^\infty(\sigma_1)\|_1}{\|\sigma_1\|_1} \tag{A.68}$$

$$\leq 1 + \max_{\mathrm{Tr}[\sigma_2]=0} \frac{\left\|\sum_{k=0}^{\infty} \mathcal{P}^k(\sigma_2)\right\|_1}{\|\sigma_2\|_1} \max_{\sigma_1} \frac{\|\mathcal{P} - \mathcal{P}^\infty(\sigma_1)\|_1}{\|\sigma_1\|_1}. \tag{A.69}$$

We now upper bound both terms separately. We note that $\mathcal{P} - \mathcal{P}^\infty = \mathcal{P}(\mathbb{1} - \mathcal{P}^\infty)$.

$$\max_{\sigma_1} \frac{\|(\mathcal{P} - \mathcal{P}^\infty)(\sigma_1)\|_1}{\|\sigma_1\|_1}$$

$$= \max_{\sigma_1} \frac{\|\mathcal{P}(\mathbb{1} - \mathcal{P}^\infty(\sigma_1))\|_1}{\|\mathbb{1} - \mathcal{P}^\infty(\sigma_1)\|_1} \frac{\|\mathbb{1} - \mathcal{P}^\infty(\sigma_1)\|_1}{\|\sigma_1\|_1} \tag{A.70}$$

$$\leq \max_{\mathrm{Tr}[\sigma_1]=0} \frac{\|\mathcal{P}(\mathbb{1} - \mathcal{P}^\infty(\sigma_1))\|_1}{\|\mathbb{1} - \mathcal{P}^\infty(\sigma_1)\|_1} \max_{\sigma_2} \frac{\|(\mathbb{1} - \mathcal{P}^\infty)(\sigma_1)\|_1}{\|\sigma_1\|_1} \tag{A.71}$$

$$\leq \tau(\mathcal{P}) \|\mathbb{1} - \mathcal{P}^\infty\|_{1\to 1} \leq \tau(\mathcal{P})(\|\mathbb{1}\|_{1\to 1} + \|\mathcal{P}^\infty\|_{1\to 1}) \tag{A.72}$$

$$\leq 2\tau(\mathcal{P}). \tag{A.73}$$

Here we have used the fact that both $\mathbb{1}$ and $\mathcal{P}^\infty$ are trace-preserving and completely

positive and Cor. (6).

$$\max_{\mathrm{Tr}[\sigma_2]=0} \frac{\left\|\sum_{k=0}^{\infty}\mathcal{P}^k(\sigma_2)\right\|_1}{\|\sigma_2\|_1} \leq \sum_{k=0}^{\infty} \max_{\mathrm{Tr}[\sigma_1]=0} \frac{\left\|\mathcal{P}^k(\sigma_1)\right\|_1}{\|\sigma_1\|_1} \tag{A.74}$$

$$= \sum_{k=0}^{\infty} \tau\Big(\mathcal{P}^k\Big) \leq \sum_{k=0}^{\infty} \tau(\mathcal{P})^k = \frac{1}{1-\tau(\mathcal{P})}. \tag{A.75}$$

Thus we can conclude that

$$\|\mathcal{Z}\|_{1\to 1} \leq 1 + \frac{2\tau(\mathcal{P})}{1-\tau(\mathcal{P})} = \frac{1+\tau(\mathcal{P})}{1-\tau(\mathcal{P})}. \tag{A.76}$$

This completes the proof of Lem. 9.

A.2.5. Change of the Ergodicity

After regauging, the ergodicity coefficient changes. This change can be bounded in the following way.

Lemma 10. *Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be trace-preserving, c.p. maps. Then we have*

$$|\tau(\mathcal{P}_1) - \tau(\mathcal{P}_2)| \leq \tau(\mathcal{P}_1 - \mathcal{P}_2).$$

Without loss of generality we assume $\tau(\mathcal{P}_1)$ to be bigger than $\tau(\mathcal{P}_2)$.

$$|\tau(\mathcal{P}_1) - \tau(\mathcal{P}_2)| = \max_{\mathrm{Tr}[\sigma_1]=0} \frac{\|\mathcal{P}_1(\sigma_1)\|_1}{\|\sigma_1\|_1} - \max_{\mathrm{Tr}[\sigma_2]=0} \frac{\|\mathcal{P}_2(\sigma_2)\|_1}{\|\sigma_2\|_1}, \tag{A.77}$$

$$\leq \max_{\mathrm{Tr}[\sigma_1]=0} \frac{\|\mathcal{P}_1(\sigma_1)\|_1}{\|\sigma_1\|_1} - \frac{\|\mathcal{P}_2(\sigma_1)\|_1}{\|\sigma_1\|_1} \tag{A.78}$$

$$\leq \max_{\mathrm{Tr}[\sigma_1]=0} \frac{\|(\mathcal{P}_1 - \mathcal{P}_2)(\sigma_1)\|_1}{\|\sigma_1\|_1} \tag{A.79}$$

$$= \tau(\mathcal{P}_1 - \mathcal{P}_2). \tag{A.80}$$

Next, we still need to bound the ergodicity coefficient for $\mathcal{X}_n\Big(\mathcal{A}_n^4 + \mathcal{P}_n\Big)\mathcal{X}_n^{-1}$. For this purpose we use Lem. 10 (in the second step).

$$\begin{aligned}
&\tau\Big(\mathcal{X}_n\Big(\mathcal{A}_n^4 + \mathcal{P}_n\Big)\mathcal{X}_n^{-1}\Big) \\
&\leq \tau\Big(\mathcal{A}_n^4\Big) + \Big|\tau\Big(\mathcal{A}_n^4\Big) - \tau\Big(\mathcal{X}_n\Big(\mathcal{A}_n^4 + \mathcal{P}_n\Big)\mathcal{X}_n^{-1}\Big)\Big| \\
&\leq \tau(\mathcal{A}_n)^4 + \tau\Big(\mathcal{A}_n^4 - \mathcal{X}_n\Big(\mathcal{A}_n^4 + \mathcal{P}_n\Big)\mathcal{X}_n^{-1}\Big).
\end{aligned} \tag{A.81}$$

Subsequently, we bound the second term.

$$\begin{aligned}
&\mathcal{A}_n^4 - \mathcal{X}_n\Big(\mathcal{A}_n^4 + \mathcal{P}_n\Big)\mathcal{X}_n^{-1} \\
&= \mathcal{A}_n^4 - \mathcal{X}_n\mathcal{A}_n^4\mathcal{X}_n^{-1} - \mathcal{X}_n\mathcal{P}_n\mathcal{X}_n^{-1} \\
&= (\mathbb{1} - \mathcal{X}_n)\mathcal{A}_n^4 + \mathcal{X}_n\mathcal{A}_n^4(\mathbb{1} - \mathcal{X}_n^{-1}) - \mathcal{X}_n\mathcal{P}_n\mathcal{X}_n^{-1}.
\end{aligned}$$

Next, we use the fact that $\tau(\mathcal{F}_1 + \mathcal{F}_2) \leq \tau(\mathcal{F}_1) + \tau(\mathcal{F}_2)$ and that $\tau(\mathcal{F}) \leq \|\mathcal{F}\|_{1\to 1}$ in the following steps,

$$\begin{aligned}
&\tau\Big(\mathcal{A}_n^4 - \mathcal{X}_n\Big(\mathcal{A}_n^4 + \mathcal{P}_n\Big)\mathcal{X}_n^{-1}\Big) \\
&\quad\leq \tau\Big((\mathbb{1} - \mathcal{X}_n)\mathcal{A}_n^4\Big) + \tau\Big(\mathcal{X}_n\mathcal{A}_n^4(\mathbb{1} - \mathcal{X}_n^{-1})\Big) + \tau\Big(\mathcal{X}_n\mathcal{P}_n\mathcal{X}_n^{-1}\Big) \\
&\quad\leq \tau(\mathbb{1} - \mathcal{X}_n)\tau\Big(\mathcal{A}_n^4\Big) + \Big\|\mathcal{X}_n\mathcal{A}_n^4(\mathbb{1} - \mathcal{X}_n^{-1})\Big\|_{1\to 1} + \Big\|\mathcal{X}_n\mathcal{P}_n\mathcal{X}_n^{-1}\Big\|_{1\to 1} \\
&\quad\leq \|\mathbb{1} - \mathcal{X}_n\|_{1\to 1}\tau(\mathcal{A}_n)^4 + \|\mathcal{X}_n\|_{1\to 1}\Big\|\mathcal{A}_n^4\Big\|_{1\to 1}\Big\|\mathbb{1} - \mathcal{X}_n^{-1}\Big\|_{1\to 1} \\
&\qquad + \|\mathcal{X}_n\|_{1\to 1}\|\mathcal{P}_n\|_{1\to 1}\Big\|\mathcal{X}_n^{-1}\Big\|_{1\to 1}.
\end{aligned}$$

Using equations (A.44) to (A.48) we conclude that

$$\begin{aligned}
&\tau\Big(\mathcal{A}_n^4 - \mathcal{X}_n\Big(\mathcal{A}_n^4 + \mathcal{P}_n\Big)\mathcal{X}_n^{-1}\Big) \\
&\leq \tau(\mathcal{A}_n)^4 3\Delta_n + \frac{1+\Delta_n}{1-\Delta_n}3\Delta_n + \frac{1+\Delta_n}{1-\Delta_n}\|\mathcal{P}_n\|_{1\to 1}.
\end{aligned} \tag{A.82}$$

Therefore,

$$\tau\Big(\mathcal{X}_n\Big(\mathcal{A}_n^4 + \mathcal{P}_n\Big)\mathcal{X}_n^{-1}\Big) \leq \tau\Big(\mathcal{A}_n^4\Big)(1 + 3\Delta_n) + \frac{1+\Delta_n}{1-\Delta_n}\left(3\Delta_n + \|\mathcal{P}_n\|_{1\to 1}\right),$$

Next, we can upper bound the new ergodicity $\tau(\mathcal{A}_{n+2})$.

$$\begin{aligned}
\tau(\mathcal{A}_{n+2}) &= \tau\left(\frac{\mathcal{X}_n\Big(\mathcal{A}_n^4 + \mathcal{P}_n\Big)\mathcal{X}_n^{-1}}{1 + \delta_R}\right) \\
&\leq \tau\Big(\mathcal{X}_n\Big(\mathcal{A}_n^4 + \mathcal{P}_n\Big)\mathcal{X}_n^{-1}\Big).
\end{aligned} \tag{A.83}$$

It follows, from assuming the worst case of a growing ergodicity, that

$$\tau(\mathcal{A}_{n+2}) \leq \tau(\mathcal{A}_n)^4(1 + 3\Delta_n) + \frac{1+\Delta_n}{1-\Delta_n}(3\Delta_n + \|\mathcal{P}_n\|_{1\to 1}). \tag{A.84}$$

A.2.6. Simplifying Parameter Space

In the following section, we will make use of the abbreviations $\tau_n = \tau(A_n)$, $Z_n = \frac{1+\tau_n^4}{1-\tau_n^4}$ and $P_n = \|\mathcal{P}_n\|_{1\to 1}$. With these definitions, the bounds are presented as follows:

$$P_n \leq 2\epsilon_n^2 + 5\epsilon_n^4, \tag{A.85}$$

$$\Delta_n \leq \frac{2Z_nP_n}{1 - 2Z_nP_n}, \tag{A.86}$$

$$\tau_{n+2} \leq \tau_n^4(1 + 3\Delta_n) + \frac{1+\Delta}{1-\Delta_n}(3\Delta_n + P_n), \tag{A.87}$$

$$\epsilon_{n+2} \leq 4\frac{1+\Delta_n}{1-\Delta_n}\Big(\epsilon_n^2 + \epsilon_n^4\Big). \tag{A.88}$$

To derive a manageable convergence threshold for ϵ_0 depending τ_0, we chose the ansatz

$$Z_n \epsilon_n \leq \frac{1}{7}. \tag{A.89}$$

Next, we show that given (A.89) for all n, ϵ_n converges to zero. Subsequently, we demonstrate that if (A.89) is fulfilled at step n, it is also fulfilled for $n+2$. First, we assert that (A.89) implies

$$\epsilon_n \leq \frac{1}{7}\frac{1-\tau_n^4}{1+\tau_n^4} \leq \frac{1}{7}. \tag{A.90}$$

Now we can upper bound P_n,

$$P_n \leq \left(2 + 5\epsilon_n^2\right)\epsilon_n^2 \leq \frac{11}{5}\epsilon_n^2. \tag{A.91}$$

Furthermore

$$\Delta_n \leq \frac{2Z_n P_n}{1 - 2Z_n P_n} \leq \frac{2\frac{11}{5} Z_n \epsilon_n^2}{1 - 2\frac{11}{5} Z_n \epsilon_n^2}, \tag{A.92}$$

$$\leq \frac{\frac{2\cdot 11}{5\cdot 7}\epsilon_n}{1 - \frac{2\cdot 11}{5\cdot 7}\epsilon_n} = \frac{\frac{22}{35}\epsilon_n}{1 - \frac{22}{35}\epsilon_n} \leq \frac{7}{10}\epsilon_n, \tag{A.93}$$

is true for all $\epsilon_n \leq \frac{1}{7}$ and leads us to

$$\frac{1+\Delta_n}{1-\Delta_n} \leq 1 + \frac{31}{20}\epsilon_n. \tag{A.94}$$

Finally, we can upper bound ϵ_{n+2}

$$\epsilon_{n+2} \leq 4\left(1 + \frac{31}{20}\epsilon_n\right)\left(1 + \epsilon_n^2\right) \leq 5\epsilon_n^2. \tag{A.95}$$

This converges to zero for $\epsilon_0 < \frac{1}{5}$ which is implied by (A.89). Next, we show that if (A.89) is true for step n it also holds for step $n+2$. We look at the update rule of τ_n, (A.87), and insert the bounds (A.91), (A.93) and (A.94).

$$\begin{aligned}\tau_{n+2} &\leq \tau_n^4\left(1 + \frac{21}{10}\epsilon_n\right) + \left(1 + \frac{31}{20}\right)\left(\frac{21}{10} + \frac{11}{5}\epsilon_n\right)\epsilon_n \\ &\leq \tau_n^4 + \tau_n^4\frac{21}{10}\epsilon_n + 3\epsilon_n \end{aligned} \tag{A.96}$$

Inserting (A.90) yields

$$\tau_n^4 + \tau_n^4\frac{3}{10}\frac{1-\tau_n^4}{1+\tau_n^4} + \frac{3}{7}\frac{1-\tau_n^4}{1+\tau_n^4}.$$

We look at two different cases, $0.5 \leq \tau_n \leq 1$ and $0 \leq \tau_n \leq 0.5$. In the first regime, we know

$$\tau_{n+2} \leq \tau_n^4 + \tau_n^4\frac{3}{10}\frac{1-\tau_n^4}{1+\tau_n^4} + \frac{3}{7}\frac{1-\tau_n^4}{1+\tau_n^4} \leq \tau_n.$$

Therefore, if (A.89) is fulfilled, $\tau_{n+2} \leq \tau_n$ and from (A.95) we know $\epsilon_{n+2} \leq \epsilon_n$. If ϵ_n and τ_n satisfy ansatz (A.89), than ϵ_{n+2} and τ_{n+2} necessarily do the same, since they are both smaller. In the second regime we know

$$\tau_{n+2} \leq \tau_n^4 + \tau_n^4 \frac{3}{10} \frac{1-\tau_n^4}{1+\tau_n^4} + \frac{3}{7} \frac{1-\tau_n^4}{1+\tau_n^4} \leq 0.5,$$

so once τ_n is smaller than 0.5 it stays smaller than 0.5. Therefore,

$$Z_{n+2}\epsilon_{n+2} \leq \frac{17}{15} 5\epsilon_n^2 \leq \frac{17}{15} \frac{5}{7} \epsilon_n \leq \frac{1}{7},$$

and the ansatz holds true in step $n+2$. By induction we can conclude that if (A.89) is satisfied for $n = 0$ it stays satisfied for all $n = 2k$ with $k \in \mathbb{N}$. The assumption for step $n = 0$ can be reformulated in the form

$$\epsilon_0 \leq \frac{1}{7} \frac{1-\tau_0^4}{1+\tau_0^4}. \tag{A.97}$$

The speed of the convergence is given by

$$\epsilon_n \leq \frac{1}{5} (5\epsilon_0)^{2^{n/2}}.$$

This concludes the derivation of a threshold for the recurrence protocol that ensures convergence to maximally entangled pure states.

