Exploring Non-Equilibrium in Ultracold Fermi Gases and Machine Learning in Physics

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Martin Link
aus
Oberwesel

Bonn 2021

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter:	Prof. Dr. Michael Köhl
2. Gutachter:	Prof. Dr. Stefan Linden
Tag der Promotion:	08. April 2021
Erscheinungsjahr:	2021

Abstract

This thesis covers two fields of research carried out in the context of ultracold Fermi gases: Non-equilibrium physics in the BEC-BCS crossover and the use of machine learning methods in physics.

The BEC-BCS crossover is of particular interest for the investigation of low-temperature superconductivity and related many-body phenomena. We realise the BEC-BCS crossover experimentally by cooling a mixture of two hyperfine states of fermionic ^{6}Li atoms far below degeneracy. To tune the interaction between the constituents, we make use of a broad magnetic Feshbach resonance between the states. We achieve temperatures far in the degenerate regime that can be tuned by a trap release heating procedure. Being able to tune temperature and interaction allows us to access the whole BEC-BCS crossover regime and investigate related phenomena.

One key quantity across the crossover is the critical temperature of the thermal to superfluid phase transition that signals condensation of pairs. The critical temperature can be calculated in the limits of weak interactions, however, for strong interactions theoretical treatment is difficult. In this thesis, we use neural networks to explore the critical temperature in the whole crossover regime with unprecedented accuracy. We show that a supervisedly trained neural network can not only detect the critical temperature, but also distinguish between Cooper pairs and Feshbach molecules, which was not possible before. Moreover, we show that an unsupervisedly trained autoencoder network is able to determine the critical temperature without any external input other than time-of-flight pictures.

Another important quantity in the crossover regime is the superfluid gap parameter. We measure the gap by exciting the Higgs mode through radio-frequency modulation of the atom population. In the BCS limit, the frequency of the Higgs mode is given by twice the gap frequency. By observing the frequency dependent response of the system we can therefore measure the gap parameter. Towards strong interactions, the observed feature broadens, signalling an instability of the mode due to the gradual violation of the required particle-hole symmetry.

Exploring the non-equilibrium behaviour of a Fermi gas in the BEC-BCS crossover might hold the key to understand the formation dynamics of pairs and how to induce superconductivity in a system that does not exhibit superconductivity in its equilibrium state. In this work, we present two ways to expose the system to a sudden change of interaction, resulting in a non-equilibrium response. Firstly, we use a fast radio-frequency transfer of one of the hyperfine states into a different, third state. We observe dynamics on several timescales in the momentum distribution and condensate fraction. Secondly, we develop a new coil system that is able to change the magnetic field at the atom position by tens of Gauss within a few microseconds. We detail the construction, testing and alignment process and present

investigations into the long-time response of the system.

Machine learning procedures can not only be used for data analysis, but also to improve upon the control over experimental systems. In this work we present a novel approach to improve the detection and correction of experimental errors exponentially. We use two machine learning approaches, principal component analysis and small, fully connected neural networks to create an empirical model of the experimental setup. Using this model allows us to reduce the required amount of data for compensation of noise sources exponentially. We demonstrate both approaches in the context of micromotion compensation on a single trapped ytterbium ion in a radio-frequency Paul trap.

Parts of this thesis have been or will be published in the following articles:

1. A. Behrle, T. Harrison, J. Kombe, K. Gao, M. Link, J.-S. Bernier, C. Kollath and M. Köhl, *Higgs mode in a strongly interacting fermionic superfluid*, Nature Physics [1]

2. T. Harrison, M. Link, A. Behrle, K. Gao, A. Kell, J. Kombe, J.-S. Bernier, C. Kollath and M. Köhl, *Decay and revival of a transient trapped Fermi condensate*, arXiv preprint [2]

3. P. Kobel, M. Link and M. Köhl, *Exponentially improved detection and correction of errors in experimental systems using neural networks*, arXiv preprint [3]

4. M. Link, K. Gao, A. Kell, B. Rauf, M. Breyer and M. Köhl, *Precision characterization of the phase diagram of Fermi condensates using neural networks*, in preparation

5. M. Link, K. Gao, A. Kell, B. Rauf, M. Breyer and M. Köhl, *Unsupervised detection of the critical temperature in a Fermi condensate using autoencoders*, in preparation

6. K. Gao, M. Link, A. Kell, A. Hoffmann, B. Rauf, M. Breyer and M. Köhl *A compact and fast magnetic coil for the interaction manipulation of quantum gases with Feshbach resonances*, in preparation

Acknowledgements

First and foremost I would like to thank Prof. Dr. Michael Köhl for trusting me with the responsibility to realise a project like this. Thank you for the support throughout the years and all the informative discussions that helped making me a better scientist. I would like to thank you especially for supporting the idea of using machine learning approaches in my project although it was a new experience for the group and it was not clear whether it would lead to meaningful results. Being able to realise a project that lies this close to my heart might be the best experience I have had in science to date.

I would also like to thank the team on the *Humphry experiment*: Dr. Alexandra Behrle, Dr. Timothy Harrison and Dr. Kuiyi Gao, the first generation of *Humphries*. You taught me a lot and it was always fun working with you. Also, thank you for building such a sturdy experimental apparatus! I also want to thank the new generation of *Humphries*, namely Andreas Kell, Moritz Breyer and Daniel Eberz. It was always fun working with you and I really enjoyed the good atmosphere in our lab. It is good to know that the experiment will be in capable hands and I am excited to see what you will make of it. Thank you also to all the people that contributed to the experiment throughout the time, such as Benjamin Rauf, Valentin Jonas, Till Leuteritz and Justas Andrijauskas.

A special thank you also to Pascal Kobel, who was the driving force in transforming the compensation project from idea to application on his experimental setup. Our discussions were very insightful and it was a pleasure working with you!

A big thank you to Johannes Kombe, Jean-Sébastien Bernier and Corinna Kollath for the fruitful collaboration on the Higgs mode and quench projects.

Thank you Marcell Gall for sharing the experience of writing a PhD thesis [4]. Talking to and having an occasional beer with you and Thorsten helped me to stay sane. Without any doubt, looking at pictures of raccoons and quokkas improved the experience even further. Thanks also to Thorsten Langerfeld who not only was part of these great beer nights, but was always available to talk and help.

A big thank you to all the people from the experimental quantum physics group that made working long days worthwile, shared the occasional pain and suffering or were excited when excitement was due. Most importantly, however, thank you for being great scientists and always willing to help when needed. Thank you to Nicola Wurz, Steffen Gohlke, Jonas Schmitz, Santhosh Surendra, Pia Fürtjes, Tobias Kree, Nick Klemmer, Janek Fleper, Luisa Tolle, Jeffrey Chan, Felix Rönchen, Leon Reiffert, Moritz Scharfstädt, Anica Hamer, Jens Samland, Ralf Berner, Kilian Kluge, Jonathan Silver, Leonardo Carcagni, Ashwin Boddeti, Eugenio Cocchi, Luke Miller, Ferdinand Brennecke, Jan Drewes, Robert Maiwald, Daniel Pertot, Max Zawierucha, Vidhya Sasidharan Nair. A special thank you to Hendrik-Marten Meyer and Tim Ballance for being my mentors during my masters thesis and ultimately

being a big part of my decision to start my PhD.

A special thank you to Tina Naggert, who made me feel welcome from day one. Without you, our workgroup would not be half as productive. Talking to you and sharing a coffee always made me feel better.

I would also like to thank Dr. Akos Hoffmann, who saved us more than once when something broke and was always helpful regarding all kinds of technical questions. Thank you also for contributing the quench coil switching circuit to this work! I am aware that I owe you a crate of beer and I stand by that promise.

A special thank you to Prof. Dr. Ian Spielman from the University of Maryland and the National Institute of Standards and Technology. Spending six months in your research group at the beginning of my PhD taught me a lot about how to be a good physicist and be inspired by what we are doing. My fascination for machine learning algorithms was amplified while working with PCA on the Kibble-Zurek project. A big thank you of course also to all the students in the research group, especially Dina Genkina, Lauren Aycock, Hsin-I Lu, Francisco Salces-Carcoba, Ana Valdes Curiel, Andika Putra, Yuchen Yue and Seiji Sugawa.

I want to thank the people who proofread parts of this work, namely Moritz Breyer, Dina Genkina, Kuiyi Gao, Andreas Kell, Steffen Gohlke, Marcell Gall, Johannes Kombe and Daniel Eberz.

Thank you to all those, who knowingly or unknowingly took part in this work and who are not mentioned by name.

Vielen Dank an meine Eltern für die jahrelange bedingungslose Unterstützung. Ohne euch hätte ich diese Arbeit niemals verfassen können.

Danke an meinen Bruder Robert, meine Lieblingsschwägerin Maria und mein Patenkind, den kleinen Sonnenschein Tom, der mir vor Augen führt, dass es einen Sinn hat die Welt verbessern zu wollen.

Zu guter Letzt: Danke Christina, dafür dass du für mich da bist und mich aufbaust, wenn die Welt unfair zu sein scheint. Ich freue mich auf unsere Zukunft. Ich liebe dich.

Contents

CHAPTER 1

Introduction

Fermions are the fundamental building blocks of everyday matter. They surround us in many different kinds and shapes: From elementary fermions described by the standard model of particle physics like quarks and leptons, to composite fermions in the form of protons and neutrons, the key building blocks of atoms. Atoms themselves can either act fermionic (as in the case of ^{6}Li within this thesis) or bosonic (as in the case of ^{23}Na within this thesis), depending on the spins of their constituents. Fermions follow the Pauli exclusion principle, which states that no two indistinguishable fermions can occupy the same quantum mechanical state. As a consequence, for zero temperature, non-interacting fermions will always occupy the lowest *unoccupied* energy state. For an ensemble of indistinguishable fermions, the energy of the highest occupied state is called Fermi energy E_F. At this point, any dynamics apart from thermal excitations at non-zero temperature is frozen out. Things change, when interactions are introduced to the system: In 1908, Kamerlingh Onnes cooled liquid helium to less than 2.17 K, observing the first signs of superfluidity [5]. Soon after, Onnes observed that mercury was able to conduct electricity without dissipation when cooled below 4.2 K [6, 7]. Only close to 50 years later, in 1957, Bardeen, Cooper and Schrieffer (BCS) formulated their theory of superconductivity [8, 9] that was finally able to explain the phenomenon by the formation of Cooper pairs: weakly bound pairs of fermions whose pairing is mediated by weak attractive interaction over a long distance. The formation of Cooper pairs is only possible in the presence of a Fermi sphere and is a pure many-body phenomenon that cannot be explained by two-particle interactions.