A.3. Bounding Observables

So far we always dealt with unnormalised MPOs, since it allows the explicit description of the MPO matrices independent of the length of the MPO. An MPO with the same matrices but of a different length will generally have a different norm.

We now show that if the transfer matrix is trace-preserving the norm of an MPO will be exponentially close to unity in L, the length of the chain.

Lemma 11. *If E is isomorphic to $\mathcal{E}$, a trace-preserving completely positive map, then*

$$1 - d^{5/2} \tau(\mathcal{E})^L \leq \mathrm{Tr}\Big[E^L\Big] \leq 1 + d^{5/2} \tau(\mathcal{E})^L,$$

where $\tau(\mathcal{E})$ is the ergodicity coefficient of $\mathcal{E}$.

Proof. The proof is based on the idea to write the trace explicitly in the matrix picture for E as a sum over a basis. We chose the generalised Pauli matrices P, which vectorised

and normalised will give us a convenient basis for the trace. We use that except for the identity, all these matrices have trace equal to zero.

$$\mathrm{Tr}\Big[E^L\Big] = \left\langle \frac{\mathbb{1}}{\sqrt{d}} \middle| E_n^{L_n-1} \middle| \frac{\mathbb{1}}{\sqrt{d}} \right\rangle + \sum_{\sigma \in P/\mathbb{1}} \langle \sigma | E_n^{L_n} | \sigma \rangle.$$

Now we change into the channel picture to use the ergodicity coefficient.

$$\mathrm{Tr}\Big[E^L\Big] \geq \frac{1}{d}\,\mathrm{Tr}\Big[\mathcal{E}^\dagger(\mathbb{1})\Big] - \max_{\mathrm{Tr}[\sigma]=0} d^2 \frac{\mathrm{Tr}\Big[\sigma \mathcal{E}_n^{L_n}(\sigma)\Big]}{\|\sigma\|_2^2}.$$

Next, we can use that the identity is a fixed point of $\mathcal{E}^\dagger$. We can freely lower bound by changing from $\|\sigma\|_2$ to $\|\sigma\|_\infty$, but we have to introduce a factor $\sqrt{d}$ to replace $\|\sigma\|_2$ with $\|\sigma\|_1$.

$$\begin{aligned}
\mathrm{Tr}\Big[E^L\Big] &\geq 1 - d^2\sqrt{d} \max_{\mathrm{Tr}[\sigma]=0} \frac{\mathrm{Tr}\Big[\sigma \mathcal{E}^L(\sigma)\Big]}{\|\sigma\|_\infty \|\sigma\|_1}, \\
&\geq 1 - d^{5/2} \max_{\mathrm{Tr}[\sigma_1]=0} \max_{\mathrm{Tr}[\sigma_2]=0} \frac{\mathrm{Tr}\Big[\sigma_1 \mathcal{E}^L(\sigma_2)\Big]}{\|\sigma_1\|_\infty \|\sigma_2\|_1}, \\
&\geq 1 - d^{5/2} \max_{\mathrm{Tr}[\sigma_2]=0} \frac{\Big\|\mathcal{E}^L(\sigma_2)\Big\|_1}{\|\sigma_2\|_1}, \\
&\geq 1 - d^{5/2} \tau(\mathcal{E})^L.
\end{aligned}$$

The upper bound can be achieved analogously. □

We now show that it is possible to upper bound the expectation value of an Hermitian operator with its 1-to-1 norm. The transfer matrix of an operator supported on r sites is denoted as E_O^r, and the isomorphic channel as $\mathcal{E}_O^r$.

Lemma 12. *If $\mathcal{E}$ is a trace-preserving completely positive map and $\mathcal{E}_O^r$ is an operator channel supported on r sites, then*

$$\frac{\mathrm{Tr}\Big[E_O^r E^{L-r}\Big]}{\mathrm{Tr}[E^L]} \leq \|E_O^r\|_{1\to 1} \frac{1 + d^{5/2}\tau(\mathcal{E})^{L-r}}{1 - d^{5/2}\tau(\mathcal{E})^L}$$

and by setting $r = 1$ and $E_O = A, B, C, D$ we find the local fidelity w.r.t $|\phi_1\rangle, |\phi_2\rangle, |\phi_3\rangle, |\phi_4\rangle$ respectively.

Proof.

$$\mathrm{Tr}\Big[E_O^r E^{L-r}\Big] = \left\langle \frac{\mathbb{1}}{\sqrt{d}} \middle| E_O^r E^{L-r} \middle| \frac{\mathbb{1}}{\sqrt{d}} \right\rangle + \sum_{\sigma \in P/\mathbb{1}} \langle \sigma | E_O^r E_n^{L-r} | \sigma \rangle$$

As before, we switch to the channel picture to deal with the first term

$$\begin{aligned}\left\langle \frac{\mathbb{1}}{\sqrt{d}}\middle| E_O^r E_n^{L_n-r}\middle|\frac{\mathbb{1}}{\sqrt{d}}\right\rangle &= \frac{1}{d}\operatorname{Tr}\left[\mathcal{E}_O^r\mathcal{E}^{L-r}(\mathbb{1})\right] \\ &= \frac{\operatorname{Tr}\left[\mathcal{E}_O^r\mathcal{E}^{L-r}(\mathbb{1})\right]}{\operatorname{Tr}[\mathbb{1}]}, \\ &\leq \left\|\mathcal{E}_O^r\mathcal{E}^{L-1}\right\|_{1\to 1}, \\ &\leq \|\mathcal{E}_O^r\|_{1\to 1}\left\|\mathcal{E}^{L-1}\right\|_{1\to 1} = \|\mathcal{E}_O^r\|_{1\to 1}.\end{aligned}$$

The second term will be handled accordingly.

$$\begin{aligned}\sum_{\sigma\in P/\mathbb{1}}\langle\sigma|E_O^r E_n^{Lr}|\sigma\rangle &\leq d^2\max_{\operatorname{Tr}[\sigma]=0}\langle\sigma|E_O^r E_n^{L-1}|\sigma\rangle \\ &= d^2\max_{\operatorname{Tr}[\sigma]=0}\frac{\operatorname{Tr}\left[\sigma\mathcal{E}_O^r\mathcal{E}_n^{L_n-1}(\sigma)\right]}{\|\sigma\|_2^2} \\ &= d^2\max_{\operatorname{Tr}[\sigma]=0}\frac{\operatorname{Tr}\left[\sigma\mathcal{E}_O^r\mathcal{E}_n^{L_n-1}(\sigma)\right]}{\|\sigma\|_2^2} \\ &\leq d^2\max_{\operatorname{Tr}[\sigma_1]=0}\max_{\operatorname{Tr}[\sigma_2]=0}\frac{\operatorname{Tr}\left[\sigma_1\mathcal{E}_O^r\mathcal{E}_n^{L_n-1}(\sigma_2)\right]}{\|\sigma_1\|_\infty\|\sigma_2\|_2}\end{aligned}$$

In the last step we made again use of $\|\sigma\|_\infty \leq \|\sigma_2\|_2$.

$$\begin{aligned}\sum_{\sigma\in P/\mathbb{1}}\langle\sigma|E_O^r E_n^{Lr}|\sigma\rangle &\leq d^2\max_{\operatorname{Tr}[\sigma_2]=0}\frac{\left\|\mathcal{E}_O^r\mathcal{E}_n^{L_n-1}(\sigma_2)\right\|_1}{\|\sigma_2\|_2} \\ &= d^2\sqrt{d}\max_{\operatorname{Tr}[\sigma_2]=0}\frac{\left\|\mathcal{E}_O^r\mathcal{E}_n^{L_n-1}(\sigma_2)\right\|_1}{\left\|\mathcal{E}_n^{L_n-1}(\sigma_2)\right\|_1}\frac{\left\|\mathcal{E}_n^{L_n-1}(\sigma_2)\right\|_1}{\|\sigma_2\|_1} \\ &= d^{5/2}\|\mathcal{E}_O^r\|_{1\to 1}\max_{\operatorname{Tr}[\sigma_2]=0}\frac{\left\|\mathcal{E}_n^{L_n-1}(\sigma_2)\right\|_1}{\|\sigma_2\|_1} \\ &= d^{5/2}\|\mathcal{E}_O^r\|_{1\to 1}\tau(\mathcal{E})^{L-r}\end{aligned}$$

For the denominator we use the upper bound from Lem. 11 and this completes the proof. □

Alas, lower bounding in the same fashion is not generally possible without knowing more about the eigenvectors of E_O^r. Lem. 12 tells us that upper bounds for physical expectation values can easily be derived by looking at the channel norms, given a long chain and that the transfer matrix is trace-preserving. This raises the question whether the transfer matrix remains trace-preserving over the course of the iteration.

This is only the case for deterministic protocols without postselection since no matrices are disregarded and thus the transfer matrix $\mathcal{E}_n$ is a power of the previous transfer matrix. Consequently, the proof of a threshold was much more straight forward in the deterministic case.

B. Code

B.1. Bounding Gaussianity using Convex Optimisation

The following code contains the minimisation of the relative entropy of a given measurement set using the symplectic diagonalisation of the given covariance matrix. It uses the modules "momentsfromtof.py", which calculates the covariance matrix for given experimental TOF measurement outcomes, and "minimizerelent.py", which contains functions needed in the minimisation procedure.

minimizegauss.py

"minimizegauss.py" transforms given measurement data using the symplectic diagonalisation of the covariance matrix γ. It then performs convex optimisation on the dual function of the minimisation problem in the single modes.

```
# -*- coding: utf-8 -*-
import numpy
#import scipy
#import math
from numpy import *
from numpy.linalg import *
#from scipy.optimize import fmin
import minimizerelent as minimize
import momentsfromtof as getmoments
#import cvxopt, cvxopt.solvers
#import pylab as p
import sys
import os.path

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
# Simulated data contains b†b (2mom) and b†bb†b (4mom)
h = int(sys.argv[3])
mode = int(sys.argv[2])
bosonic_mom = numpy.load('/net/data/janina/npy/input/bosonic_moments_L'
        +str(mode)+'_h'+str(h)+'.npy').item()
num = int(sys.argv[1])
```

```
path = '/net/data/janina/npy/'+str(mode)+'mode/fast/h_'+str(h)+'/'

length = int(sys.argv[3])
assert(length<=mode)

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
#for free bosons: [num] exists from 0 to 9.
for length in range(1,mode):
  append = '_'+str(num)+'_'+str(length)+'_'+str(h)+'.npy'
  if os.path.exists(path+'M2m'+append) and  os.path.exists(path+'M4m'+append):
    M2m = numpy.load(path+'M2m'+append)
    M4m = numpy.load(path+'M4m'+append)
    print 'loading old data'
  elif length < mode:
    M2m = bosonic_mom[num]['2mom']
    M4m_array = bosonic_mom[num]['4mom']
    print('Taking cut from the middle...')
    a = mode-length
    M2m = M2m[a/2:a/2+length,a/2:a/2+length]
    M4m_array = M4m_array[a/2:a/2+length,a/2:a/2+length,a/2:a/2+length,a/2:a/2+length]
                #If needed, calculate missing entries
    #######M4m = getmoments.getM4fromnumerics(M2m,M4m_array)
    #save data
    numpy.save(path+'M2m'+append, M2m)
    numpy.save(path+'M4m'+append, M4m_array)
  else:
    M2m = bosonic_mom[num]['2mom']
    M4m_array = bosonic_mom[num]['4mom']
    #calculate data, if needed
    ######M4m = getmoments.getM4fromnumerics(M2m,M4m_array)
    #save data
    numpy.save(path+'M2m'+append, M2m)
    numpy.save(path+'M4m'+append, M4m_array)

  gamma = minimize.covfromdata(M2m)

  #%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

  # Symplectic Diagonalization of the Gaussian State
  S,eta,gammadiag = minimize.symplectic(gamma)
  numpy.save('npy/'+str(mode)+'mode/eta_'+str(num)+'.npy',eta)
  # if wanted, check if S is symplectic:
  minimize.checkS(S)

  #Transform moments in the symplectic basis
  if 0: #os.path.exists(path+'M2'+append) and  os.path.exists(path+'M2'+append):
```

```
68      M2 = numpy.load(path+'M2'+append)
69      M4 = numpy.load(path+'M4'+append)
70      print 'loading old data'
71    else:
72      #Calculation (For more complicated systems)
73                  #this will take a while for big N (40*2.8 min for N=40)
74      #####M2t = minimize.getM2(M2m,S)
75      #####M4t = minimize.getM4eff(M4m,S)
76      #####numpy.save(path+'M2t'+append,M2t)
77      #####numpy.save(path+'M4t'+append,M4t)
78
79      #Calculation for simulated data
80      M2, U = numpy.linalg.eigh(M2m)
81      perm = argsort(-M2)
82      M2 = M2[perm]
83      U = U[:,perm]
84      M4 = minimize.getM4fast(M4m_array,U)
85
86      #this was effort. save for later use.
87      numpy.save(path+'M2'+append,M2)
88      numpy.save(path+'M4'+append,M4)
89
90    #make calculation on smaller part, where M2, M4>0
91    ind = M2>10**-4 #basically empty mode
92    eta = 2*M2 + 1
93    etas = eta[ind]
94    M2s = M2[ind]
95    M4s = M4[ind]
96
97
98    #now calculate the minimized entropy
99    x0 = array([1.0,M2[0],M4[0]])
100   min_entropy,xopt = minimize.minrelent(M2s,M4s,etas,x0)
101   dualentropy = minimize.dualentropy(M2,eta)
102
103   numpy.save(path+'entropies'+append.format(40),
104                 {'RelEnt':-sum(min_entropy) + sum(dualentropy),
105                  'MinEnt':min_entropy,
106                  'GaussEnt':dualentropy})
107   print length
```

momentsfromtof.py

The following code contains the functions and routines needed to transform given moments from the measurement basis to the basis in which the minimisation of the dual

function in performed.

```
# -*- coding: utf-8 -*-
import numpy
import time
from numpy import linalg
import matplotlib
import math
from matplotlib import pyplot as p

#----------------------Don't write over this line-------------------------------

def energie(datall):
  """Do everything one could want to do with this kind of data."""
  atomn,atomnred,noise,datsmall=sortdata(datall)
  convav,avconv,convdat,corr=getmoments.fouriercorr(datsmall)
  adaav, ada, avada = getmoments.getM2fromexp(datsmall)
  M4 = getmoments.getM4fromexp(ada)
  return avada, corr

# Sort experimental data for atom number
def sortdata(datall,NX1=170,NX2=200,NY1=50,NY2=75,X1=103,X2=333,Y1=77,Y2=307):
  """We integrate over a small noise area to reduce noise in the picture. then
  we integrate over a certain section of the image, to get the atom number of
  each image. index contains the pictures, where the atom number is between
  Nmin and Nmax. For further calculations, the subpicture should be squared.

  Parameters
  ----------
  NX1,NX2;NY1,NY2: noise area
  X1,X2;Y1,Y2: cloud area
  Nmin,Nmax: accepted particle range
  outputs atomnumber for different slices in datall and smaller data array with
  particle number in Nmin,Nmax array
  """
  factor=numpy.divide(float((X2-X1)*(Y2-Y1)),float((NX2-NX1)*(NY2-NY1)))
  noise = numpy.sum(datall[:,NX1:NX2,NY1:NY2],axis=1)
  noise = numpy.sum(noise,axis=1)
  atomn = numpy.sum(datall[:,X1:X2,Y1:Y2],axis=1)
  atomn = numpy.sum(atomn,axis=1)
  atomnred = atomn - factor*noise
  n,bins,patch=p.hist(atomnred)
  Nmin=bins[n.argmax()-1]
  Nmax=bins[n.argmax()+2]
  datsmall=datall[(atomnred>=Nmin)&(atomnred<=Nmax),X1:X2,Y1:Y2]
  return atomn,atomnred,noise,datsmall
```

```
45
46  def fouriercorr(datsmall): #,X1=110,X2=327,Y1=77,Y2=307):
47    """Calculate fourth moments (quasimomentum space) from time-of-flight
48    pictures using convolution via fft of the data.
49
50    Parameters
51    ----------
52    X1,X2;Y1,Y2: boundary on data subset to use (already done by sortdata)
53    datall: complete data set
54    outputs convoluted average of data, average of convoluted data,
55    convoluted data, correlationfunction
56    todo: output moments in format from minimizing code; include prefactors
57    """
58  #  datsmall = datall[:,X1:X2,Y1:Y2]
59    N = datsmall.shape[0]
60    avdat = numpy.mean(datsmall,axis=0)
61    convav = numpy.fft.fftshift(numpy.fft.ifft2(numpy.square(
62      numpy.abs(numpy.fft.fft2(avdat)))))
63    X = convav.shape[0]
64    Y = convav.shape[1]
65    convdat = numpy.zeros([N,X,Y])
66    for i in range(N):
67      convdat[i] = numpy.fft.fftshift(numpy.fft.ifft2(numpy.square(
68        numpy.abs(numpy.fft.fft2(datsmall[i])))))
69
70    avconv = numpy.mean(convdat,axis=0)
71    corr=numpy.divide(avconv,convav)
72    corr=corr.real
73    p.pcolor(corr-1,vmin=0.0004,vmax=0.0016)
74    p.colorbar()
75    p.show()
76    return convav,avconv,convdat,corr
77
78  def picto1D(datsmall,N=113): #oder 115
79    '''only use snippet from 2D picture and average over it. Use the middle.
80
81    Parameters
82    ----------
83    datsmall: array with TOF data, different pictures.
84    N: middle of picture in the second direction (y)
85    only use one period of k (one peak + surrounding at 115 \pm 43
86    '''
87    section = datsmall[:,72:158,N-3:N+3]
88    #this should find the mean in y-direction and keep individual pictures.
89    flatpic = numpy.mean(section, axis=2)
90    return flatpic
```

```
91
92  def getM2fromexp(datsmall,N):
93    #[2D:X1=103, X2=333, Y1=77 , Y2=307 (squared); 1D:230]
94    """Calculate second moments and single shots from time of flight pictures.
95    output: adaav, ada, avada (last one is M2m = <a_i^\dag a> (d) with d=(i-j))
96    todo: fix prefactors; trace(ada) should be around 40 but is tiny.
97    careful: we assume translational invariance here, not invertible otherwise.
98    lattice 1D, makes factor easier, data achievable, if not yet present.
99    """
100   #some constants we need because this is real data.
101   al = 425 * 10**(-9) # m lattice constant (\lambda/2)
102   px = 2.43 * 10**(-6) # m pixelsize
103   hbar = 1.054571726 * 10**(-34) # m^2 kg / s
104   t = 17 * 10**(-3) # s
105   m = 1.44316082 * 10**(-25) # kg (Rb87: 86.9091835 u)
106
107   nq1 = picto1D(datsmall) #average over middle section
108   #need to adjust this to fit the k-space picture.
109   nq = nq1/px #only 1D, no correction for r->k
110
111   n = nq.shape[1] #1D
112 #  N = 15    #lattice dimension (1D) (now user input)
113
114   P = numpy.zeros(N) #double counting for same distance between lattice points
115   L = numpy.zeros([n,N]) #phasefactor in transformation
116
117   #1D formulation: calculate P and L s.t. nq = L * P * M2
118   A = m*px*al/(hbar*t) #prefactor in the exponent
119   for k in range(N):
120     P[k] = (N-k)
121     for l in range(n):
122       L[l][k] = 2 * math.cos(A * k*l)
123
124    #no double counting while inverting indices!
125   P[0] = 0.5*P[0]
126
127   #Pseudo invert L to solve for M2
128   Linv = numpy.linalg.pinv(L)
129
130   #get single shot values for M2
131   nl = numpy.zeros([nq.shape[0],N])
132   for i in range(nq.shape[0]):
133     nl[i,:] = 1./P * numpy.dot(Linv , nq[i,:])
134
135   #get average = expectation value of M2, for covariance matrix
136   nqav = numpy.mean(nq,axis=0)
```

```
137   nlav = 1./P * numpy.dot(Linv,nqav) # 1/P * multiplies P[i] to dot(L,n)[i]
138
139   #build a^\dag a covariance matrix with translational and inversion invariance
140   #first averaged version, needed for covariance matrix
141   avada = numpy.diag(nlav[0]*numpy.ones(N),0)
142   for i in range(1,N):
143     avada += numpy.diag(nlav[i]*numpy.ones(N-i),i)
144     avada += numpy.diag(nlav[i]*numpy.ones(N-i),-i)
145
146   #then also single-shot version, later needed for fourth moments
147   #CAREFUL: pictures now in last dimension.
148   ada = numpy.zeros([N,N, nl.shape[0]])
149   for j in range(nl.shape[0]):
150     ada[:,:,j] += numpy.diag(nl[j,0]*numpy.ones(N),0)
151     for i in range(1,N):
152       ada[:,:,j] += numpy.diag(nl[j,i]*numpy.ones(N-i),i)
153       ada[:,:,j] += numpy.diag(nl[j,i]*numpy.ones(N-i),-i)
154
155   #Prefactor: W^d/(2pi) * e^(m/(h*t))
156 #  hbar = 1.054571726 * 10**(-34) # m^2 kg / s
157 #  t = 17 * 10**(-3) # s
158 #  m = 1.44316082 * 10**(-25) # kg (Rb87: 86.9091835 u)
159   wt = 250 # 1/s Trap frequency
160 #  al = 425 * 10**(-9) # m lattice constant (\lambda/2)
161   Er = hbar**2*numpy.pi**2/(al**2*2*m) #m**2 kg /s**2 Energy. good.
162 #  J = 1.39666*16**1.051*math.e**(-2.12104 * 4) * Er # Er (really that small?)
163 #  a01 = math.sqrt(math.sqrt(2*J)*al/(wt*m))# m**2  width of wannier state
164   V0 = 16*Er # depth of potential
165   who = math.sqrt(Er*V0)/hbar # 1/s frequency of HO
166   a0 = math.sqrt(hbar/(m*who)) # m wannier width
167 #  px = 2.43 * 10**(-6) # m pixelsize
168   W=(hbar*t)/(a0*m) # m nq has unit 1/m
169
170   #add prefactor to matrices
171   ada = W * ada
172   adaav = numpy.mean(ada, axis=0)
173   avada = W * avada
174
175   return adaav, ada, avada
176
177 def kd(i,j):
178   """Kronecker delta used in commutation relations for M4 calculation."""
179   if i==j:
180     return 1
181   else:
182     return 0
```

```
183
184 def getM4fromexp(ada):
185   """calculate all the M4 expectation values from the ada single shot values.
186   All entries with nonequal number of ad/a are zero
187
188   Parameters
189   ----------
190   ada: single shot values from TOF measurement, in lattice space.
191   """
192   t0 = time.clock() # this can run long; show time as entertainment
193   N = ada.shape[0] # lattice dimension
194   M = ada.shape[2] # number of pictures
195   M4 = numpy.zeros(16*N**4) # number of different combinations of a/ad
196
197   #find all entries through commutation relations
198   for i in range(N):
199     for j in range(N):
200       for k in range(N):
201         for l in range(N):
202           #I   a a ad ad
203           M4[8*N**3*i + 4*N**2*j + 2*N*(k+N) + (l+N)] = \
204               numpy.mean(ada[k,i] * ada[l,j] +
205               kd(j,k) * ada[l,i] + kd(j,l) * ada[k,i] + kd(i,k) * ada[l,j] +
206               kd(i,l) * kd(j,k) + kd(i,k) * kd(j,l) )
207           #II  a ad a ad
208           M4[8*N**3*i + 4*N**2*(j+N) + 2*N*k + (l+N)] = \
209               numpy.mean(ada[j,i] * ada[l,k] +
210               kd(i,j) * ada[l,k] + kd(k,l) * ada[j,i] + kd(k,l) * kd(i,j) )
211           #III a ad ad a
212           M4[8*N**3*i + 4*N**2*(j+N) + 2*N*(k+N) + l] = \
213               numpy.mean(ada[j,i] * ada[k,l] + kd(i,j) * ada[k,l] )
214           #IV  ad a a ad
215           M4[8*N**3*(i+N) + 4*N**2*j + 2*N*k + (l+N)] = \
216               numpy.mean(ada[i,j] * ada[l,k] + kd(k,l) * ada[i,j] )
217           #V   ad a ad a
218           M4[8*N**3*(i+N) + 4*N**2*j + 2*N*(k+N) + l] = \
219               numpy.mean(ada[i,j] * ada[k,l])
220           #VI  ad ad a a
221           M4[8*N**3*(i+N) + 4*N**2*(j+N) + 2*N*k + l] = \
222                   numpy.mean(ada[i,k] * ada[j,l] - kd(j,k) * ada[i,l] )
223     print(str(i)+' of '+str(N)+'; current runtime: '+
224         str(round((time.clock()-t0)/60.,3))+' min')
225   print('Total runtime: '+str(round((time.clock()-t0)/60.,3))+' min')
226   return M4
227
228 def getM4fromnumerics_ferm(ada,M4):
```

```
    """calculate all the M4 expectation values from the adaada and ada exp values.
    All entries with nonequal number of ad/a are zero
    Fermionic: Use anticommutator instead of commutator.

    Parameters
    ----------
    ada: single shot values from TOF measurement, in lattice space.
    """
    t0 = time.clock() # this can run long; show time as entertainment
    N = ada.shape[0] # lattice dimension
    M4m = 1j*numpy.zeros(16*N**4) # number of different combinations of a/ad

    #find all entries through commutation relations
    for i in range(N):
      for j in range(N):
        for k in range(N):
          for l in range(N):
            #I   a a ad ad
            M4m[8*N**3*i + 4*N**2*j + 2*N*(k+N) + (l+N)] = \
                M4[k,i,l,j] -  kd(j,k) * ada[l,i] - kd(j,l) * ada[k,i] - \
                kd(i,k) * ada[l,j] - kd(i,l) * kd(j,k) - kd(i,k) * kd(j,l)
            #II  a ad a ad
            M4m[8*N**3*i + 4*N**2*(j+N) + 2*N*k + (l+N)] = \
                M4[j,i,l,k] - \
                kd(i,j) * ada[l,k] - kd(k,l) * ada[j,i] - kd(k,l) * kd(i,j)
            #III a ad ad a
            M4m[8*N**3*i + 4*N**2*(j+N) + 2*N*(k+N) + l] = \
                M4[j,i,k,l] - kd(i,j) * ada[k,l]
            #IV  ad a a ad
            M4m[8*N**3*(i+N) + 4*N**2*j + 2*N*k + (l+N)] = \
                M4[i,j,l,k] - kd(k,l) * ada[i,j]
            #V   ad a ad a
            M4m[8*N**3*(i+N) + 4*N**2*j + 2*N*(k+N) + l] = \
                M4[i,j,k,l]
            #VI  ad ad a a
            M4m[8*N**3*(i+N) + 4*N**2*(j+N) + 2*N*k + l] = \
                 M4[i,k,j,l] + kd(j,k) * ada[i,l]
      #print(str(i)+' of '+str(N)+'; current runtime: '+
      #    str(round((time.clock()-t0)/60.,3))+' min')
    #print('Total runtime: '+str(round((time.clock()-t0)/60.,3))+' min')
    return M4m


def getM4fromnumerics(ada,M4):
    """calculate all the M4 expectation values from the adaada and ada exp values.
    All entries with nonequal number of ad/a are zero
```

```
275
276   Parameters
277   ----------
278   ada: single shot values from TOF measurement, in lattice space.
279   """
280   t0 = time.clock() # this can run long; show time as entertainment
281   N = ada.shape[0] # lattice dimension
282   M4m = 1j*numpy.zeros(16*N**4) # number of different combinations of a/ad
283
284   #find all entries through commutation relations
285   for i in range(N):
286     for j in range(N):
287       for k in range(N):
288         for l in range(N):
289           #I   a a ad ad
290           M4m[8*N**3*i + 4*N**2*j + 2*N*(k+N) + (l+N)] = \
291               M4[k,i,l,j] +  kd(j,k) * ada[l,i] + kd(j,l) * ada[k,i] + \
292               kd(i,k) * ada[l,j] + kd(i,l) * kd(j,k) + kd(i,k) * kd(j,l)
293           #II  a ad a ad
294           M4m[8*N**3*i + 4*N**2*(j+N) + 2*N*k + (l+N)] = \
295               M4[j,i,l,k] + \
296               kd(i,j) * ada[l,k] + kd(k,l) * ada[j,i] + kd(k,l) * kd(i,j)
297           #III a ad ad a
298           M4m[8*N**3*i + 4*N**2*(j+N) + 2*N*(k+N) + l] = \
299               M4[j,i,k,l] + kd(i,j) * ada[k,l]
300           #IV  ad a a ad
301           M4m[8*N**3*(i+N) + 4*N**2*j + 2*N*k + (l+N)] = \
302               M4[i,j,l,k] + kd(k,l) * ada[i,j]
303           #V   ad a ad a
304           M4m[8*N**3*(i+N) + 4*N**2*j + 2*N*(k+N) + l] = \
305               M4[i,j,k,l]
306           #VI  ad ad a a
307           M4m[8*N**3*(i+N) + 4*N**2*(j+N) + 2*N*k + l] = \
308                M4[i,k,j,l] - kd(j,k) * ada[i,l]
309     print(str(i)+' of '+str(N)+'; current runtime: '+
310         str(round((time.clock()-t0)/60.,3))+' min')
311   print('Total runtime: '+str(round((time.clock()-t0)/60.,3))+' min')
312   return M4m
```