Experiments with ultracold atomic Fermi gases like the Na-Li machine within this thesis provide a way to investigate a plethora of quantum mechanical phenomena like the aforementioned BCS superfluidity in a laboratory environment. To reach ultracold temperatures in the order of tens of nanokelvin, far in the degenerate regime with $T \ll T_F = E_F/k_B$, we use laser cooling and evaporative cooling techniques. In the final stage of the experimental preparation, we load ^{6}Li atoms into a dipole trap, formed by high power laser beams. At these low temperatures, the Pauli principle hinders thermalisation between indistinguishable fermions as the probability for scattering decreases. We circumvent this effect by creating a 50/50 mixture of two different hyperfine states. Since these states are distinguishable, the Pauli principle does not apply and the states can thermalise with each other. Where solid state experiments often suffer from the availability of material that suits the purpose and

from ways to directly observe any dynamics, Fermi gas experiments provide high tunability: Density, trapping potential, temperature and even dimensionality. Notably, by using magnetic Feshbach resonances between the states in the lithium mixture, we are able to set the scattering length, which characterises the interactions between the fermions, arbitrarily: Negative scattering length corresponds to attractive interactions, positive scattering length to repulsive interactions. By changing the magnetic field we can tune the scattering length from the limit of weakly attractive interactions to the limit of weakly repulsive interactions, across the Feshbach resonance, where the scattering length diverges.

The BEC-BCS crossover

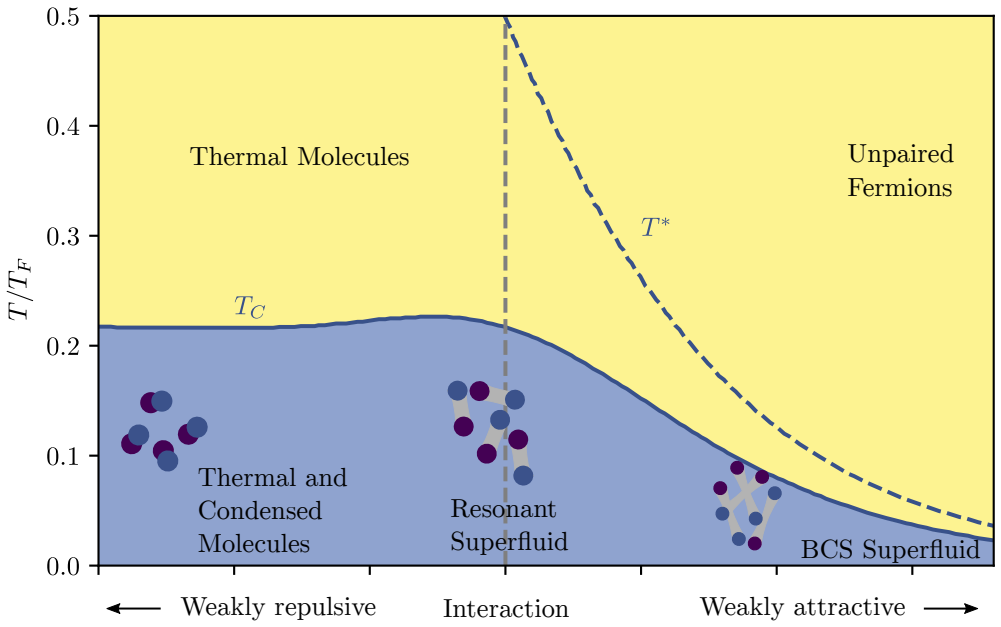

Figure 1.1: BEC-BCS phase diagram. The critical temperature T_C and pairing temperature T^* across the BEC-BCS crossover. Data adapted from [10, 11]. The shape of the curve can be understood as a guide to the eye, as the exact shape is not known and object of investigation.

The limit of weakly attractive interactions is described by the aforementioned model of Bardeen, Cooper and Schrieffer and is often simply called BCS limit. It is characterised by a second order phase transition from a thermal gas to a superfluid with a complex order parameter Δ. Its absolute value $|\Delta|$ is the minimum energy to create an excitation in the superfluid and can be understood as a measure for the binding energy of the Cooper pairs. $|\Delta|$ decreases exponentially with interactions and so does the critical temperature of the phase transition.

For repulsive interactions a two-body ground state exists. This allows for the formation of dimers of fermions that can act as composite bosons and are increasingly tightly bound

for weaker interactions. The bosonic dimers can then undergo Bose-Einstein condensation, which lends the weakly repulsive regime its name: BEC-regime [12–14]. For tightly bound dimers, the critical temperature for condensation approaches the critical temperature for Bose-Einstein condensation of bosons with two times the fermion mass and half the density.

Connecting both regimes is a region of diverging scattering length, that can be described by unitary parameters that are independent of microscopic properties. Fittingly, this regime is named unitarity regime. Although the scattering length diverges, the transition from tightly bound dimers to weakly bound Cooper pairs is smooth. Consequently, the critical temperature varies smoothly from limit to limit, with a maximum value predicted around unitarity. However, calculations are difficult when interactions become too strong to treat them perturbatively, which makes ultracold Fermi gases the ideal platform to investigate the whole BEC-BCS crossover shown in figure 1.1.

Exploring the crossover

Since the first experiments with degenerate Fermi gases in 2003 [15–17], the crossover has been the target of many investigations aiming to understand the physics of strongly interacting fermions. A key to understanding the dynamics and formation of the condensate below the critical temperature T_C is being able to determine T_C and the fraction of atoms in the condensate accurately. Although early measurements [17, 18] explored the critical temperature and condensate fraction in the crossover, the exact behaviour of T_C and the condensate fraction still remains elusive. There are theoretical approaches that aim to predict the behaviour of T_C [19–21], however, accurate experimental verification is still missing. At unitarity, where the equation of state is known, the critical temperature has been measured and provides a cross-check for theory [22].

Condensation in atomic samples is usually detected by measuring the momentum distribution in time of flight measurements. While a thermal gas shows a Boltzmann distribution, the condensate manifests itself as an inverted parabola around zero momentum signaling the macroscopic population of the momentum ground state. The resulting bimodal distribution consisting of said parabola with thermal background is a smoking gun for the presence of a condensate and can be used to determine the condensate fraction. Following this, being able to detect small amounts of condensate fraction and measuring the associated temperature allows for a precise determination of T_C. However, determining the condensate fraction in the BEC-BCS crossover is made difficult by the fragility of the pairs towards the BCS regime. When the binding energy of the pairs becomes comparable to the thermal energy of the system, the pairs break when released from the trapping potential in time of flight. This obstructs the observation of the condensate in momentum space. One way around this is to project the Cooper pairs onto tightly bound molecules by ramping the magnetic field across the Feshbach resonance in time of flight [17, 18]. Since the binding energy of the dimers then is much higher, a bimodal distribution can be observed. However, this technique comes with its own problems: For one, the conversion efficiency from Cooper Pairs to dimers is not known a priori and may depend on the exact characteristics of the used ramp. In combination with inaccurate fitting for low condensate fractions where the inverted parabola is smaller than the thermal background, this prevents an accurate measurement of small

condensate fractions. Secondly, the change in interaction during time of flight may obstruct phenomena that are critically dependent on the interaction. Lastly, measuring T_C in physical units requires a reliable thermometry to be established in the whole crossover regime.

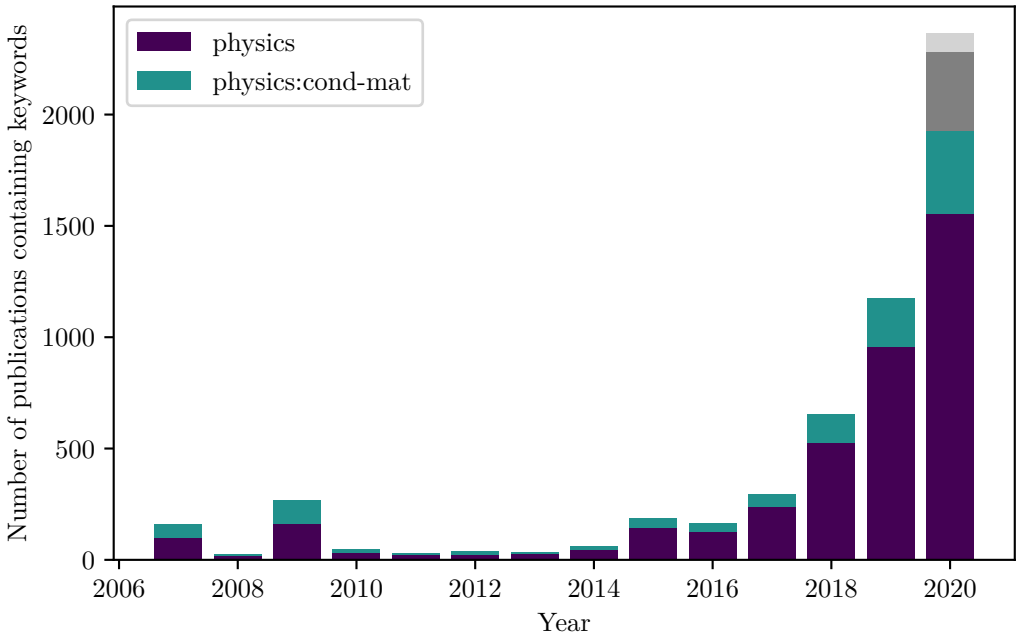

Figure 1.2: arXiv publications regarding machine learning. The total number of publications on arXiv containing one of the keywords *Machine Learning*, *Artificial Intelligence*, *Deep Learning* or *Neural Network* in the title for sections physics and physics:cond-mat. Shown in grey is a linear projection until the end of the year. Data is from October 24th, 2020.