minimizerelent.py

The following code contains the functions and routines needed to perform the symplectic diagonalisation and minimisation.

```
1  # -*- coding: utf-8 -*-
2  import math
```

```
import scipy
import numpy
import time
from numpy import *
from numpy.linalg import *
from scipy.optimize import fmin
#import cvxopt, cvxopt.solvers

#----------------------Don't write over this line------------------------------

#-------Second Part: Reading and Preparation of Data-----------
# Transform measured moments in Symplectic base z -> Sz; M -> SMS.T
# Linear mapping between creation/annihilation and x/p base q=Tr
def rtoq(N):
  '''builds transform from a^\dag a to x p basis (and back)
  (a, a^\dag) = T * (x, p)
  output: T, Tinv'''
  n = zeros((N,N))
  I = eye(N,N)
  T = 1./sqrt(2) * vstack([ hstack([ I, I*1j ]), hstack([ I, -I*1j ]) ])
  Tinv = 1./sqrt(2) * vstack([ hstack([ I, I ]), hstack([ -I*1j, I*1j ]) ])
  return T,Tinv

def covfromdata_ferm(M2m):
  '''takes fermionic a_i^\dag a_j data and returns covariance matrix.'''
  N = M2m.shape[0]
  antidiag = numpy.array([[0,-1],[1,0]])
  gamma = kron(2*M2m - eye(N) ,antidiag)
  #Find transformation on the level of (a,a^\dag)
  return gamma

#data means: a_i^d a_j - data.
def covfromdata(M2m):
  '''takes a_i^\dag a_j data and transforms covariance matrix in x/p basis
  output: gamma'''
  N = M2m.shape[0]
  gammaq = numpy.vstack([numpy.hstack([numpy.zeros([N,N]),M2m
    + numpy.eye(N)]),numpy.hstack([M2m,numpy.zeros([N,N])])])
  T,Tinv = rtoq(N)
  gammax = dot(dot(Tinv, gammaq), Tinv.T)
  gamma = 2*real(gammax)
  return gamma

def kd(i,j):
        """Kronecker delta used in commutation relations for M4 calculation."""
        if i==j:
```

```
                return 1
        else:
                return 0

def orthogonal(gamma):
        N = gamma.shape[0]/2
        [eta,S] = eigh(1j*gamma) #gamma is skew-symmetric. 1j*gamma is hermitian.
        perm = argsort(-eta)[:N]
        eta = 1j*eta[perm]
        S = S[:,perm]
        gammadiag = dot(dot(S.T.conjugate(),1j*gamma),S)
        return S,eta,gammadiag

def symplectic(gamma):
  '''finds the symplectic transform that diagonalizes gamma;
  output: S, eta, diagonalized gamma'''
  N = gamma.shape[0]/2
  n = zeros((N,N))
  I = eye(N,N)
  J = vstack([hstack([n,I]),hstack([-I,n])])
  [eta, W] = eig(-1j * dot(gamma, J))
  eta = eta.real
  perm = argsort(-eta)[:N]
  eta = eta[perm]
  Ws = -sqrt(2) * 1j * dot(J, W[:, perm])
  X = Ws.real
  Y = Ws.imag
  S = hstack([X, Y])
  gammadiag = dot(dot(S.transpose(), gamma), S)
  return S,eta,gammadiag

def checkS(S):
  '''check whether S really is symplectic; mostly for debugging'''
  N = S.shape[0]/2
  n = zeros((N,N))
  I = eye(N,N)
  J = vstack([hstack([n,I]),hstack([-I,n])])
  A = S[:N,:N]
  B = S[N:,:N]
  C = S[:N,N:]
  D = S[N:,N:]
  check1 = dot(A.transpose(), D) - dot(C.transpose(), B)
  check2 = dot(A, D.transpose()) - dot(B, C.transpose())
  check3 = dot(dot(S.transpose(), J), S)
  check4 = dot(dot(S, J), S.transpose())
  check = max(abs(check1 - I).max(), abs(check2 - I).max(),
```

```
95                    abs(check3 - J).max(), abs(check4 - J).max())
96    if check<10**(-10):
97      print("S is symplectic.")
98    else:
99      print("S is not symplectic.")
100   return check
101
102 #-------Fourth Part: Minimization of relative Entropy-----------
103
104 #----Minimizing - Definitions-----
105
106 def state(mu0, mu2, mu4, N):
107     N_array=arange(N)
108     return e**(-1.0*(1.0+mu0+mu2*N_array+mu4*(N_array**2)))
109
110 def g(x,M0,M2,M4,N):
111     mu0, mu2, mu4 = x
112     st=state(mu0,mu2,mu4,N)
113     gval=sum(st)+mu0*M0+mu2*M2+mu4*M4
114     return gval
115
116 def gradg(x,M0,M2,M4,N):
117     mu0, mu2, mu4 = x
118     st=state(mu0,mu2,mu4,N)
119     dgdmu0=M0-sum(st)
120     dgdmu2=M2-sum(st*arange(N))
121     dgdmu4=M4-sum(st*arange(N)**2)
122     return array([dgdmu0,dgdmu2,dgdmu4])
123
124 def check_constraints(opt, M0, M2, M4, N):
125     err0=sum(state(opt[0],opt[1],opt[2],N))-M0
126     err2=sum(state(opt[0],opt[1],opt[2],N)*arange(N))-M2
127     err4=sum(state(opt[0],opt[1],opt[2],N)*arange(N)**2)-M4
128     return array([err0,err2,err4])
129
130 def compute_optval(x,N):
131     st=state(x[0],x[1],x[2],N)
132     ind=(st>10E-15) #cut off in fock basis
133     optval=sum(st[ind]*log(st[ind]))
134     return optval
135
136 def compute_optimal_x_brutal(x0,M0,M2,M4,N,tol=10e-3):
137     """Compute initial guess for optimal point of objective function
138        in the middle of the intervall of possible values for M4 """
139     xstart=1.0*x0
140     x_opt=zeros((len(M2), 3))
```

```
141     """Compute a better initial guess for the first element M2[0] using
142        simplex downhill. This is then used as an initial guess for the actual
143        computation using the fmin_bfgs algorithm. """
144     x0=scipy.optimize.fmin(g,xstart,args=(M0,M2[0],M4[0],N),disp=False)
145     for i in arange(len(M2)):
146          print('\n\nMinimizing mode '+str(i+1)+' of '+str(len(M2))+'\n')
147          if M4[i]>M2[i]**2:
148              opt=scipy.optimize.fmin_bfgs(g,x0,fprime=gradg,\
149                  args=(M0,M2[i],M4[i],N),full_output=True,disp=False)
150              err=linalg.norm(check_constraints(opt[0],M0,M2[i],M4[i],N))
151              """Automatically check if the algorithm converges."""
152              if opt[6]!=0 or err>=tol:
153                  x0 = xstart
154                  print("Initial guess didn't work, try again...")
155                  #x0=array([M0,1/M2[i],1/M4[i]])
156                  opt=scipy.optimize.fmin_bfgs(g,x0,fprime=gradg,\
157                   args=(M0,M2[i],M4[i],N),full_output=True,disp=False)
158                  err=linalg.norm(check_constraints(opt[0],M0,M2[i],M4[i],N))
159                  if err>=tol:
160                      n=0
161                      print("...and keep on trying...")
162                      while True:
163                          x0 = random.rand(3)
164                          opt=scipy.optimize.fmin_bfgs(g,x0,fprime=gradg,\
165                         args=(M0,M2[i],M4[i],N),full_output=True,disp=False)
166                          err=linalg.norm(check_constraints(opt[0],M0,M2[i],M4[i],N))
167                          n = n+1
168                          if err<=tol or n>600:
169                              if err>tol:
170                                  print("Failed to find minimum! Current error: "
171                                      +str(err)+'\nMode '+str(i)+' of '+str(len(M2)-1))
172                                  print("M2 "+str(M2[i])+" and M4 "+str(M4[i])+
173                                      " will not contribute to the bound.")
174                              else:
175                                  print("Stage 3 converged.")# Current error: "+str(err))
176                                  x_opt[i,:]=opt[0]
177                              break
178 #                 print("Current xopt: "+str(opt[0]))
179 #                 print("Unable to find minimum for given M2, M4, set min_ent to zero")
180                  else:
181                      print("Stage 2 converged.")# Current error: "+str(err))
182                      x_opt[i,:]=opt[0]
183              else:
184                  print("Stage 1 converged.")# Current error: "+str(err))
185                  x_opt[i,:]=opt[0]
186          else:
```

```
187             print("M2 and M4 incompatible. Bound 0.")
188     return x_opt
189
190
191 def compute_optimal_x(x0,M0,M2,M4,N,tol=10e-3):
192     """Compute initial guess for optimal point of objective function
193        in the middle of the intervall of possible values for M4 """
194     xstart=1.0*x0
195     x_opt=zeros((len(M2), 3))
196     """Compute a better initial guess for the first element M2[0] using
197        simplex downhill. This is then used as an initial guess for the actual
198        computation using the fmin_bfgs algorithm. """
199     x0=scipy.optimize.fmin(g,xstart,args=(M0,M2[0],M4[0],N))
200     for i in arange(len(M2)):
201        opt=scipy.optimize.fmin_bfgs(g,x0,fprime=gradg,\
202                                     args=(M0,M2[i],M4[i],N),full_output=True)
203        err=linalg.norm(check_constraints(opt[0],M0,M2[i],M4[i],N))
204        """Automatically check if the algorithm converges."""
205        if opt[6]!=0 or err>=tol:
206           x0 = xstart
207           #x0=array([M0,1/M2[i],1/M4[i]])
208           opt=scipy.optimize.fmin_bfgs(g,x0,fprime=gradg,\
209                                        args=(M0,M2[i],M4[i],N),full_output=True)
210           err=linalg.norm(check_constraints(opt[0],M0,M2[i],M4[i],N))
211           if err>=tol:
212              #raise Exception("Unable to find minimum for given M2, M4")
213              print("Current error: "+str(err))
214              print("Current xopt: "+str(opt[0]))
215              print("Unable to find minimum for given M2, M4, set min_ent to zero")
216              x_opt[i,:]=0*opt[0]
217           else:
218              x_opt[i,:]=opt[0]
219        else:
220           x_opt[i,:]=opt[0]
221     return x_opt
222
223 def minrelent(M2,M4,eta,x0=array([1.,0.1,0.1]),M0=1.0,it=1000):
224   beta = 2./eta #Normalize, since gamma = 2* real(r_i r_j)
225   x_opt = []
226   g_opt = []
227   constraints = []
228   dualentropy = []
229   f_opt = []
230   #it = 600 #counting treshold for summation over fock space
231   x_opt = compute_optimal_x_brutal(x0,M0,M2,M4,it,tol=10e-3)
232   for k in range(M2.shape[0]):
```

```
233                 if x_opt[k].all() == 0:
234                     g_opt.append( 0 )
235                 else:
236                     g_opt.append( compute_optval(x_opt[k],it))
237   min_entropy = - real(g_opt) #entropy of rho
238   return min_entropy,x_opt
239
240 def dualentropy(M2,eta):
241   beta = 2./eta
242   dualentropy=( beta * M2 - log(1.0 - e**(-beta))) #entropy of sigma
243   return dualentropy
244
245 def gaussian_entropy(eta):
246   mu = 1./(eta)
247   gaussian_ent=numpy.zeros(len(eta))
248   for i in range(len(eta)):
249             if abs(mu[i]-1)>10**-3:
250     gaussian_ent[i]=(((1-mu[i])/(2*mu[i])) * math.log((1+mu[i])/(1-mu[i]))
251                     - math.log((2*mu[i])/(1+mu[i])))
252             else:
253                 gaussian_ent[i] = 0
254   return array(gaussian_ent)
255
256 #---------Minimizing - Calculation----------------
257 # compute all minimized entropies for M2, M4
258 def getM2(M2m,S):
259   N = M2m.shape[0]
260   T,Tinv = rtoq(N)
261   ST = dot(dot(T,S.T),Tinv)
262   gammaq = numpy.vstack([numpy.hstack([numpy.zeros([N,N]),M2m
263     + numpy.eye(N)]),numpy.hstack([M2m,numpy.zeros([N,N])])])
264   M2t = dot(dot(ST, gammaq), ST.T.conjugate())
265 #  M2t = dot(dot(ST.T.conjugate(), gammaq), ST)
266 #  M2t = dot(dot(dot(dot(T,S.T),gamma),S),T.T) #equivalent, but already real
267   M2 = real(diag(M2t[N:,:N]))
268   return M2,M2t
269
270 def getM2_ferm(M2m,S): #Not right yet. Need trf from ada to cov. Replace T by this trf.
271   N = M2m.shape[0]
272   T,Tinv = rtoq(N)
273   ST = dot(dot(T,S.T),Tinv)
274   gammaq = numpy.vstack([numpy.hstack([numpy.zeros([N,N]),-M2m
275     + numpy.eye(N)]),numpy.hstack([M2m,numpy.zeros([N,N])])])
276   M2t = dot(dot(ST, gammaq), ST.T)
277 #  M2t = dot(dot(dot(dot(T,S.T),gamma),S),T.T) #equivalent, but already real
278   M2 = real(diag(M2t[N:,:N]))
```

```
279   return M2
280
281 def getM4fast(M4m_array,U):
282   N = M4m_array.shape[0]
283   Ud = U.T.conjugate()
284   Ut = U.T
285   M4 = numpy.zeros(M4m_array.shape[0])
286   M4m_array = numpy.tensordot(Ud,M4m_array,(1,0))
287   M4m_array = numpy.tensordot(Ut,M4m_array,(1,1)).transpose((1,0,2,3))
288   M4m_array = numpy.tensordot(Ud,M4m_array,(1,2)).transpose((1,2,0,3))
289   M4m_array = numpy.tensordot(Ut,M4m_array,(1,3)).transpose((1,2,3,0))
290   for i in range(M4m_array.shape[0]):
291     M4[i] = real(M4m_array[i,i,i,i])
292   return M4
293
294 def getM4eff(M4m_flat,S): #runtime: 40*2.9 min for N=40
295   t0 = time.clock()
296   N = int(0.5*sqrt(sqrt(M4m_flat.shape[0])))
297   T,Tinv = rtoq(N)
298   ST = dot(dot(T,S.T),Tinv)
299   M4t = (1+1j)*numpy.zeros([N])
300   for z in range(N):
301     ST4it = (1+1j)*numpy.zeros([M4m_flat.shape[0]])
302     for i in range(2*N):
303       for j in range(2*N):
304         for k in range(2*N):
305           for l in range(2*N):
306             ST4it[(2*N)**3*i + (2*N)**2*j + 2*N*k + l] = \
307               ST[z+N,i] * ST[z,j] * ST[z+N,k] * ST[z,l]
308     M4t[z] = dot(ST4it, M4m_flat)
309     print(str(z)+' of '+str(N-1)+'; current runtime: '+
310         str(round((time.clock()-t0)/60.,3))+' min')
311   M4 = real(M4t)
312   print('Total runtime: '+str(round((time.clock()-t0)/60.,3))+' min')
313   return M4
314
315 def getM4_ferm(M4m_flat,S): #not right yet.
316   t0 = time.clock()
317   N = int(0.5*sqrt(sqrt(M4m_flat.shape[0])))
318   T,Tinv = rtoq(N) #Replace with the right T
319   ST = dot(dot(T,S.T),Tinv)
320   M4t = (1+1j)*numpy.zeros([N])
321   for z in range(N):
322     ST4it = (1+1j)*numpy.zeros([M4m_flat.shape[0]])
323     for i in range(2*N):
324       for j in range(2*N):
```

```
325         for k in range(2*N):
326           for l in range(2*N):
327             ST4it[(2*N)**3*i + (2*N)**2*j + 2*N*k + l] = \
328               ST[z+N,i] * ST[z,j] * ST[z+N,k] * ST[z,l]
329     M4t[z] = dot(ST4it, M4m_flat)
330     print(str(z)+' of '+str(N-1)+'; current runtime: '+
331         str(round((time.clock()-t0)/60.,3))+' min')
332   M4 = real(M4t)
333   print('Total runtime: '+str(round((time.clock()-t0)/60.,3))+' min')
334   return M4
335
336 def getM4evenmoreeff(M4m_flat,S): #restrict to columns where M4m != 0
337   t0 = time.clock()
338   N = int(0.5*sqrt(sqrt(M4m_flat.shape[0])))
339   T,Tinv = rtoq(N)
340   ST = dot(dot(Tinv.T,S),T.T)
341   M4t = (1+1j)*numpy.zeros([N])
342   for z in range(N):
343     ST4it = (1+1j)*numpy.zeros([M4m_flat.shape[0]])
344     for i in range(N):
345       for j in range(N):
346         for k in range(N):
347           for l in range(N):
348             ST4it[8*N**3*i + 4*N**2*j + 2*N*(k+N) + (l+N)] = \
349                 ST[i,z+N] * ST[j,z] * ST[k+N,z+N] * ST[l+N,z]
350             ST4it[8*N**3*i + 4*N**2*(j+N) + 2*N*k + (l+N)] = \
351                 ST[i,z+N] * ST[j+N,z] * ST[k,z+N] * ST[l+N,z]
352             ST4it[8*N**3*i + 4*N**2*(j+N) + 2*N*(k+N) + l] = \
353                 ST[i,z+N] * ST[j+N,z] * ST[k+N,z+N] * ST[l,z]
354             ST4it[8*N**3*(i+N) + 4*N**2*j + 2*N*k + (l+N)] = \
355                 ST[i+N,z+N] * ST[j,z] * ST[k,z+N] * ST[l+N,z]
356             ST4it[8*N**3*(i+N) + 4*N**2*j + 2*N*(k+N) + l] = \
357                 ST[i+N,z+N] * ST[j,z] * ST[k+N,z+N] * ST[l,z]
358             ST4it[8*N**3*(i+N) + 4*N**2*(j+N) + 2*N*k + l] = \
359                 ST[i+N,z+N] * ST[j+N,z] * ST[k,z+N] * ST[l,z]
360     M4t[z] = dot(ST4it, M4m_flat)
361     print(str(z)+' of '+str(N-1)+'; current runtime: '+
362         str(round((time.clock()-t0)/60.,3))+' min')
363   M4 = real(M4t)
364   print('Total runtime: '+str(round((time.clock()-t0)/60.,3))+' min')
365   return M4
```

B.2. Non-linear Tomography in Optical Lattices

The following code contains the code used to simulate the measurement setup for the reconstruction of atom positions in an optical lattice. The reconstruction is done for real as well as simulated data ("Data/Simulated_reconstruction.m"). The Fourier approach is contained in "Atomic_reconstruction.m". The used subfunctions "Lmatrix.m", "Measurement.m" and "Reconstruct.m" are added in following.

Execute.m

In the following, "Execute.m" is called to choose one of several data arrays containing different atom configurations. The pictures are then prepared for reconstruction and later reconstructed, when "Data_reconstruction.m" is called as the main subroutine. Possible errors are investigated (empty pictures and missing peaks) and the reconstructions are compared to the reconstructions performed by the experimental group using a two way reconstruction process.

```
%% Load Data

%load Data
load('New Data/BenchmarkImageDataset.mat') % already contains LSF

%Process pictures with few atoms
%PicN = size(images_2atoms_d_1site);
%PicN = size(images_2atoms_d_2sites);
PicN = size(images_2atoms_d_25sites); %

%PicN = size(sortingSummary.imageStack);  %9Images

I = zeros(PicN(1),PicN(3));
for i = 1:PicN(1)
    % results from Bonn (fitted_positions) for d_1 and d_2
    %I(i,:) = squeeze(mean(images_2atoms_d_1site(i,:,:),2));
    I(i,:) = squeeze(mean(images_2atoms_d_25sites(i,:,:),2));
    %I(i,:) = squeeze(mean(images_2atoms_d_2sites(i,:,:),2));
end

%% Initialize
%Initialize LSF
[Lbin,LL2] = Lmatrix(LSF,length(I(1,:)));
'Lbin initialized.' %#ok<NOPTS>
s = length(LSF(:,1)); % Datapoints in LSF (in superresolution)
t = floor(abs(-LSF(1)+LSF(s))); % Number of pixels covered by LSF

```

```
28 % %Initialize arrays and fill them
29 %Height of background noise
30 epsilon = 2.5*std(I(30,1:50)); % 60 for Benchmark, 20 for old pictures
31
32 %half width of LSF
33 w = floor(0.5*t)+7; % 20 for Benchmark, 50 for old LSF
34
35 Ip = zeros(PicN(1),PicN(3)); %processed picture (baseline removed)
36 Iroi = cell(1,PicN(1)); %ROIS of each picture
37 rois = cell(1,PicN(1)); %positions of rois of each picture
38 eps = zeros(1,PicN(1)); %noise std in each picture
39 err = zeros(1,PicN(1)); %peaks at borders. 1 if there's a peak.
40
41 for i = 1:PicN(1)
42     [Ip(i,:),Iroi{i}, rois{i}, eps(i), err(i)] = PreparePictures(I(i,:),epsilon,w);
43 end
44 'Pictures processed.'
45
46
47 %% Reconstruction (And some information it needs.)
48
49
50 NoP = 10; %Number of Pictures to evaluate
51 NoA = 2; %Expected Number of Atoms
52
53 % xx contains the reconstructed amplitudes at superresolution positions,
54 % Sri contains the Source function in the resolution of Iroi
55 % Iroi contains the Region of interest containing the peak
56 % errorbar contains the errorbar estimated for this region
57 % usablePic is a index set of Pictures with the same number of atoms (No
58 % more half atoms..) on which the reconstruction is performed
59 % samedist is a index set of indices that contain atoms with the same
60 % distance in the lattice, up to errorbars.
61 [xx,Sri,Iroi,errorbar, usablePic, recPic, fitted_pos, dist, samedist] =
62   Data_reconstruction(PicN, Iroi, rois, 2*eps, Lbin, s, t, NoA, NoP);
63
64 {'Reconstruction success in ',length(recPic), 'of' NoP, 'pictures'}
65 NoPs = length(recPic);
66 %% further Information  (mostly useful for benchmarking for large distances)
67 %Information on Errorbars:
68 {'Errorbars calculated. Mean:',
69 mean(errorbar(usablePic(find(errorbar(recPic(1:NoPs))<10))))}
70
71 %Something went wrong:
72 if length(find(errorbar(recPic(1:NoPs))>10))>0
73     'Problems (Errorbar>10) in pictures:' %#ok<NOPTS>
```

```
74      find(errorbar(:)>10)
75  end
76
77  %Which pictures contain atoms with the same distance in the lattice
78  {'Main distance', mean([dist(samedist)]), length(samedist), 'of' NoPs}
79
80
81  %lattice distance in pixel:  %from earlier estimations: about 0.68
82  DoA = 30; %attempted distance
83  {'Estimated lattice distance:', DoA/mean([dist(samedist)])} %pixel/lattice site
84
85  %% Compare Reconstructions; (Only for d_1 site and d_2 sites)
86  % If reconstructions of the data arrays exist, we can compare them to ours.
87  load('New Data/results_d_1site.mat')
88  %load('New Data/results_d_2sites.mat')
89
90  compare = zeros(1,PicN(1));
91  for i = recPic(1:NoPs)
92          for k = 1:length(fitted_pos{i})
93              dist_rec{i}=fitted_pos{i}(k)-fitted_positions{i}{1};
94             if any(abs(fitted_pos{i}(k)-fitted_positions{i}{1}) <= errorbar(i))
95                  compare(i) = 1;
96             else
97                  compare(i) = 0;
98                  'reconstruction not in errorbars in picture ' %#ok<NOPTS>
99                  i%#ok<NOPTS>
100            end
101         end
102 end
103
104 'Pictures reconstructed in accordance to Experimentalists:'
105 {sum(compare(recPic)) 'of' length(recPic)}
106 %'of'
107 if length(find(compare(recPic(1:NoPs))<1))>0
108 'Reconstruction failed in pictures: (position, deviation, errorbar)'
109 end
110 for i =[find(compare(recPic(1:NoPs))<1)]
111     for j = 1:length(fitted_pos{i})
112         err = max(abs(fitted_pos{i}(j) - fitted_positions{i}{1}));
113     end
114     {i err errorbar(i)}
115 end
116
117 %% Visual output for certain picture
118 % rather translate Sri to Si (without region of interest offset) and show
119 % reconstruction in the whole picture. (could also do that with fitted_pos)
```

```
120 i=20;
121 j=1;
122
123 plot(1:length(Sri{i}{j}),Sri{i}{j},1:length(Iroi{i}{j}),Iroi{i}{j})
124 'reconstructed positions:' %#ok<NOPTS>
125 fitted_pos{i} %#ok<NOPTS>
126
127 load('New Data/results_d_1site.mat')
128 %load('New Data/results_d_2sites.mat')
129
130 %'reconstructed positions from Bonn:' %#ok<NOPTS>
131 fitted_positions{i}{1} %#ok<NOPTS>
132 fitted_positions_discrete{i}{1}
```