In the first part of this thesis, we present a new measurement scheme to accurately determine the critical temperature across the BEC-BCS crossover with neural networks. In the recent years, neural networks have been established as powerful tools in the context of physics [23]. This is especially visible in the increasing number of machine learning related publications in the physics sections of arXiv, see figure 1.2. Moreover, machine learning methods have been shown to be very effective in the detection of phase transitions on theoretical data [24–30], as well as on experimental data [31]. Here, we use a neural network to determine the condensate fraction from single time-of-flight pictures throughout the whole BEC-BCS crossover. For training the neural networks, we provide labels generated by a rapid ramp sequence, however, the detection itself does not rely on the rapid ramp. Tuning the temperature by changing the trap confinement for a variable amount of time and using the generalisation abilities of the network provides us with a clear onset of condensation, allowing for a high precision determination of the critical point in terms of heating time. To convert heating time to Kelvin, we employ a thermometry based on a measurement of the density on the cloud surface, where interaction effects can be neglected. This provides us with a thermometry that is viable throughout the whole crossover and allows us to measure the critical temperature

with unprecedented accuracy in the whole crossover regime. Surprisingly, we find that the network is not only able to detect a condensate fraction from single shots without obvious experimental signature, but can also distinguish between tightly bound dimers and Cooper pairs, which was not possible before. Additionally, we present the first unsupervised detection of a phase transition on experimental data with no obvious experimental signature. For this, we generate a low-dimensional representation of time of flight pictures at different temperatures throughout the BEC-BCS crossover. By analysing this representation, we are able to extract the critical point in agreement with results from the supervisedly trained network. This circumvents the use of the rapid ramp technique altogether, allowing for the determination of T_C without any underlying assumptions.

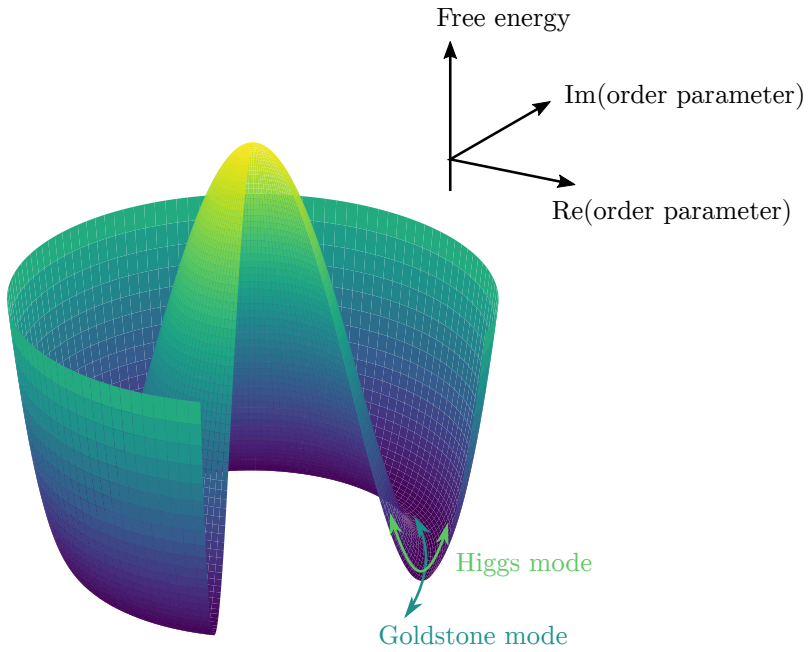

Figure 1.3: Mexican hat potential. When crossing the phase boundary from a thermal gas to a superfluid, the energy landscape deforms to a mexican hat with broken phase symmetry. This landscape has two possible modes, the Higgs mode along the radial direction and the Goldstone mode along the rim of the hat.

When transitioning the phase border from a thermal gas to a superfluid, the system undergoes a second order phase transition to a state with broken phase symmetry. The resulting free energy landscape has the shape of a mexican hat with a minimum at non-zero $|\Delta|$ and exhibits two distinct collective modes: The Goldstone mode, which corresponds to changes in the phase, along the valley of the mexican hat and the Higgs mode, which corresponds to oscillations along the radial direction, see figure 1.3. The stability of the Higgs mode against decay into the Goldstone mode has to be secured by additional symmetries. In the case of a BCS superfluid, particle-hole symmetry can take this role. Experimentally, a stable Higgs mode has been observed in conventional BCS superconductors [32–34], however,

only indirectly through non-linear terahertz spectroscopy and coupling to charge-density waves. The energy of the Higgs mode is directly proportional to the gap energy $\hbar\omega_{\text{Higgs}} = 2|\Delta|$, so measuring the energy of a stable Higgs mode amounts to a measurement of the superfluid gap $|\Delta|$. This is of great interest, as the superfluid gap is a key parameter for understanding the physics of the superfluid. It was investigated numerically [19, 35–39], but so far had only limited experimental access [40, 41]. In this thesis, we present a novel scheme to directly excite and measure the Higgs mode. For this, we modulate the population of one of the two hyperfine states in the two state Fermi mixture with variable frequency. We verify numerically that this scheme results in a modulation of $|\Delta|$ and therefore is able to excite the Higgs mode. By scanning the modulation frequency and observing a drop in the condensate fraction, we are then able to measure the frequency at which the Higgs mode occurs for several values of the interaction strength. We compare the corresponding frequencies with numerically and experimentally determined values of the superfluid gap and find good agreement. Towards the BEC regime, where particle-hole symmetry is not fulfilled, we observe a broadening of the Higgs mode signalling its increasing instability.

Non-equilibrium physics in the crossover

Of special interest within the BEC-BCS crossover are phenomena that require the system to be out of equilibrium. Although these phenomena are hard to treat theoretically [42, 43], they cause a variety of interesting physics not accessible in equilibrium [44], such as superconductivity induced by strong ultrashort laser pulses in condensed matter physics [45, 46]. In the context of ultracold Fermi gases, non-equilibrium can be investigated by measuring the response of the system to changes in the interaction strength that are fast compared to the intrinsic timescales of the system. However, beating the most relevant timescale – the time corresponding to the Fermi energy τ_F on the order of 5 µs – comes with obvious experimental difficulties. Controlling the interaction in the Fermi gas via Feshbach resonances requires a strong controllable magnetic field (on the order of $\approx 1\,000$ G). Creating such a field requires large coils and high currents that are intrinsically slow. This is further complicated by the presence of a metallic vacuum chamber that is prone to eddy currents that act against fast changes of the magnetic field. In this thesis, we present two realisations of interaction quenches that attempt to circumvent these experimental difficulties.

For all combinations of two hyperfine states of the lowest three hyperfine states of ^{6}Li, the corresponding Feshbach resonances are reasonably close to each other. Our first scheme uses a fast radio-frequency transfer of one of the two states in the Fermi mixture to another, third state. This amounts to transferring the mixture from one Feshbach resonance to another Feshbach resonance without changing the magnetic field. Since center field and width of the resonances are different, this can result in an effective interaction quench. The characteristic time of the quench is equal to the time needed to do the full rf flip $\tau \approx 28(2)$ µs, which is still slower than $\tau_F \approx 5$ µs, but faster than the typical timescales for magnetic field quenches and in the order of the timescale characterising excitations in the superfluid. Following the quench, we observe dynamics on several time scales in the momentum distribution and condensate fraction of the sample. For the momentum distribution, we observe the excitation of the monopole mode within the trap potential. On a shorter timescale, we observe a reshuffling

of quasiparticle excitations. We also observe a fast decay of the condensate fraction with a subsequent unexpected revival. We compare all of our findings to a theoretical model and find good qualitative agreement.

In order to quench faster than the Fermi time, we develop a system of small coils that is able to rapidly change the magnetic field by several Gauss. Close to a Feshbach resonance, this amounts to a reasonable quench in interaction strength. The new coil system is specifically designed for fast magnetic field changes by keeping the coils small to limit the mutual inductance to the Feshbach coils and prevent eddy currents. The final design consists of two coils that are wound onto 3D printed bodies and slid into each other. The generated magnetic field at the atom position is $\approx 36\,\mathrm{G}$ with vanishing gradient. A specifically designed circuit is able to rapidly turn off the current through the coil. We achieve quench times of $3\,\mu\mathrm{s} < \tau_\mathrm{F} \approx 4.7\,\mu\mathrm{s}$. We detail the construction, testing and alignment of the coil system. We give an overview on the long time dynamics after a rapid quench, which shows an interaction dependent delay between density and condensate fraction dynamics. Finally, we give a detailed outlook on how these quenches may be used to directly detect the Higgs mode in the future.

Improving experiments

Machine learning techniques can not only be used to analyse data, but also to optimise experimental systems [47–50]. This is usually done by using reinforcement learning techniques on experimental control parameters, however, the process of learning then takes a long time because of the large state space. In this thesis we show how techniques from supervised and unsupervised learning can be used to improve the detection and correction of experimental errors exponentially as compared to general approaches. For this, we build an empirical model of the experimental system from data used for previous error corrections. We then use this model to speed up the process of error correction exponentially. We test these new approaches on the example of stray field compensation of a single trapped ytterbium ion. They are, however, applicable to the general problem of experimental error correction. We furthermore show that our methods are only limited by the measurement error and not by the method itself.

The structure of this thesis

In order to establish a basis for understanding the investigations that are presented in this work, the first chapters provide a general introduction to the theory and experiment. Chapter 2 provides an introduction to the theory of ultracold Fermi gases. The goal of this chapter is to establish the theoretical knowledge that is necessary to understand the investigations in later chapters. Apart from general properties of Fermi gas systems, this includes the mechanism behind Feshbach resonances and the effect of non-zero temperature. Chapter 3 provides an introduction to machine learning methods with focus on neural networks. It is designed such that it provides a comprehensive overview and aims to provide all information that is required to understand the machine learning techniques that are used throughout the investigations conducted in this work. The experimental system and techniques that are used

to create the ultracold ^{6}Li mixture are introduced in chapter 4. This also includes the rapid ramp technique and its experimental implementation within our experiment.