Data_reconstruction.m

In this subfunction, the pictures are preselected according to their number of atoms and number of reconstructed peaks, in order to get rid of faulty pictures. The reconstruction is then performed on all usable pictures. Afterwards, the mean peak distance of the reconstructed atoms is evaluated to benchmark the reconstructed lattice distance. To do that, the median of reconstructed positions within the error bar is used.

```
1  %% Reconstruct from Data array and LSF
2  function [xx,Sri,Iroi,errorbar, usablePic, recPic, fitted_pos, dist, samedist] =
3  Data_reconstruction(PicN, Iroi, rois, eps, Lbin, s, t, NoA, NoP)
4
5  %Determine Normalization and Pictures with right number of Atoms
6  [usablePic, normalization] = peakheight(Iroi, Lbin, NoA);
7  tr = .01*normalization;   %threshold for xx to be taken into account.
8
9  %Reconstruct
10 xx = cell(1,PicN(1));  %cell containing reconstructed positions in ROI.
11 Sri = cell(1,PicN(1)); %cell containing reconstructed source distributions.
12 for i = usablePic(1:NoP)
13     [xx{i}, Sri{i}] = reconstruct_data(normalization, s,t, tr, Iroi{i},
14                                        rois{i}, eps(i), Lbin);
15 end
16
17 errorbar = Errorbar(Lbin, eps, Iroi, usablePic, NoP);
18 errorbar = errorbar*t/s; %scale down to pixelsize
19
20 %Find indices where reconstruction failed
21 recPic = [];
22 for i = usablePic(1:NoP)
23     for j = length(rois{i})
```

```
24          if sum(xx{i}{j}>0)>0
25              recPic = [recPic, i];
26          end
27      end
28  end
29
30  NoP = min(NoP,length(recPic));
31
32  fitted_pos = cell(1,length(rois));
33  dist = zeros(1,length(rois));
34  for i = recPic(1:NoP)
35      [fitted_pos{i}, dist(i)] = Fit_Positions(rois{i}, xx{i}, t, s, tr);
36  end
37
38  samedist = Index_samedist(dist, recPic, NoP, errorbar);
39  end
40
41
42  %% Single particle peak height
43  %Figures out the height of one atom. This only works, if a set of pictures
44  %with mostly one-atom-peaks exists or if the atom number that is present in
45  %most of the pictures is known. No sophisticated step function evaluation
46  %yet, as there are many cases where the atoms seem to leave the lattice.
47  %(Compare boxplot)
48  function [usablePic, normalization] = peakheight(Iroi,Lbin, NoA)
49  %If no sorting is going to be needed, set
50  %usablePic = [1:PicN(1)];
51
52  PicN(1) = length(Iroi);
53  particleN = zeros(1,PicN(1));
54
55  for i = 1:PicN(1)
56      particleN(i) = sum([Iroi{i}{:}]);
57      %particleN(i) = sum(Ip(i,:));
58  end
59
60  boxplot(particleN);
61  %h = findobj(gcf,'tag','Box');
62  %upper = get(findobj(gcf,'tag','Upper Whisker'),'YData');
63  %lower = get(findobj(gcf,'tag','Lower Whisker'),'YData');
64  box = get(findobj(gcf,'tag','Box'),'YData');
65
66  usablePic = intersect(find(particleN>min(box)), find(particleN<max(box)));
67  %If number of atoms per picture are known:
68  %NoA = 2;
69  %normalization = (mean(box)/NoA)/max(Lbin(:,1));
```

```
normalization = (max(particleN(usablePic))/NoA)/sum(Lbin(:,floor(length(Lbin(1,:))/2)));

end

%% Reconstruct
function [xx, Sri] = reconstruct_data(normalization, s,t, tr, Iroi, rois, eps, Lbin)
top = normalization; %max(Iroi{1}{1})/max(Lbin(150,:))

for l = 1:length(rois)
    if isempty(rois{l})
        xx{l} = NaN;
        Sri{l} = NaN;
    else
        [xx{l}, Sri{l}] = Reconstruct(Iroi{l}, Lbin, 1*eps, tr, top, s,t);
    end
end
end

%% Errorbars
function errorbar = Errorbar(Lbin, eps, Iroi, usablePic, NoP)
line = floor(length(Lbin(:,1))/2);
[MaxVal,Index] = max(Lbin(line,:));
%Lbin(line,Max) is maximal. proceed going left and right from there.

errorbar = zeros(1,length(Iroi));
for j = usablePic(1:NoP)
    i = 0;
    while norm(Lbin(line,Index) - Lbin(line,Index+i),2)
        *max(Iroi{j}{1})/max(Lbin(line,:))
        <= eps(j)*sqrt(length(transpose(Iroi{j}{1})));
        if i>=(length(Lbin)-Index)
            break
        else
            i = i+1;
        end
    end
    errorbar(j) = i;
end

end

%% Make reconstructed positions comparable to other reconstruction.
% Gives positions and peak distance in pixels of the original image
function [fitted_pos, dist] = Fit_Positions(rois, xx, t, s, tr)

fitted_pos = [];
```

```
116 for j = 1:length(rois)
117     if length(rois{j})>0
118         fitted_pos = [fitted_pos; t/s*(find(xx{j}>tr)+0.5*s) + rois{j}(1) - 1];
119     end
120 end
121 dist = max(fitted_pos) - min(fitted_pos);
122 
123 %'Positions fitted.' %#ok<NOPTS>
124 end
125 
126 %% Mean Peak Distance
127 % estimate the lattice constant in pixels from the distance that occures
128 % mostly in the pictures with some attempted atom distance in lattice sites.
129 
130 function samedist = Index_samedist(dist, usablePic, NoP, errorbar)
131 %get rid of empty matrices
132 Ind = find(abs([dist(usablePic(1:NoP))] - mean([dist(usablePic(1:NoP))]))
133   < 1.5*std([dist(usablePic(1:NoP))]));
134 %boxplot([dist{usablePic(1:NoP)}]);
135 boxplot([dist(usablePic(Ind))]);
136 box2 = get(findobj(gcf,'tag','Box'),'YData');
137 
138 samedist = usablePic(intersect(find([dist(usablePic(1:NoP))]>
139  (min(box2)-mean(errorbar(usablePic(1:NoP)))))), find([dist(usablePic(1:NoP))]
140  <(max(box2)+mean(errorbar(usablePic(1:NoP)))))));
141 
142 end
```

Simulation_reconstruction.m

The following code uses the LSF from experimental data to simulate the fluorescence measurement on an optical lattice. It takes wanted resolution, sparsity to simulate, noise level and possible amplitude range as input values. The positions can be chosen at random or can be set to specific values to benchmark the reconstruction behaviour in critical situations. After calculating a noiseless picture using the LSF, noise is added. This noisy picture afterwards undergoes the same reconstruction as described in the reconstruction of experimental data (Reconstruct.m). Afterwards, error bars are estimated and the reconstruction is compared with the real positions used in the simulation.

```
1 %% Simulate the measurement of atoms in an optical lattice
2 load('New Data/BenchmarkImageDataset.mat') % already contains LSF
3 s = length(LSF(:,1));
4 t = floor(abs(-LSF(1)+LSF(s)));
5 
```

```
n = 313;
Lbin = Lmatrix(LSF,n); %Benchmark
'Lbin initialized.' %#ok<NOPTS>

n = 2;                %sparsity to simulate.
eps = 0.1;            %Noise level.
at = 0.99;            %Threshold value for lowest amplitudes

%pos = randi(N,n,1);

pos = [200,207];
[ak, Xk, I_sim] = Measurement(n, 1, eps, at, Lbin, pos, 0,1);

w = floor(0.5*t)+1;  %LSF width
epsilon = 2.5*std(I_sim(1:10)); %Noise content
[Ip_sim, Iroi_sim, rois_sim, eps_sim, err_sim] = PreparePictures(100*I_sim', epsilon, w);

xx_sim = cell(1,length(rois_sim));
Sri_sim = cell(1,length(rois_sim));
fitted_pos_sim = cell(1,length(rois_sim));

%Plot:
%plot(1:length(Iroi_sim{1}),Iroi_sim{1})

%% Reconstruct

%eps_sim = 0.5;
for i = 1:length(rois_sim)
    [xx_sim{i}, Sri_sim{i}] = Reconstruct(Iroi_sim{i}, Lbin, eps_sim, 0.5, 100*at, s, t);
    %[xx_sim{i}, Sri_sim{i}] = ITReconstruct(Iroi_sim{i}, Lbin, 0.2*eps_sim, 0.5, 10000);
    %fitted_pos_sim{i} = 1.4*(0.1*find(xx_sim{i}>0.3) + rois_sim{i}(1)-1);
    fitted_pos_sim{i} =find(xx_sim{i}>0.3);% + 10*rois_sim{i}(1)-10; %
end

%Errorbar
line = floor(length(Lbin(:,1))/2);
[MaxVal,Index] = max(Lbin(line,:));

i= 0;
while norm(Lbin(line,Index) - Lbin(line,Index+i),2)* max(Iroi_sim{1})/max(Lbin(line,:))
  <= eps_sim*sqrt(length(transpose(Iroi_sim{1})));
    if i >= (length(Lbin)-Index)
        break
    else
        i = i+1;
    end
```

```
end
errorbar = i;

errorbar = errorbar*t/s; %scale down to pixelsize

'Results: Real positions vs. fitted positions.'
Xk - floor(s/t)*(rois_sim{1}(1)-1), transpose(fitted_pos_sim{:}), errorbar

%% compare picture with reconstruction
a = length(Iroi_sim{1});
b = min(floor(s/t)*a,max(floor(s/t),floor(s/t)*a - s));
Lbinr = Lbin(1:a,1:b);
Irec = Lbinr*xx_sim{1};
plot(1:length(Iroi_sim{1}),Iroi_sim{1},1:length(Irec),Irec)
```

Atomic_reconstruction.m

In this approach, the Fourier approach for reconstructing atom positions in fluorescence microscopy is implemented. First, the measurement is simulated, where the measurement plane is shifted to the Fourier plane of the image. Discretisation and convolution with the LSF are added in this approach (see Measurement.m). Subsequently, the Fourier transform of the LSF is used to deconvolve the signal and afterwards estimate the frequencies using the atomic norm approach. As described in the main text, the current LSF is not useful for this approach, as it only cuts out a very thin band of the signal which is insufficient to reconstruct frequencies. However, the approach is added here, as it is the natural starting point for implementing filtering with coded diffraction patterns.

```
%% Reconstruction of Atom positions using atomic norm

% 1. Compute DFT of baseline corrected image
% 2. Compute DFT of all shifted versions of the binned LSF
% 3. Divide result of (1) by all results of (2), producing a set of I_fc
% (fourier+cleaned)

%% Simulate Pictures with measurement in fourier plane

[ak, Xk, Data] = Measurement(n, A, e, at, Lbin, LL2, LSF, pos, map,d);

n = 1;                %sparsity to simulate.
eps_sim1 = 0.1;  %Noise level.
at = 0.9;             %Threshold value for lowest amplitudes
pos = [547];
[ak, Xk, I_sim] = Simulate(n, eps_sim1, at, Lbin, pos);
[Ip_sim, Iroi_sim, rois_sim, eps_sim, err_sim] = PreparePictures(10000*I_sim');
```

```
Ip = Ip_sim;

%% Get rid of LSF contribution in pictures. (1-3)
% Ip: Picture containing LSF convoluted with delta
N = size(Ip);
M = size(LL2);
%M = size(Lbin);

%Initialization
If = cell(1,N(1));
Ifc = cell(N(1),10);
Ic = cell(N(1),10);
Iip = zeros(N(1),M(2));
Lf = zeros(10,M(2));

% 2.
for i = 1:10
    Lf(i,:) = fft(LL2(M(1)/2+i,:));
    %Lf(i,:) = fft(Lbin(M(1)/2+i,:));
end

%Interpolate and fourier transform the signal (1.)
for i = 1:N(1)
    Iip(i,:) = interp1([1:N(2)],Ip(i,:),linspace(1,N(2),M(2)));
    If{i} = fft(Iip(i,:));
    % Get rid of LSF contribution (3.)
    for j = 1:10
        Ifc{i,j} = If{i}./Lf(j,:);  %Picture containing cleaned fourier
        %Lfi = Lf(j,:)\1;
        %Ifc{i,j} = If{i}.*Lfi';
        Ic{i,j} = ifft(Ifc{i,j}); %Picture containing cleaned version
    end
end
plot(Ic{1,1})


%%
%6. ...

%load('2peaks.mat')
%load('Lbin2.mat')
N = size(Ip);
M = size(Lbin);
Lf = fft(Lbin(M(1)/2+1,:));
Iip = interp1([1:N(2)],Ip,linspace(1,N(2),M(2)));
If = fft(Iip);
```

```
64
65  epsilon = eps;
66  for j = 1:M(2)
67      if abs(Lf)>=epsilon
68          Ifc(j) = If(j)/Lf(j);
69      else
70          Ifc(j) = If(j);
71      end
72  end
73
74  %Ifc = If./Lf;
75  Ic = ifft(Ifc);
76  figure, plot(real(Ic))
```

Lmatrix.m

The short routine transforms the array containing the LSF in superresolution into the matrix used in the linear mapping. Note that the matrix is not simply a Toeplitz matrix, as the border effects as well as the shift between lattice space and image space are crucial.

```
1   %% Line-spread function to matrix without boundary effects.
2   function [Lbin,LL2] = Lmatrix(LSF,n)
3   %n: (wanted) resolution of pictures
4
5   z = LSF(:,2);      % contains entries of LSF only
6   b = length(z);     % length of big L != 0
7   a = floor(b/(-LSF(1)+LSF(b)))*n;  % dim(LL)
8
9   [s1, c1] = min(abs(LSF(:,1))); %find entry corresponding to pixel 0.
10                                 %This will be on the diagonal of LL.
11  %[s2, c2] = sort(z,'descend'); %find the highest entry of the LSF -> put that to pixel 0.
12  %c2 = c2(1);
13
14  bigL = zeros(a,1);
15  bigL(a/2 -c1+1: a/2 + (b-c1)) = z; %maybe more physical, since it comes
16  %from the measurements.
17  %bigL(a/2 -c2+1: a/2 + (b-c2)) = z;%maximum in the middle of the array.
18
19  z1 = zeros(a,1);
20  z2 = flipud(z);                 %Reverse z in filling the lower diags of LL2
21  z1(1) = z2(1);                  %Make first entry equal to avoid warning in toeplitz
22  bigL2 = zeros(a,1);
23  bigL2(1:length(z2)) = z2;       %Build huge vector with reverse LSF
24
25  LL2 = zeros(a);
```

```
26 LL2 = toeplitz(bigL2,z1);        %Build a matrix with the LSF contribution on lower diag.
27 LL2 = LL2(:,1:a-b+1);            %cut off the part that would lead to incomplete peaks.
28
29 %LL2 will be (a)x(a - a/n*b) in size.
30
31 %% Bin down LSF: Lbin \in R^{nxm}
32
33 BN = zeros(n,a);
34 %How to bin:
35 bn = ones(a/n,1);
36
37 for i = 1:n
38     BN(i,a/n*(i-1)+1:a/n*i) = bn;
39 end
40
41 %Binned L:
42 Lbin = BN*LL2;
43
44 end
```

Measurement.m

In Measurement.m, three different kinds of measurement maps are implemented. The first map is the standard measurement, which simulates the measurement performed in the laboratory. It performs the convolution with the superresolution linespread function, adds noise and bins down to a specified resolution, simulating the detector behaviour. The second map simulates a moving camera, where a specified number of pictures are taken with the camera moving in subpixel distances. For the given LSF, the error bars are dominated by noise and not by the offset, so this approach does not improve the reconstruction. For different measurement maps however, this approach can help to reduce the error introduced by the discretisation through the detectors. The third approach is performing a measurement in the Fourier plane using a Fourier transform of the LSF. To get this Fourier transform, the LSF was fitted by a Gaussian and the Fourier transform was then calculated analytically. The measurement consists of a simple product of the discretised Fourier transforms of LSF and signal, where noise is added afterwards.

```
1 %% Simulation of Measurement using different Measurement Maps
2 function [ak, Xk,S, Data] = Measurement(n, A, e, at, Lbin, LL2, LSF, pos, map,d)
3 if map == 2
4     Res = length(Lbin(:,1));
5     [Dataft, Data,S, ak, Xk] = FourierM(Res,n,A,at,pos,e); %Xk are frequencies here
6 else
```

```
[ak, Xk, S] = SourceDist(n, A, at, Lbin, pos);
if map == 0
    Data = StandardM(e, S, Lbin); %
elseif map == 1
    n = length(Lbin(:,1));
    Data = MovingM(e, A, S, n, LL2, d); %Moving camera approach.
    %d: Number of lattice sites to move the camera
%elseif map == 2 %Does not exist yet.
%    Data = FourierM(e, S, LSF, pos);
end
end
end
%% Generate the undisturbed signal
function [ak, Xk, S] = SourceDist(n, A, at, Lbin, pos)
%Parameters to alter:
%n = 5;             %sparsity to simulate.
%e = 0.001;         %Noise level.
%at = 0.9;          %Threshold value for lowest amplitudes

%Fixed Parameters:
l = 1./1.4;             %lattice width in pixel.
Rs = length(Lbin(:,1)); %resolution of the picture.
%A = s/t;                %Expansion factor for S to have max. res. of LSF

bigRes = floor(length(Lbin));  %column dimension of LSF matrix.
N = floor(bigRes/(A*l));       %number of lattice spaces.

%Sampling:
ak = at + (1-at)*rand(n,1);    %Sample peak heights.
%Xk = randi(N,n,1);     %Sample n positions from 1..N
%(NOTE: samples with replacement, values can double.)
Xk = pos;
%Xk = randperm(N,n);    %Sample (uniquely) n positions from 1..N
a0 = 0;                 %Add stray light background.

%Calculate signal vector:
%Ss = ones(N,1)*a0;     %Signal in lattice resolution.
%Ss(Xk) = ak;           %Fill in sampled data.

S = ones(bigRes,1)*a0;  %Signal in max. res. of LSF
%S(floor(Xk*A)) = ak;   %Fill in sampled data.
S(Xk) = ak;
end

%% Calculate picture:
function Data = StandardM(e, S, Lbin)
```

```
%A = s/t;
Rs = length(Lbin(:,1)); %resolution of the picture.
Data = Lbin*S;        %Simulated picture (without noise)
%eps = e*randn(Rs,1);  %Compute noise.
eps = e*randn(Rs,1).*(1+log(1+Data));
%eps = zeros(Rs,1);
Data = Data + eps;    %Add noise.

%to be able to compare Data and Input:
%xsi = 0.1*(Xk*A+0.5*s);   %align sampled positions to image.
%Si = ones(Rs,1)*a0;
%Si(floor(xsi)) = 0.02*ak;  %scale by max(LSF)
end
%% Moving camera:
function Data = MovingM(e, A, S, n, LL2, d)
%Data = Lbin*S;        %Simulated picture (without noise)

%Do the following in a shifted way for shifts from 1:d. Concatenate the
%results.
Data = [];
for j = 0:d-1
BN = zeros(n,n*A);
bn = ones(A,1);
BN(1,1:A-d) = ones(A-d,1);
for i = 2:n
    BN(i,A*(i-1)-d+1:A*i-d) = bn;
end
%Binned L:
Lbin = BN*LL2; %*BM
Datad = Lbin*S;
Data = [Data; Datad];
end
%Add some noise
eps = e*randn(d*n,1);  %Compute noise.
Data = Data + eps;    %Add noise.
end
%% Fourier type measurement
function [Data,Dataclean,S, ak, Xk] = FourierM(Res,n,A,at,pos1,e)

%For some other function than LSF calculate fourier transform of peaks und
%multiply with fourier trf of lsf sample.

%build an artificial lsf
a = 0.1232; b = 0; c = 50; %This comes quite close to the real LSF with
%b=232 (in superresolution)
a = 0.1232; b = 0; c = 50/A;  %(no superresolution)
```

```
dim1 = 477; %477 is the length of the LSF(:)
%%x = linspace(-dim1/2,dim1/2,dim1)';
x = linspace(1,dim1,dim1)';
a = 0.1232; b = 0; c = 5;
y = a*exp(-((x-b)/c).^2); %Could use y as LSF(:,2)
%plot(y)

%fourier trf: dim2 should be choosen large to simulate binning.
dim2 = A*Res;
supp = pi;
w = linspace(-supp,supp,dim2)';
yft = a*c/sqrt(2) * exp(-(c*w).^2/4 + 1i*b*w);
%b is just the offset. It is problematic later.

%fourier trf of source fn (at * delta(x-pos))
S = zeros(dim2,1);

% Rescale positions to normal space (no superresolution)
pos = pos1/A;%positions

for i = 1:length(pos)
    S = S + exp(1i*pos(i)*w);
end

Xk = pos/(2*pi);
ak = at/sqrt(2*pi);
S = at/sqrt(2*pi) * S;

%use convolution theorem to compute the image
Dataft = S.*yft;

%Add noise (how big should it be?)
Databig = Dataft + e*randn(dim2,1);

%Bin to make the atomic norm thing harder. (Another error gets introduced)
BN = zeros(Res,dim2);

%How to bin:
bn = ones(A,1);
for i = 1:Res
    BN(i,A*(i-1)+1:A*i) = bn;
end

%Binned L:
Data = BN*Databig;
%only the real part will be measured!
```

```
145 Data = real(Data);
146
147 %compute the inverse of the binned fourier(LSF) to get rid of its
148 %contribution. This will be a problem we have to face.
149 yftbin = BN*yft;
150 %for current values: relevant data is in yftbin(130:185) (10**-2). Find
151 %these edges automatically for whenever yftbin<e
152 %start = 130;
153 %stop = 185;
154
155 factor = 0.2;
156 %yftinvcut = yftbin(start:stop)
157 yftinv = 0*yftbin;%(yftbin>factor*e);
158 yftinv(abs(real(yftbin))>factor*e) = 1./yftbin(abs(real(yftbin))>factor*e);
159 yftinv(abs(real(yftbin))<factor*e) = 0;
160
161 % analytic calculation
162 %yftinv = sqrt(2)/(a*c) * exp((c*w).^2/4 -1i*b*w);
163 %%%yftinv .* yft is equal to 1.
164 %yftinvbin = BN*yftinv;
165
166 Sbin = BN*S;
167
168 Dataclean = Data.*yftinv; %this is obviously not good yet.
169 t = linspace(0,Res,Res)';
170 plot(t,real(Data),t,real(Dataclean),t,real(Sbin))
171 center = floor(size(Dataclean)/2.);
172 center = center(1);
173 length = 15;
174 Dataclean = Dataclean(center-length:center+length+1);
175 plot(real(Dataclean))
176
177 w2 = linspace(-supp,supp,Res)';
178 plot(w2,Data);
179
180 %use the real lsf
181 f = fit(LSF(:,1), LSF(:,2),'fourier5');
182 coeff = coeffvalues(f);
183 FourierL = zeros(1,A);
184 N = length(coeff)
185
186 end
```

Reconstruct.m

In this last routine, atom positions are reconstructed from given arrays containing regions of interest and the binned down version of the experimental line spread function. It is a standard compressed sensing approach to find the optimal sparse estimate for the signal vector.

```
%% Reconstruction with nonsquared L (current)
%Data should contain regions of interest of whole pictures without any baseline.

function [xx, Sri] = Reconstruct(Iroi, Lbin, eps, tr, top, s,t)
%Data = Iroi; %Take Data from Image.
%Data from simulation: Run first part of Simulate.m.
Data = Iroi';
a = length(Data);
b = min(min(floor(s/t)*a,max(2*floor(s/t),floor(s/t)*a - s)),length(Lbin(1,:)));
Lbinr = Lbin(1:a,1:b); %adjust LSF to size of ROI

%Set noise level:
epsilon = sqrt(eps)*a;

cvx_begin %quiet

   variable xx(b); %complex

   minimize( norm(xx,1) );
   subject to
   norm(Data - Lbinr*xx,2) <= epsilon %Lbin: (a x b) Data: (a) xx: (b)
        xx <= top;
        xx >= 0; % lower bound on reconstructed amplitudes

cvx_end

find(xx>tr)
sort(floor(A*Xk))        %for comparison with simulated Xk.

xr = zeros(b,1);
xr([find(xx>tr)])=xx([find(xx>tr)]);  %write biggest values in new vector
I_rec = Lbinr*xr;  %Reconstructed Image.
norm(I_rec-Data,2)

%Find reconstructed source function to compare with data.
Sri = zeros(a,1);
Sri(floor(t/s*(find(xx>tr)+0.5*s))) = xx(find(xx>tr));
end
plot([1:length(Data)],Data,[1:length(Sri)],Sri)
```

Author's Contributions

[1] J. Gertis, M. Friesdorf, C. A. Riofrío and J. Eisert. 'Estimating strong correlations in optical lattices'. In: *Phys. Rev. A* 94 (5 2016), p. 053628.

[2] S. Waeldchen, J. Gertis, E. T. Campbell and J. Eisert. 'Renormalizing Entanglement Distillation'. In: *Physical Review Letters* 116.2 (2016).

Bibliography

[1] A. Ahlbrecht, F. Richter and R. F. Werner. 'How long can it take for a quantum channel to forget everything?' In: *International Journal of Quantum Information* 10 (2012), p. 1250057.

[2] A. Ahmed, B. Recht and J. Romberg. 'Blind Deconvolution Using Convex Programming'. In: (2012). ArXiv: `1211.5608`.

[3] A. Alberti, C. Robens, W. Alt et al. 'Super-resolution microscopy of single atoms in optical lattices'. In: *New Journal of Physics* 18.5 (2016), p. 053010.

[4] E. Altman, E. Demler and M. D. Lukin. 'Probing many-body states of ultra-cold atoms via noise correlations'. In: *Physical Review A* 70 (2004), p. 013603.

[5] R. Arnon-Friedman and R. Renner. 'De Finetti reductions for correlations'. In: *Journal of Mathematical Physics* 56.5 (2015), p. 052203. ArXiv: `1308.0312`.

[6] K. M. R. Audenaert and M. B. Plenio. 'When are correlations quantum?' In: *New Journal of Physics* 8 (2006), p. 266.

[7] C. H. Bennett and G. Brassard. 'Quantum cryptography: Public key distribution and coin tossing'. In: *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing* 175 (1984), p. 8.

[8] C. H. Bennett, G. Brassard, S. Popescu et al. 'Purification of Noisy Entanglement and Faithful Teleportation via Noisy Channels'. In: *Physical Review Letters* 76 (1996), p. 722.

[9] C. H. Bennett, G. Brassard, C. Crépeau et al. 'Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels'. In: *Physical Review Letters* 70.13 (1993), pp. 1895–1899.

[10] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin and W. K. Wootters. 'Mixed-state entanglement and quantum error correction'. In: *Physical Review A* 54.5 (1996), pp. 3824–3851.

[11] R. Bhatia. *Matrix analysis*. Springer, 1997.

[12] I. Bloch, J. Dalibard and S. Nascimbene. 'Quantum simulations with ultracold quantum gases'. In: *Nature Physics* 8 (2012), p. 267.

[13] I. Bloch, J. Dalibard and W. Zwerger. 'Many-body physics with ultracold gases'. In: *Reviews of Modern Physics* 80 (3 2008), pp. 885–964.

[14] T. Blumensath and M. E. Davies. 'Iterative Hard Thresholding for Compressed Sensing'. In: *Applied and Computational Harmonic Analysis* 27 (2009), p. 265.

[15] F. G. S. L. Brandao and J. Eisert. 'Correlated entanglement distillation and the structure of the set of undistillable states'. In: *Journal of Mathematical Physics* 49.4 (2008), p. 042102.

[16] S. Braun, M. Friesdorf, S. S. Hodgman et al. 'Emergence of coherence and the dynamics of quantum phase transitions'. In: *Proceedings of the National Academy of Sciences* 112 (2015), p. 3641.

[17] H. J. Briegel, W. Dür, J. I. Cirac and P. Zoller. 'Quantum Repeaters: The Role of Imperfect Local Operations in Quantum Communication'. In: *Physical Review Letters* 81 (1998), p. 5932.

[18] F. Buscemi and N. Datta. 'Distilling entanglement from arbitrary resources'. In: *Journal of Mathematical Physics* 51.10 (2010), p. 102201.

[19] E. J. Candes. 'The Restricted Isometry Property and its Implications for Compressed Sensing'. In: *Compte Rendus de l'Academie des Sciences* 346.Paris, Serie I (2008), pp. 589–592.

[20] E. J. Candes and T. Tao. 'Decoding by Linear Programming'. In: *IEEE Transactions on Information Theory* 51.12 (2005), pp. 4203–4215.

[21] A. C. Cassidy, C. W. Clark and M. Rigol. 'Generalized Thermalization in an Integrable Lattice System'. In: *Physical Review Letters* 106 (2011), p. 140405.

[22] D. Chen, M. White, C. Borries and B. DeMarco. 'Quantum Quench of an Atomic Mott Insulator'. In: *Physical Review Letters* 106 (23 2011), p. 235304.

[23] S. Chen, Y.-A. Chen, B. Zhao et al. 'A Robust Atom-Photon Entanglement Source for Quantum Repeaters'. In: *Physical Review Letters* 99.18 (2007). ArXiv: 0706.2327.

[24] L. Childress, J. M. Taylor, A. S. Sörensen and M. D. Lukin. 'Fault-Tolerant Quantum Communication Based on Solid-State Photon Emitters'. In: *Physical Review Letters* 96 (2006), p. 070504.

[25] J. F. Clauser and A. Shimony. 'Bell's theorem. Experimental tests and implications'. In: *Reports on Progress in Physics* 41.12 (1978), p. 1881.

[26] M. Cramer, A. Bernard, N. Fabbri et al. 'Spatial entanglement of bosons in optical lattices'. In: *Nature Communications* 4 (2013), p. 3161.

[27] M. Cramer, C. M. Dawson, J. Eisert and T. J. Osborne. 'Exact Relaxation in a Class of Nonequilibrium Quantum Lattice Systems'. In: *Physical Review Letters* (2008), p. 030602.