Following these introductory chapters, we show investigations that were conducted during this work. Firstly, chapters 5 and 6 detail how we use neural networks to explore the BEC-BCS crossover phase diagram. We show how a network can be supervisedly trained to predict the condensate fraction and the critical temperature. Additionally, we show how an autoencoder can be used to create a low-dimensional representation of the phase diagram data and how this can be used to determine the critical point. In chapter 7, we establish a method to directly excite the Higgs mode by modulation of the population of one state in the Fermi mixture. We measure the superfluid gap in the BCS regime and compare our results to experimental and theoretical predictions. In chapter 8, we present two realisations of interaction quenches. The first one is based on fast radio-frequency transfers of one of two hyperfine state into a third state. We investigate the resulting dynamics in the condensate fraction and momentum distribution of the sample and find dynamics on several timescales that are cross-checked with a theoretical model. In the second part of this chapter, we detail the design, construction and implementation of a coil system that is able to realise interaction quenches by a fast change of the magnetic field. We show the long-time dynamics after such a quench. In chapter 9, we establish a new experimental technique that is able to exponentially decrease the amount of data needed for the compensation of experimental errors. For this, we create an empirical model of the experimental system with machine learning approaches that is able to predict the correct settings for error compensation from a strongly reduced amount of measurements.

In chapter 10, we will recapitulate and suggest future investigations that utilise the results of this work.

CHAPTER 2

Fermi Gases

This chapter aims to introduce basic concepts that are necessary for understanding the experiments performed throughout this thesis. It will start by recapitulating the basic properties of fermions at ultracold temperatures close to 0 K. It then proceeds to increase the complexity of the model by introducing interactions between the particles, and how to realise them experimentally. In order to understand the effects measured in the experiment, where zero temperature can not be reached, finite temperatures will be investigated. The goal in the end is to establish all necessary concepts that are shared throughout the investigations in later chapters. In the scope of this thesis, this chapter can only serve to be a short introduction to the basic concepts. There are several much more detailed reviews to (some of) the topics touched here and on which this chapter is based, see e.g. refs [51–54].

2.1 Fermionic atoms and cold temperatures

Fundamentally, matter as we know it is divided into two different kinds of particles: bosons and fermions. The spin-statistics theorem states that the wave function of identical particles with an integer spin are symmetric under exchange, whereas the wave function of identical particles with half-integer spin is antisymmetric. The first kind of particle is called *boson* and the latter is called *fermion*. Let $\Psi(x, y)$ be the many particle wave function of two identical particles at position x and y, rephrasing the spin-statistics theorem for fermions in terms of the wave function gives

$$\Psi(x, y) = -\Psi(y, x). \tag{2.1}$$

This does not change the probability to find the particles at either x or y, as $|\Psi|^2$ is not changed, however, if $x = y$, the above equation is only true if $\Psi(x, x) = 0$ and consequently $|\Psi(x, x)|^2 = 0$. Phrased differently: Two identical particles can not be at the same position at the same time. This principle was first phrased by Wolfgang Pauli in 1925 and hence has become famous as *Pauli principle*. The Pauli principle only holds for fermionic matter, because in the case of bosons, equation 2.1 does not have the minus sign and therefore the wave function of two identical particles at the same position can have a finite value.

Fermions come in a variety of kinds: From elementary particles like quarks and leptons, to composite particles like protons, neutrons or atoms. Bosons are represented by force carriers, such as photons and the Higgs/W/Z-bosons, but composite systems such as atoms can acquire a total integer spin, making them effective bosons. In our experiment, we investigate samples of ^{6}Li atoms, which behave fermionic. In the process of preparing the lithium however, we also use bosonic ^{23}Na atoms, and by exploiting their bosonic properties, we cool the lithium down to ultracold temperatures.

2.1.1 Fermions: The Fermi-Dirac distribution

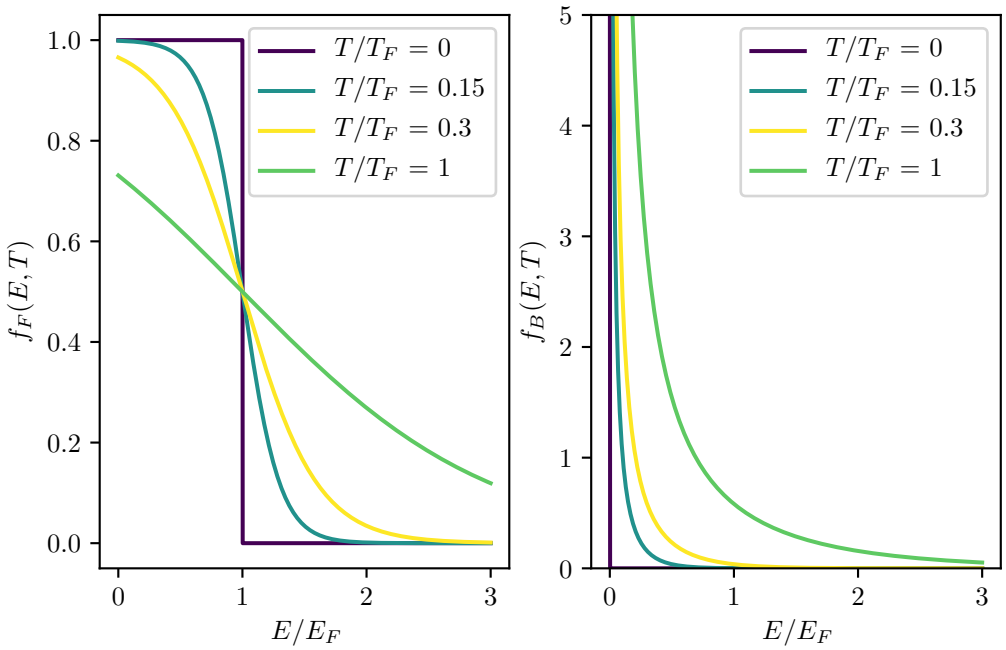

Figure 2.1: Difference between fermions and bosons. The Fermi-Dirac distribution (left) and the Bose-Einstein distribution (right) for non-interacting particles at different temperatures. In case of the Fermi-Dirac distribution, the occupancy of an energy state is never larger than one due to Pauli blocking. For the Bose-Einstein distribution, the occupation of the lowest state diverges towards zero temperature - all particles condense into the ground state. Note that for both distributions the integral over all energies always equates to unity: The particle has to be found in some state.

One important consequence of the Pauli principle is that even at zero temperature, in a system with identical fermions, not all particles can occupy the single particle ground state. Rather, the particles are *stacked up*, always occupying the lowest unoccupied state. The energy of the highest occupied state at zero temperature is known as *Fermi energy* E_F. The associated Fermi wave vector $k_F = \sqrt{2mE_F}/\hbar$ and Fermi temperature $T_F = E_F/k_B$ define the properties of the Fermi gas at zero temperature and provide a well defined reference frame for several physical quantities. At non-zero temperature the occupation probability of a state with Energy E is given by the Fermi-Dirac distribution [55]

$$f_F(E,T) = \frac{1}{\exp\left((E-\mu)/k_BT\right)+1}. \tag{2.2}$$

Here, E denotes the energy, T the temperature, μ is the chemical potential and k_B is Boltzmanns constant. As can be seen in figure 2.1, at zero temperature the probability to find a state with energy less than the Fermi energy occupied goes to 1, whereas the occupation of all states with higher energy goes to 0. This edge is called Fermi edge, and it washes out for non-zero temperatures due to the increasing probability to find particles in states with $E > E_F$. For temperatures $T \gg T_F$, the Fermi-Dirac distribution tends towards a Maxwell-Boltzmann distribution that describes a thermal Fermi gas.

2.1.2 Density distribution

Equation (2.2) can be rewritten in terms of the kinetic energy $\mathbf{p}^2/2m$ and potential energy $V(\mathbf{r})$ to yield

$$f_F(\mathbf{r},\mathbf{p},T) = \frac{1}{\exp\left((\mathbf{p}^2/2m + V(\mathbf{r}) - \mu)/k_BT\right)+1}. \tag{2.3}$$

The density distribution for a fixed temperature T can now be calculated by integrating out the momentum dependency

$$n(\mathbf{r}) = \int_{-\infty}^{\infty} \frac{d^3p}{(2\pi\hbar)^3} f_F(\mathbf{r},\mathbf{p}) \tag{2.4}$$

$$= -\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{3/2} \mathrm{Li}_{3/2}\left(-e^{(\mu - V(\mathbf{r}))/(k_BT)}\right). \tag{2.5}$$

Here $\mathrm{Li}_{3/2}(x)$ denotes the polylog function, defined via $\mathrm{Li}_s(x) = \sum_{k=1}^{\infty} \frac{x^k}{k^s}$. Importantly, for small arguments the polylog function approaches identity $\mathrm{Li}_{3/2}(x) \approx x$. As a consequence, near the surface of a trapped cloud, where the density is low and the chemical potential is only a small correction to the potential, $n(\mathbf{r})$ can be approximated by an exponential function $\propto \exp(-V(\mathbf{r})/(k_BT))$. From the slope of this function it is possible to determine the temperature of the cloud if $V(\mathbf{r})$ is known. In case of absorption imaging along one spatial direction, the direction of imaging is integrated out, however, for sufficiently low density in the wings of the cloud, this integration only modifies the amplitude of the expression, which is not needed in order to determine the temperature. In chapter 5, we use fits to the surface of an atomic cloud to determine the temperature of the sample in this way.

2.1.3 Bosons: Bose-Einstein statistics

Due to the Pauli principle not applying, the distribution function for bosons looks vastly different than the Fermi-Dirac distribution. Going towards zero Temperature, bosons will all occupy the lowest possible state and the average occupancy of the ground state diverges. The average occupancy is described by Bose-Einstein statistics

$$f_B(E,T) = \frac{1}{\exp\left((E-\mu)/k_BT\right) - 1}. \tag{2.6}$$

For low temperature, f_B diverges and all bosons will occupy the same state, enabling Bose-Einstein condensation, see figure 2.1. Although this thesis is mainly concerned with physics in fermionic systems, bosons play an important role as well: For increasingly low temperature, collisions of indistinguishable fermions are suppressed by the Pauli principle - due to the simple fact, that in order to collide, the well localised cold atoms have to be at the same position. For bosons, this is not true. In order to cool our cloud of ^{6}Li, we therefore use a very large cloud of bosonic ^{23}Na which then cools the lithium sympathetically. Another occurrence of bosons in our experiments is the existence of bosonic dimers, or Feshbach molecules (see section 2.3), bound states of two distinguishable fermions that act as a composite boson and can condense into a Bose-Einstein condensate.

2.1.4 Harmonically trapped fermions

In an experimental environment, atoms are oftentimes trapped in potentials generated by laser beams. For cold samples, the size of the cloud is often much smaller than the waist of the trapping laser beam. In this case, the trapping potential is formed by the center of the Gaussian beam and is well approximated by a harmonic potential

$$V = \frac{m}{2}(\omega_x^2 x^2 + \omega_y^2 y^2 + \omega_z^2 z^2), \tag{2.7}$$

with m being the mass of the particle and ω_i being the harmonic oscillator frequency in i-direction. The energy of state n_x, n_y, n_z is then given by the sum of the energies for every spatial direction

$$E_{x,y,z} = \hbar(\omega_x(n_x + 1/2) + \omega_y(n_y + 1/2) + \omega_z(n_z + 1/2)). \tag{2.8}$$

Ignoring zero-point energy contributions, all states with energy less than E lie within a pyramid whose defining edges in n-space have length $n_{\mathrm{i,\ max}} = E/\hbar\omega_i$. The number of available states below E is then given by the volume of this pyramid, which is

$$N = \frac{1}{6} n_{\mathrm{x,\ max}} n_{\mathrm{y,\ max}} n_{\mathrm{z,\ max}} = \frac{E^3}{6\hbar^3\omega_x\omega_y\omega_z}. \tag{2.9}$$

Derivation with respect to E gives the density of states, which describes the number of states contained in an infinitesimal energy interval δE

$$\rho(E) = \frac{\partial N}{\partial E} = \frac{1}{2}\frac{E^2}{\hbar^3 \bar{\omega}^3}, \tag{2.10}$$

with the geometric mean of the trap frequencies $\bar{\omega} = (\omega_x \omega_y \omega_z)^{1/3}$. Moreover, the Fermi energy for a harmonic trap filled with N Fermions can be found by rearranging equation (2.9) to yield

$$E_F = (6N)^{1/3} \hbar \bar{\omega}. \tag{2.11}$$

2.2 Scattering and interactions

In the following section, we will explore the interaction between two atoms. As Pauli blocking prevents dynamics due to the dynamics freezing out at low temperature, we will investigate the behaviour of two distinguishable species (or states) of atoms. In this work, those distinguishable atoms are represented by ^{6}Li atoms in two different hyperfine states. First, a simple model for scattering in ultracold, ultra-dilute atomic samples will be presented. This will show that for low temperatures, the scattering process is characterised by a single variable, the *scattering length* a. In the second part of this section, Feshbach resonances will be introduced. These resonances present a tool to tune the scattering length and therefore the interactions between atoms by changing an experimentally readily accessible parameter: the magnetic field. More detailed explanations of these concepts can be found in refs [51, 54, 56, 57].

2.2.1 Interacting ultracold Fermi gases

In order to investigate the scattering of two distinguishable atoms, we presume the wave functions of both atoms can be written as plane waves in momentum eigenstates $\mathbf{k}_1$ and $\mathbf{k}_2$. To describe the collision, we switch to the center-of-mass system with relative coordinates. In this case, the relative motion is completely determined by the Schrödinger Equation

$$(\nabla^2 + k^2)\Psi(\mathbf{r}) = v(r)\Psi(\mathbf{r}), \tag{2.12}$$

with $k^2 = mE/\hbar^2$ and $v(r) = mV(r)/\hbar^2$. Here, $V(r)$ is a central potential, only dependent on the relative distance r. Far away from the scattering center, the solution to equation (2.12) can be written as sum of the incident plane wave and a scattered spherical wave,

$$\lim_{r\to\infty} \Psi = \underbrace{e^{ikz}}_{\text{incident}} + \underbrace{f(\theta)\frac{e^{ikr}}{r}}_{\text{scattered}}. \tag{2.13}$$

The scattering amplitude $f(\theta)$ represents the probability amplitude for scattering into a given direction and θ denotes the scattering angle. Since the incoming wave is chosen to be along the z-direction, $f(\theta)$ does not depend on the azimutal angle ϕ, which defines the scattering plane. Since $V(\mathbf{r})$ is a central potential, $\Psi(\mathbf{r})$ can be expanded into radial and angular eigenfunctions

$$\Psi(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{m=l} c_{lm} R_l(k,r) Y_l^m(\theta,\phi). \tag{2.14}$$

This expansion is known as *partial wave expansion* with $R_l(k,r)$ being the radial wave function and $Y_l^m(\theta,\Phi)$ being the angular wave function. By rewriting the scattered wave from equation (2.13) with the partial wave expansion, it can be shown that the scattering amplitude takes the form

$$f(\theta) = \frac{1}{2ik} \sum_{l=0}^{\infty} (2l+1)(e^{2i\eta_l} - 1) P_l(\cos(\theta)), \tag{2.15}$$

with $P_l(\cos(\theta))$ being the Legendre Polynomials [54]. The coefficients c_{lm} have been replaced by a phase factor $e^{2i\eta_l}$. In the limit of ultracold atomic gases, it is viable to assume that the wave vector of the particle is much smaller than the inverse of the range of the interatomic potential r_0, $k \ll 1/r_0$. In this case, it can be shown that the dominant term of the partial wave expansion is the $l = 0$ term [54, 57] and $f(\theta)$ can be rewritten to yield

$$f(\theta) \approx f_s = \frac{1}{k\cot(\eta_0) - ik}. \tag{2.16}$$

This limit is called s-wave limit. Again using the limit of low temperature with small k, the $k\cot(\eta_0)$ term can be expanded to order k^2, which gives

$$k\cot(\eta_0) \approx -\frac{1}{a} + r_{\text{eff}} \frac{k^2}{2}, \tag{2.17}$$

with r_{eff} being the effective range of the interaction, which for low momenta can be approximated by $r_{\text{eff}} \approx r_0$ [54]. a is the *scattering length*, defined through

$$a = -\lim_{k \ll 1/r_0} \frac{\tan(\eta_0)}{k}. \tag{2.18}$$

The scattering length is related to the phase shift between incoming and scattered wave and can be thought of as an effective range for the interaction. Importantly, positive a represents attractive interactions and negative a repulsive interactions. Substituting equation (2.17) into equation (2.16) gives

$$f(k) = \frac{1}{-\frac{1}{a} + r_{\text{eff}}\frac{k^2}{2} - ik}. \tag{2.19}$$

Notably, $f(k)$ does not depend on the angle due to the symmetry of the s-wave scattering. Generally, the quantity of interest is the scattering cross section σ, which describes the probability of a scattering event. Since the squared absolute of the scattering amplitude $|f(\theta)|^2$ describes the probability of scattering into a given direction, σ can be determined by integration over the solid angle via

$$\sigma = \int_\Omega d\Omega\, |f(\theta)|^2 = \int_0^\pi 2\pi |f(\theta)|^2 \sin^2\theta\, d\theta. \tag{2.20}$$

Evaluating the above integral gives the *optical theorem*, which connects the imaginary part of forward scattering ($\theta = 0$) with the cross section [54, 57]

$$\sigma = \frac{4\pi}{k}\text{Im}f(0). \tag{2.21}$$

Using the low-momentum effective-range expansion from equation (2.17), it is possible to evaluate the s-wave scattering cross section, which yields

$$\sigma_s = 4\pi a^2 \frac{1}{1 + k^2 a^2 (1 - r_{\text{eff}}/a) + \ldots}. \tag{2.22}$$

Interestingly, for small k, σ_s is the same for attractive ($a > 0$) and repulsive ($a < 0$) interactions, although thermodynamics in both cases are vastly different. In the limit of weak interactions $k|a| \ll 1$, the cross section becomes $\sigma_s \approx 4\pi a^2$, which is equal to scattering of two hard spheres with radius $|a|$. In the limit of large $k|a| \gg 1$ and low momenta $r_{\text{eff}} \ll 1/k$, the cross section becomes independent of all potential-specific quantities $\sigma \approx 4\pi/k^2$. This means, that even for diverging scattering length, the cross section can not become larger than this value. This limit is called *unitarity limit*.