[28] M. Cramer and J. Eisert. 'A quantum central limit theorem for non-equilibrium systems: Exact local relaxation of correlated states'. In: *New Journal of Physics* 12 (2010), p. 055020.

[29] M. Cramer, M. B. Plenio, S. T. Flammia et al. 'Efficient quantum state tomography'. In: *Nature Communications* 1 (2010), p. 149.

[30] M. Curty, M. Lewenstein and N. Lütkenhaus. 'Entanglement as Precondition for Secure Quantum Key Distribution'. In: *Physical Review Letters* 92 (2004), p. 217903.

[31] D. Deutsch, A. Ekert, R. Jozsa et al. 'Quantum Privacy Amplification and the Security of Quantum Cryptography over Noisy Channels'. In: *Physical Review Letters* 77.13 (1996), pp. 2818–2821.

[32] S. J. Devitt, K. Nemoto and W. J. Munro. 'Quantum Error Correction for Beginners'. In: *Reports on Progress in Physics* 76.7 (2013), p. 076001. ArXiv: 0905.2794.

[33] D. L. Donoho. 'Compressed Sensing'. In: *IEEE Transactions on Information Theory* 52.4 (2006), pp. 1289–1306.

[34] L. M. Duan, M. D. Lukin, J. I. Cirac and P. Zoller. 'Long-distance quantum communication with atomic ensembles and linear optics'. In: *Nature* 414 (2001), p. 413.

[35] A. Einstein, B. Podolsky and N. Rosen. 'Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?' In: *Physical Review* 47.10 (1935), pp. 777–780.

[36] J. Eisert. 'Entanglement and tensor network states'. In: *Modeling and Simulation* 3 (2013), p. 520. ArXiv: 1308.3318.

[37] J. Eisert, M. Friesdorf and C. Gogolin. 'Quantum many-body systems out of equilibrium'. In: *Nature Phys* 11 (2 2015), pp. 124–130.

[38] J. Eisert, F. G. S. L. Brandao and K. M. R. Audenaert. 'Quantitative entanglement witnesses'. In: *New Journal of Physics* 9.3 (2007), pp. 46–46. ArXiv: quant-ph/0607167.

[39] M. Endres, M. Cheneau, T. Fukuhara et al. 'Single-site- and single-atom-resolved measurement of correlation functions'. In: (2013). ArXiv: 1303.5652.

[40] M. Fannes, B. Nachtergaele and R. F. Werner. 'Finitely correlated states on quantum spin chains'. In: *Communications in Mathematical Physics* 144 (1992), p. 443.

[41] S. Fölling. 'Quantum noise correlation experiments with ultra-cold atoms'. In: (2014). ArXiv: 1403.6842.

[42] S. Fölling, F. Gerbier, A. Widera et al. 'Spatial quantum noise interferometry in expanding ultra-cold atom clouds'. In: *Nature* 434 (2005), pp. 481–484.

[43] M. Fornasier and H. Rauhut. 'Compressive Sensing'. In: *Handbook of Mathematical Methods in Imaging*. Ed. by O. Scherzer. Springer, 2011, pp. 187–228.

[44] S. Foucart. 'A Note on Guaranteed Sparse Recovery via l1-Minimization'. In: *Applied and Computational Harmonic Analysis* (2010).

[45] C. W. Gardiner and P. Zoller. *The Quantum World of Ultra-Cold Atoms and Light*. London : Hackensack, New Jersey: Imperial College Press, 2015. 475 pp.

[46] M. G. Genoni and M. G. A. Paris. 'Quantifying non-Gaussianity for quantum information'. In: *Physical Review A* 82 (5 2010), p. 052341.

[47] M. G. Genoni, M. G. A. Paris and K. Banaszek. 'Quantifying the non-Gaussian character of a quantum state by quantum relative entropy'. In: *Physical Review A* 78 (6 2008), p. 060303.

[48] J. Gertis, M. Friesdorf, C. A. Riofrío and J. Eisert. 'Estimating strong correlations in optical lattices'. In: *Phys. Rev. A* 94 (5 2016), p. 053628.

[49] N. Gisin and R. Thew. 'Quantum communication'. In: *Nature Photonics* 1 (2007), p. 165.

[50] N. Gisin, G. Ribordy, W. Tittel and H. Zbinden. 'Quantum cryptography'. In: *Reviews of Modern Physics* 74.1 (2002), pp. 145–195.

[51] M. Gluza, C. Krumnow, M. Friesdorf, C. Gogolin and J. Eisert. 'Equilibration via Gaussification in fermionic lattice systems'. In: (2016). ArXiv: 1601.00671.

[52] D. Gottesman, H.-K. Lo, N. Lütkenhaus and J. Preskill. 'Security of quantum key distribution with imperfect devices'. In: *Quantum Information & Computation* 4 (2004), p. 325.

[53] D. Gross, F. Krahmer and R. Kueng. 'Improved Recovery Guarantees for Phase Retrieval from Coded Diffraction Patterns'. In: (2014). ArXiv: 1402.6286.

[54] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker and J. Eisert. 'Quantum state tomography via compressed sensing'. In: *Physical Review Letters* 105.15 (2010). ArXiv: 0909.3304.

[55] O. Gühne, M. Reimpell and R. F. Werner. 'Estimating Entanglement Measures in Experiments'. In: *Physical Review Letters* 98 (2007), p. 110502.

[56] L. Hartmann, B. Kraus, H.-J. Briegel and W. Dür. 'Role of memory errors in quantum repeaters'. In: *Physical Review A* 75 (2007), p. 032310.

[57] S. W. Hell and M. Kroug. 'Ground-state-depletion fluorscence microscopy: A concept for breaking the diffraction resolution limit'. In: *Applied Physics B* 60.5 (1995), pp. 495–497.

[58] R. Horodecki, P. Horodecki and M. Horodecki. 'Quantum α-entropy inequalities: independent condition for local realism?' In: *Physics Letters A* 210.6 (1996), pp. 377–381.

[59] L. Isserlis. 'On a Formula for the Product-Moment Coefficient of any Order of a Normal Frequency Distribution in any Number of Variables'. In: *Biometrika* 12.1 (1918), pp. 134–139.

[60] L. Jiang, J. M. Taylor, K. Nemoto et al. 'Quantum repeater with encoding'. In: *Physical Review A* 79 (2009), p. 032325.

[61] M. Karski. 'State-selective transport of single neutral atoms'. PhD thesis. 2010.

[62] M. Kliesch, D. Gross and J. Eisert. 'Matrix-Product Operators and States: NP-Hardness and Undecidability'. In: *Physical Review Letters* 113.16 (2014), p. 160503.

[63] D. Kretschmann and R. F. Werner. 'Quantum Channels with Memory'. In: *Physical Review A* 72 (2005), p. 062323.

[64] K. Li, L. Gan and C. Ling. 'Convolutional Compressed Sensing using Deterministic Sequences'. In: *IEEE Transactions on Signal Processing* 61.3 (2013), pp. 740–752.

[65] Y. Li, S. D. Barrett, T. M. Stace and S. C. Benjamin. 'Hybrid-system approach to fault-tolerant quantum communication'. In: *Physical Review A* 87.5 (2013).

[66] A. Mari, K. Kieling, B. M. Nielsen, E. Polzik and J. Eisert. 'Directly Estimating Nonclassicality'. In: *Physical Review Letters* 106 (2011), p. 010403.

[67] P. Marian and T. A. Marian. 'Relative entropy is an exact measure of non-Gaussianity'. In: *Physical Review A* 88 (1 2013), p. 012322.

[68] F. J. Narcowich. 'Geometry and uncertainty'. In: *Journal of Mathematical Physics* 31.2 (1990), pp. 354–364.

[69] J. V. Neumann. *Mathematical Foundations of Quantum Mechanics*. Princeton University Press, 1955. 462 pp.

[70] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. 1 edition. Cambridge ; New York: Cambridge University Press, 2000. 675 pp.

[71] M. Ohliger, V. Nesme and J. Eisert. 'Efficient and feasible state tomography of quantum many-body systems'. In: *New Journal of Physics* 15 (2013), p. 015024.

[72] M. Ohliger, V. Nesme, D. Gross, Y.-K. Liu and J. Eisert. 'Continuous-variable quantum compressed sensing'. In: (2011). ArXiv: `1111.0853`.

[73] R. Orus. 'A practical introduction to tensor networks: Matrix product states and projected entangled pair states'. In: *Annals of Physics* 349 (2014), p. 117.

[74] A. Peres. 'Separability Criterion for Density Matrices'. In: *Physical Review Letters* 77.8 (1996), pp. 1413–1415.

[75] D. Perez-Garcia, F. Verstraete, M. M. Wolf and J. I. Cirac. 'Matrix product state representations'. In: *Quantum Information & Computation* 5&6 (2006), p. 401.

[76] M. B. Plenio and S. Virmani. 'Spin Chains and Channels with Memory'. In: *Physical Review Letters* 99 (2007), p. 120504.

[77] R. Raussendorf and H. J. Briegel. 'A One-Way Quantum Computer'. In: *Physical Review Letters* 86.22 (2001), pp. 5188–5191.

[78] B. Recht, M. Fazel and P. A. Parrilo. 'Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization'. In: *SIAM Rev.* 52.3 (2010), pp. 471–501.

[79] R. Renner, N. Gisin and B. Kraus. 'An Information-Theoretic Security Proof for QKD Protocols'. In: *Physical Review A* 72 (2005), p. 012332.

[80] M. Rigol, V. Dunjko, V. Yurovsky and M. Olshanii. 'Relaxation in a Completely Integrable Many-Body Quantum System: An Ab Initio Study of the Dynamics of the Highly Excited States of 1D Lattice Hard-Core Bosons'. In: *Physical Review Letters* 98 (2007), p. 050405.

[81] J. P. Ronzheimer, M. Schreiber, S. Braun et al. 'Expansion Dynamics of Interacting Bosons in Homogeneous Lattices in One and Two Dimensions'. In: *Physical Review Letters* 110 (2013), p. 205301.

[82] R. Roth and K. Burnett. 'Superfluidity and interference pattern of ultracold bosons in optical lattices'. In: *Physical Review A* 67 (2003), p. 031602.

[83] C. Schoen, E. Solano, F. Verstraete, J. I. Cirac and M. M. Wolf. 'Sequential Generation of Entangled Multiqubit States'. In: *Physical Review Letters* 95 (2005), p. 110503.

[84] U. Schollwoeck. 'The density-matrix renormalization group in the age of matrix product states'. In: *Annals of Physics* 326 (2011), p. 96.

[85] M. Schreiber, S. S. Hodgman, P. Bordia et al. 'Observation of many-body localization of interacting fermions in a quasi-random optical lattice'. In: *Science* 349 (2015), p. 842.

[86] N. Schuch. 'Condensed Matter Applications of Entanglement Theory'. In: (2013). ArXiv: `1306.5551`.

[87] M. O. Scully and M. S. Zubairy. *Quantum Optics*. Cambridge University Press, 1997. 662 pp.

[88] C. E. Shannon. 'A Mathematical Theory of Communication'. In: *ACM SIGMOBILE Mobile Computing and Communications Review* 5.1 (2001), pp. 3–55.

[89] P. W. Shor and J. Preskill. 'Simple Proof of Security of the BB84 Quantum Key Distribution Protocol'. In: *Physical Review Letters* 85 (2000), p. 441.

[90] A. Steffens, M. Friesdorf, T. Langen et al. 'Towards experimental quantum field tomography with ultracold atoms'. In: *Nature Communications* 6 (2015), p. 7663.

[91] J. Stolze and D. Suter. *Quantum Computing, Revised and Enlarged*. 2 edition. Weinheim: Wiley-VCH, 2008. 282 pp.

[92] O. Szehr and M. M. Wolf. 'Perturbation bounds for quantum Markov processes and their fixed points'. In: *Journal of Mathematical Physics* 54.3 (2013), p. 032203.

[93] G. Tarozzi and A. Merwe. *Open Questions in Quantum Physics: Invited Papers on the Foundations of Microphysics*. Springer Science & Business Media, 2012. 422 pp.

[94] C. G. Timpson. 'Quantum Information Theory and the Foundations of Quantum Mechanics'. In: (2004). ArXiv: `quant-ph/0412063`.

[95] S. Trotzky, Y.-A. Chen, A. Flesch et al. 'Probing the relaxation towards equilibrium in an isolated strongly correlated one-dimensional Bose gas'. In: *Nature Physics* 8 (2012), p. 325.

[96] S. Trotzky, L. Pollet, F. Gerbier et al. 'Suppression of the critical temperature for superfluidity near the Mott transition: validating a quantum simulator'. In: *Nature Physics* 6 (2010), p. 998.

[97] V. Vedral. 'The role of relative entropy in quantum information theory'. In: *Reviews of Modern Physics* 74 (1 2002), pp. 197–234.

[98] F. Verstraete, J. I. Cirac, J. I. Latorre, E. Rico and M. M. Wolf. 'Renormalization group transformations on quantum states'. In: *Physical Review Letters* 94 (2005), p. 140601.

[99] F. Verstraete, J. J. García-Ripoll and J. I. Cirac. 'Matrix Product Density Operators: Simulation of Finite-Temperature and Dissipative Systems'. In: *Physical Review Letters* 93.20 (2004), p. 207204.

[100] S. Waeldchen, J. Gertis, E. T. Campbell and J. Eisert. 'Renormalizing Entanglement Distillation'. In: *Physical Review Letters* 116.2 (2016).

[101] R. F. Werner. 'Quantum Information Theory - an Invitation'. In: (2001). ArXiv: `quant-ph/0101061`.

[102] G. C. Wick. 'The Evaluation of the Collision Matrix'. In: *Physical Review* 80.2 (1950), pp. 268–272.

[103] J. Williamson. 'On the Algebraic Problem Concerning the Normal Forms of Linear Dynamical Systems'. In: *American Journal of Mathematics* 58.1 (1936), pp. 141–163.

[104] M. M. Wolf. *Quantum channels and operations – a guided tour*. 2012.

[105] W. K. Wootters and W. H. Zurek. 'A single quantum cannot be cloned'. In: *Nature* 299.5886 (1982), pp. 802–803.

[106] E. Zohar, J. I. Cirac and B. Reznik. 'Cold-Atom Quantum Simulator for SU(2) Yang-Mills Lattice Gauge Theory'. In: *Physical Review Letters* 110 (2013), p. 125304.

[107] M. Zwolak and G. Vidal. 'Mixed-State Dynamics in One-Dimensional Quantum Lattice Systems: A Time-Dependent Superoperator Renormalization Algorithm'. In: *Physical Review Letters* 93.20 (2004), p. 207205.