2.2.2 Tuning interactions: Feshbach resonances

In order to access a variety of different phenomena that occur for different scattering lengths a, it is necessary to tune the scattering length over a wide range. Generally, a is governed by the interatomic potential and therefore is a fixed quantity, labelled the background scattering length a_{BG}. However, for certain magnetic fields it is possible to scatter resonantly into a bound state of an energetically higher channel. These resonances are called (Fano-) Feshbach resonances [58, 59] and provide means of tuning the scattering length a away from its background value a_{BG}. This section gives a short overview of Feshbach resonances. Far more detailed explanations can be found in refs [51, 54, 56, 60]. Here, we focus on the case of ^{6}Li,

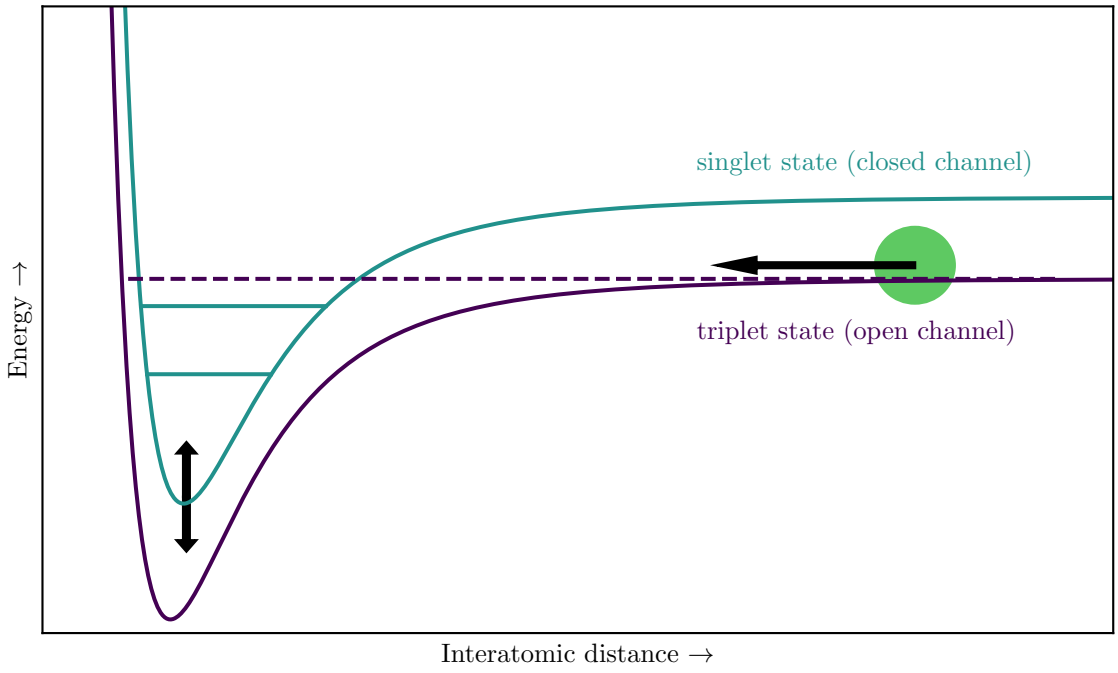

Figure 2.2: Mechanism for Feshbach resonances. The atom enters the interatomic potential in the open channel. Initially, the energy of the open channel is too high for the atom to be excited. By tuning the magnetic field, the relative energy distance between the two channels can be tuned, such that the entry energy (dashed line) coincides with the energy of a bound state. The value of the resulting scattering length is determined by the energy difference between entrance channel and bound state.

however, the general principle is applicable to different systems as well.
Generally, far away from collision, atoms are prepared in specific hyperfine states. The specific combination in which the atoms are prepared is called the collisional entrance channel. Due to the rich internal structure of hyperfine states, atoms have a variety of channels they can scatter into. For large distances $r \gg a_B$, the energy of the channels is determined by the Zeeman and hyperfine energy structure. For closer distances on the order of the Bohr radius a_B, it is governed by molecular potentials [56]. Channels, in which atoms can separate after collision due to to energy confinement are open collisional channels, whereas channels that are energetically too high are called closed collisional channels. A Feshbach resonance is present, if a closed channel contains a bound state to which scattering can occur. By tuning the energetic distance between both channels the scattering length can then be tuned, see figure 2.2.

In the case of ^{6}Li and high magnetic fields ($B > 140\,\mathrm{G}$) the lowest three hyperfine states all have a spin projection of $m_s = -1/2$. Coupling of two hyperfine states is possible into a singlet state with total spin $S = 0$ that is unaffected by the magnetic field, or a triplet state with $S = 1$ and $m_S = -1, 0, +1$ with a Zeeman splitting of $\Delta E = g_s m_S \mu_B B$. Due to this Zeeman splitting, for low temperatures $k_B T \ll g_s m_S \mu_B B$ and large distance, the only

open collisional channel is is the lowest triplet state $|S, m_S\rangle = |1, -1\rangle$. However, due to the different Zeeman energies – which determine the energy far away from the collision – the relative energy shift between the channels can be tuned by tuning the magnetic field. For collisions of ultracold ^{6}Li at high magnetic field, the total spin $F = S + I + L$ is a conserved quantity, with $L = 0$ for alkali atoms. $S = s_1 + s_2$ and $I = i_1 + i_2$ however are not conserved separately due to the hyperfine interaction. This provides the mechanism to scatter from the triplet state into a bound state of the singlet state and therefore a Feshbach resonance. If the energy of the open channel is tuned slightly below the energy of the bound state, the scattering length will be large and negative. If tuned slightly above the energy of the bound state, it will be large and positive [53]. Feshbach resonances can be characterised by a simple functional form [51]

$$a(B) = a_{\mathrm{BG}} \left(1 - \frac{\Delta_B}{B - B_0}\right), \tag{2.23}$$

where Δ_B describes the width and B_0 the center position of the Feshbach resonance. Due to lack of knowledge regarding the short range molecular potentials, the characteristic parameters B_0 and Δ_B are usually determined experimentally. If the kinetic energy of the sample $\hbar^2 k^2/(2m)$ is small compared to the energy width of the Feshbach resonance, we speak of a *broad* Feshbach resonance. This means, all atoms in the sample experience the same scattering length $a(B)$. As shown in figure 2.3 and table 2.1, between the lowest three hyperfine states of lithium (for convenience labeled $|1\rangle, |2\rangle \,\&\, |3\rangle$) there are three exceptionally broad (between 100 G and 300 G) Feshbach resonances between 600 G and 900 G which provide very precise tuning of the scattering length [61, 62].

States	$a_{\mathrm{BG}}\,[a_B]$	$B_0\,[\mathrm{G}]$	$\Delta_B\,[\mathrm{G}]$
$\|1\rangle \,\&\, \|2\rangle$	-1405	834.15	-300
$\|1\rangle \,\&\, \|3\rangle$	-1727	690.43	-122.3
$\|2\rangle \,\&\, \|3\rangle$	-1490	811.22	-222.3

Table 2.1: Characteristics of Feshbach resonances between the lowest three hyperfine states of ^{6}Li, as reported in ref [61].

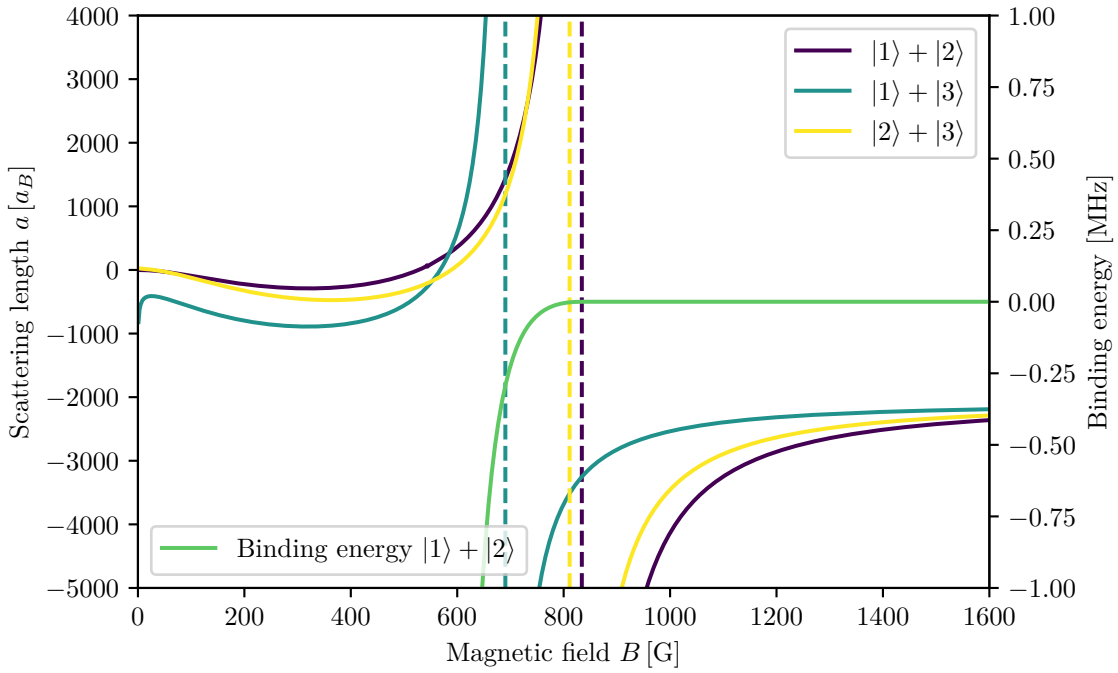

Figure 2.3: Feshbach resonances in ^{6}Li. The scattering length a versus magnetic field B of the broad Feshbach resonances between 600 G and 900 G. Data is taken from [63]. Shown in green is the binding energy of the two-particle bound state.

2.3 BEC-BCS crossover

Ultracold samples of interacting fermions in the same state form a Fermi fluid, which showcases the effect of Pauli blocking and fermion statistics, however, there are no phase transitions to study in such samples. Things become more interesting, when studying ultracold clouds of distinguishable fermions, such as samples of ^{6}Li in mixtures of two of the lowest three hyperfine states. The Feshbach resonances described in section 2.2.2 provide a convenient tuning knob for the interaction, which enables experimental investigation of several different phenomena:

- In the limit of weak repulsive interactions $k_F a \to 0_+$, the existence of a bound state leads to formation of tightly bound dimers which act as composite bosons. This enables Bose-Einstein-Condensation of molecular clouds in the vicinity of Feshbach resonances (hence, this regime is called *BEC regime*) [64, 65]. These clouds have comparably long lifetimes compared to pure bosonic clouds, due to the suppression of collisional losses by the Pauli exclusion principle.

- In the limit of weak attractive interactions $k_F a \to 0_-$, no such bound state exists. However, in 1911, the experiments of Kammerlingh Onnes showed a phase transition of metals who lost all electrical resistance when cooled to Temperatures of few Kelvin [6, 7]. This phase transition was later attributed to the formation of very weakly

bound pairs of electrons over a large distance, called Cooper pairs [66]. These pairs form a superfluid, which in turn enables the electrical resistance to drop. This pairing is possible even with vanishingly weak attractive interactions, however, only in the presence of a Fermi sphere that modifies the density of states. Consequently, Cooper pairing is a pure many body phenomenon. Since the theory describing this kind of superconductivity is known as Bardeen-Cooper-Schrieffer theory after their founders, this limit is usually called *BCS limit* [8, 9].

- In the crossover regime between both limits, where $1/(k_F a) \to 0$, the scattering length becomes much larger than the interparticle distance. Hence, all measurable quantities must be independent of the microscopic properties of the system, such as for example the cross section described in section 2.2.1. Due to the dependence on universal quantities, this limit is called *unitarity limit*. In this regime no analytical solutions are known and the descriptions from the above limits break down. Yet, the transition from the BEC limit to the BCS limit is smooth.

In the following all of the above limits will be explained in more detail. Key quantities will be derived and explained.

2.3.1 Repulsive interactions: The BEC regime

On the repulsive side of the Feshbach resonance, a molecular bound state between the two different fermion states exists [56, 60, 67]. These dimers have a binding energy of

$$E_B = -\frac{\hbar^2}{ma^2}, \tag{2.24}$$

and a characteristic size of a. The binding energy is plotted in figure 2.3 for states $|1\rangle$ and $|2\rangle$. If the size becomes small, these dimers act as composite Bosons and can therefore undergo Bose-Einstein condensation below a critical temperature T_C [64, 65]. These dimers, called Feshbach molecules, are created in a high rotovibrational state. Lifetimes of mixtures of these dimers are much longer than the lifetime of their bosonic counterparts, as relaxation into deep bound states necessary for inelastic three-body scattering is strongly suppressed by Pauli blocking [52, 64]. The scattering length between dimers is $a_{dd} = 0.6\,a$. [64, 68, 69].

2.3.2 Attractive interactions: The BCS regime

In order to understand the BCS regime it is instructive to first understand why no bound state exists for arbitrarily small attractive interactions. From this argument it will become apparent, why the presence of a Fermi sea is necessary to form Cooper pairs, which in turn condense into a superfluid.

The absence of a two-body bound state for attractive interactions

The starting point for the investigation of the existence of a bound state is the Schrödinger equation in momentum space [51, 56]

$$\frac{\hbar^2(q^2+\kappa^2)}{2m}\Psi_{\mathbf{q}} = \int_{-\infty}^{\infty} \frac{d^n q'}{(2\pi)^n}\Psi_{\mathbf{q}'}\tilde{V}_{\mathbf{q}-\mathbf{q}'}. \tag{2.25}$$

Here, we assume a binding energy of $E_B = \frac{\hbar^2\kappa^2}{2m}$ and $\Psi_{\mathbf{q}}$ and $V_{\mathbf{q}}$ are the Fourier transforms of the wave function and potential, respectively. The potential mainly acts on wave vectors that are comparable to the inverse of its characteristic length scale $|\mathbf{q}| \approx 1/r_0$. Going to the usual limit of small characteristic length scale r_0, the Fermi gas only probes a very small part of the potential in momentum space, since $k_F \ll 1/r_0$. It is therefore viable to assume a constant potential across the range of wave vectors in the Fermi gas, which is determined by its value around zero momentum $V_{\mathbf{q}} \approx V_{\mathbf{q}=0} = V_0$. Using this and $|\mathbf{q}| \approx 1/r_0$ as cut-off for the integral in the above equation and rewriting the integral in terms of energy, we arrive at [51, 56]

$$\frac{1}{V_0} = \int_0^{E_{r_0}} dE \frac{\rho(E)}{2E+|E_B|}. \tag{2.26}$$

This equation provides a link between the potential depth V_0, the density of states $\rho(E)$ and the binding energy E_B. Importantly, this equation holds true even for 2D and 1D systems, if the density of states is chosen accordingly. Assuming the density of states is proportional to E^α, for $\alpha > 0$, the right hand side will be finite [56]. For a 3D square well potential (which is a good approximation the case of small characteristic length scale) $\alpha = 1/2$, so in all cases the right hand side will be finite. However, if the potential depth is decreased, the left hand side will diverge. Therefore, the above equation does not have a solution for arbitrary small potentials. Consequently, in this case there is no bound state. By using a pseudopotential approach it can be shown that only in the case of repulsive interactions a bound state always exists with the binding energy given in equation (2.24). If a Fermi sphere is present the density of states is modified to be approximately constant on the Fermi surface ($\alpha = 0$), even in 3D. In this case the right hand side of the above integral diverges as well and there is always a solution for a bound state. This is the mechanism which enables the existence of Cooper pairs, even for vanishingly small attractive interactions. Since the Fermi sphere is required, Cooper pairing is a purely many-body process which can not be understood in terms of single particles. This a key difference to the BEC limit where the bound state can be understood from just two particles interacting with each other.

Cooper pairing

The full Hamiltonian of the BCS problem in its most general form is given by

$$\hat{H} - \mu\hat{N} = \underbrace{\sum_{\mathbf{k}} \frac{\hbar^2 k^2}{2m} \left(c^\dagger_{\mathbf{k},\uparrow} c_{\mathbf{k},\uparrow} + c^\dagger_{\mathbf{k},\downarrow} c_{\mathbf{k},\downarrow} \right)}_{\text{Kinetic Energy}} + \underbrace{\frac{1}{2} \sum_{\mathbf{k}_1,\mathbf{k}_2,\mathbf{q},\sigma_1,\sigma_2} \hat{\tilde{V}}_{\mathbf{q}} c^\dagger_{\mathbf{k}_1+\mathbf{q},\sigma_1} c^\dagger_{\mathbf{k}_2-\mathbf{q},\sigma_2} c_{\mathbf{k}_2,\sigma_2} c_{\mathbf{k}_1,\sigma_1}}_{\text{Interaction}} - \mu N. \tag{2.27}$$

Here, $c^{(\dagger)}_{\mathbf{k},\sigma}$ denotes an annihilation (creation) operator for a fermion with momentum $\mathbf{k}$ and spin σ. The first term sums up the kinetic energies for all particles and spins, where a particle with momentum $\mathbf{k}$ contributes a kinetic energy of $\hbar^2 k^2/(2m)$ (equally for both spin states). The second term describes the process of two particles with spins σ_1 and σ_2 and momenta $\mathbf{k}_1$ and $\mathbf{k}_2$ scattering through a phonon of momentum $\mathbf{q}$ within a momentum dependent interaction $\tilde{V}_{\mathbf{q}}$ and preserved spin. The effect of this Hamiltonian was first investigated by Bardeen, Cooper and Schrieffer in the 1950's in the context of electrons in a metal [8, 9]. In order to simplify equation (2.27), a few key assumption can be made: Firstly, due to Pauli blocking for short range interactions, it is useful to assume that interactions only take place between unequal spin particles $\sigma_1 \neq \sigma_2$. Secondly, it is assumed that only particles of equal and opposite momentum $\mathbf{k}_1 = -\mathbf{k}_2$ interact with each other. This assumption can be justified by estimating the number of final states after scattering. For this, imagine two particles with momenta $\mathbf{k}_1$ and $\mathbf{k}_2$ on top of the Fermi surface. Since all states inside of the Fermi sphere are occupied, they are not available as final states. In case of scattering with zero center-of-mass momentum, the whole Fermi surface is available, whereas with finite momentum only a small ring on the surface is available. In this case, the number of final states is therefore restricted. Hence, it is expected that most pairing takes place with zero center-of-mass momentum. The last simplifying assumption is that the interaction is not momentum dependent $\tilde{V}_{\mathbf{q}} = \tilde{V}_0$. With all these simplifications equation 2.27 takes the form

$$\hat{H} - \mu\hat{N} = \sum_{\epsilon_{\mathbf{k}}} \epsilon_{\mathbf{k}} \left(c^\dagger_{\mathbf{k},\uparrow} c_{\mathbf{k},\uparrow} + c^\dagger_{\mathbf{k},\downarrow} c_{\mathbf{k},\downarrow} \right) + \hat{\tilde{V}}_0 \sum_{\mathbf{k},\mathbf{q}} c^\dagger_{\mathbf{k}+\mathbf{q},\uparrow} c^\dagger_{-\mathbf{k}-\mathbf{q},\downarrow} c_{-\mathbf{k},\downarrow} c_{\mathbf{k},\uparrow}. \tag{2.28}$$

Here, $\epsilon_{\mathbf{k}} = \hbar^2 k^2/(2m) - \mu$ has been defined. In order to diagonalise this Hamiltonian, we use mean-field theory. For this, the pair annihilation operator $c_{\mathbf{k},\uparrow} c_{\text{-}\mathbf{k},\downarrow}$ is rewritten in terms of its expectation value and fluctuations around the expectation value,

$$c_{\mathbf{k},\uparrow} c_{\text{-}\mathbf{k},\downarrow} = \left\langle c_{\mathbf{k},\uparrow} c_{\text{-}\mathbf{k},\downarrow} \right\rangle + \underbrace{c_{\mathbf{k},\uparrow} c_{\text{-}\mathbf{k},\downarrow} - \left\langle c_{\mathbf{k},\uparrow} c_{-\mathbf{k},\downarrow} \right\rangle}_{\text{Fluctuations}}. \tag{2.29}$$

Inserting equation (2.29) into equation (2.28), we can neglect all terms quadratic in the fluctuations (hence the name mean-field). This finally yields the mean-field BCS Hamiltonian

$$\hat{H} - \mu\hat{N} = \sum_{\mathbf{k}} \epsilon_{\mathbf{k}} \left(c^\dagger_{\mathbf{k},\uparrow} c_{\mathbf{k},\uparrow} + c^\dagger_{\mathbf{k},\downarrow} c_{\mathbf{k},\downarrow} \right) - \left(\Delta^* c_{-\mathbf{k},\downarrow} c_{\mathbf{k},\uparrow} + \Delta c^\dagger_{\mathbf{k},\uparrow} c^\dagger_{-\mathbf{k},\downarrow} \right). \tag{2.30}$$

Here, we defined the superfluid order parameter as

$$\Delta = -\tilde{V}_0 \sum_{\mathbf{k}} \left\langle c_{\mathbf{k},\uparrow} c_{-\mathbf{k},\downarrow} \right\rangle . \tag{2.31}$$

The order parameter will be of great importance and its relevance will be discussed below. In order to diagonalise equation (2.30), we employ a Bogoliubov-Valatin transformation [70, 71] and define

$$c_{\mathbf{k},\uparrow} = u^*_{\mathbf{k}} \gamma_{\mathbf{k},\uparrow} + v_{\mathbf{k}} \gamma^\dagger_{-\mathbf{k},\downarrow}, \tag{2.32}$$

$$c^\dagger_{-\mathbf{k},\downarrow} = u_{\mathbf{k}} \gamma^\dagger_{-\mathbf{k},\downarrow} - v^*_{\mathbf{k}} \gamma_{\mathbf{k},\uparrow}, \tag{2.33}$$

$$|u_{\mathbf{k}}|^2 + |v_{\mathbf{k}}|^2 = 1, \tag{2.34}$$

where $\gamma^{(\dagger)}_{\mathbf{k},\sigma}$ is a fermionic annihilation (creation) operator for a new quasiparticle excitation with momentum $\mathbf{k}$ and spin σ. By minimising the free energy $\{H - \mu N\}$, we find

$$|u_{\mathbf{k}}|^2 = \frac{1}{2}\left(1 + \frac{\epsilon_{\mathbf{k}}}{\sqrt{\epsilon^2_{\mathbf{k}} + \Delta^2}}\right) \qquad |v_{\mathbf{k}}|^2 = \frac{1}{2}\left(1 - \frac{\epsilon_{\mathbf{k}}}{\sqrt{\epsilon^2_{\mathbf{k}} + \Delta^2}}\right). \tag{2.35}$$

Using this, equation (2.30) becomes

$$H - \mu N = \sum_{\mathbf{k}} \underbrace{(\epsilon_{\mathbf{k}} - E_{\mathbf{k}})}_{\text{Kinetic energy}} + \underbrace{E_{\mathbf{k}}(\gamma^\dagger_{\mathbf{k},\uparrow}\gamma_{\mathbf{k},\uparrow} + \gamma^\dagger_{\mathbf{k},\downarrow}\gamma_{\mathbf{k},\downarrow})}_{\text{Quasiparticle excitations}}, \tag{2.36}$$

which is diagonal in the quasiparticle number operators $\gamma^\dagger_{\mathbf{k},\sigma}\gamma_{\mathbf{k},\sigma}$ with ground state energy $E_{\mathbf{k}} = \sqrt{\epsilon^2_{\mathbf{k}} + \Delta^2}$. The first term in the above sum provides the ground state energy, whereas the second term is the excitation spectrum given by the number of spin-up quasiparticles and spin-down quasiparticles times their respective energy. As opposed to the excitation spectrum of a Fermi gas, the creation of an excitation in the BCS regime always requires at least an energy of Δ. This energy gap is the reason for superfluidity within the BCS regime, as a minimum energy needs to be put into the system in order to create an excitation. Excitations are Bogoliubov quasiparticles described by the γ-operators which describe superpositions of particles and holes. The superfluid gap can be written in experimentally accessible parameters for $T = 0$ to yield

$$\Delta = \frac{8E_F}{e^2} e^{-\pi/(2k_F|a|)}, \tag{2.37}$$

which connects the value of Δ to the s-wave scattering length [51]. Since Δ presents the

minimum energy to create an excitation which at the same time destroys a Cooper pair, it can be thought of as the binding energy of a Cooper pair. This showcases that Cooper pairs get increasingly more fragile, the weaker the interaction. The ground state wave function of the mean field BCS Hamiltonian can be written as [9, 51, 56]

$$|\Psi_{\mathrm{BCS}}\rangle = \prod_{\mathbf{k}} \left(u_{\mathbf{k}} + v_{\mathbf{k}} c^{\dagger}_{\mathbf{k},\uparrow} c^{\dagger}_{-\mathbf{k},\downarrow} \right) |0\rangle \,, \tag{2.38}$$

which leaves the interpretation of $v_{\mathbf{k}}$ ($u_{\mathbf{k}}$) as the probability amplitude to find (not find) a Cooper pair at momentum $\mathbf{k}$. Moreover, Δ acts as an order parameter for a second order phase transition from a thermal Fermi gas to a BCS superfluid. The critical temperature for the onset of Δ for weak interactions can be approximated (beyond mean field) by

$$k_B T_C \approx 0.277 E_F e^{-\pi/(2k_F a)}, \tag{2.39}$$

which is directly proportional to the superfluid gap Δ, see section 2.4.2. This approximation is known as Gorkov approximation [72].

2.3.3 The crossover: Unitarity

Although the scattering length diverges at unitarity, the connection from tightly bound molecules on the BEC side to weakly bound Cooper pairs on the BCS side is smooth. The equation for the gap (2.31) can be rewritten in terms of the $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$ to yield

$$\Delta = -\frac{V_0}{\Omega} \sum_{\mathbf{k}} u_{\mathbf{k}} v_{\mathbf{k}} = -\frac{V_0}{\Omega} \sum_{\mathbf{k}} \frac{\Delta}{2E_{\mathbf{k}}}, \tag{2.40}$$

with Ω being the volume of the system. Rearranging this equation and going to the continuum limit yields the gap equation

$$-\frac{1}{V_0} = \int \frac{d^3k}{(2\pi)^3} \frac{1}{2E_{\mathbf{k}}}. \tag{2.41}$$

This equation, together with the number equation for the total particle density

$$n = \frac{N}{\Omega} = 2 \int \frac{d^3k}{(2\pi)^3} v^2_{\mathbf{k}}, \tag{2.42}$$

can be solved simultaneously to yield the two unknown variables μ and Δ across the crossover. This yields two equations [51]

$$-\frac{1}{k_F a} = \frac{2}{\pi}\left(\frac{2}{3I_2\left(\frac{\mu}{\Delta}\right)}\right)^{1/3} I_1\left(\frac{\mu}{\Delta}\right) \tag{2.43}$$

$$\frac{\Delta}{E_F} = \left(\frac{2}{3I_2\left(\frac{\mu}{\Delta}\right)}\right)^{2/3}, \tag{2.44}$$

with

$$I_1(z) = \int_0^\infty dx\, x^2 \left(\frac{1}{\sqrt{\left(x^2 - z\right)^2 + 1}} - \frac{1}{x^2}\right) \tag{2.45}$$

$$I_2(z) = \int_0^\infty dx\, x^2 \left(1 - \frac{x^2 - z}{\sqrt{\left(x^2 - z\right)^2 + 1}}\right). \tag{2.46}$$

These equation can either be solved analytically in terms of elliptical integrals [73] or, alternatively, numerically by inverting equation (2.43) and inserting into equation (2.44) to get $\frac{\Delta}{E_F}\left(\frac{1}{k_F a}\right)$. This in turn can be used to calculate $\frac{\mu}{E_F}\left(\frac{1}{k_F a}\right)$. The numerical mean-field solutions for the gap and the chemical potential are shown in figure 2.4. However, these solutions are known to overestimate Δ, as for large interactions higher order processes play a non-neglectable role [72]. Nonetheless, they provide a reference and show that the transition from the BEC to the BCS regime across unitarity is smooth.

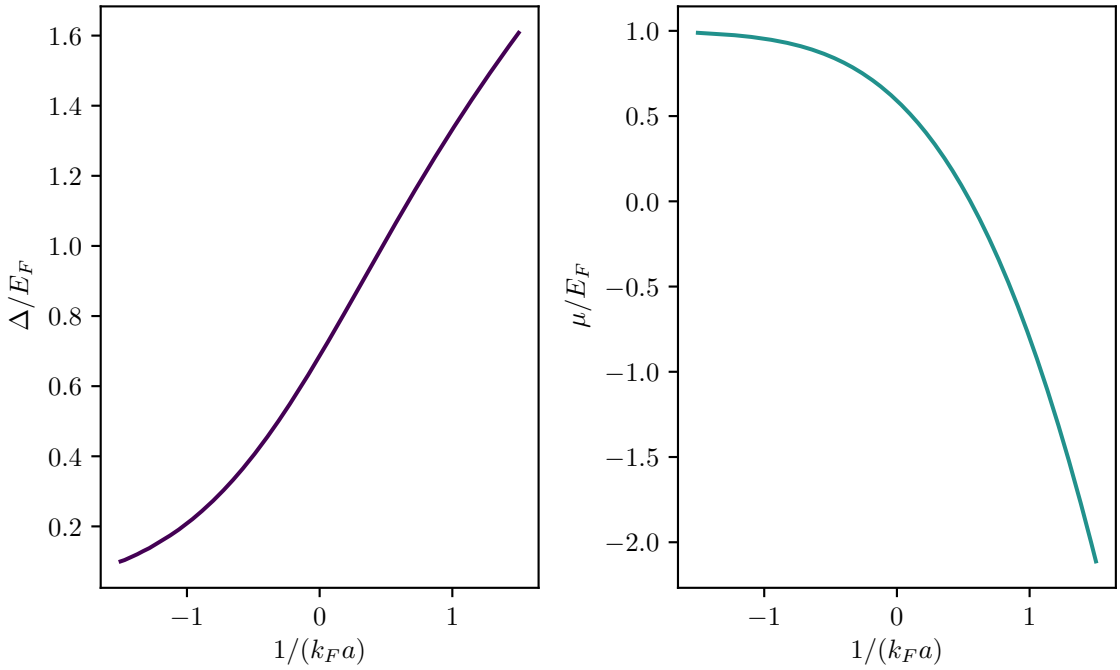

Figure 2.4: Mean field solutions. Mean field solution for the superfluid gap Δ and the chemical potential μ in units of the Fermi energy E_F versus interaction parameter $1/(k_F a)$ at $T = 0$.

2.4 Interacting Fermi gases at finite temperatures

In the end, to explore the phase diagram of the BEC-BCS crossover, two separate parameters are of interest: interaction and temperature. The effect of different interactions and their experimental realisation by means of Feshbach resonances has been investigated in the previous section. Now, the focus will be on the effect of non-zero temperature. In case of finite temperature, single particle excitations in the system play a non-vanishing role and alter the dynamics of the system. In order to understand how these excitations change the dynamics, we first have to focus a bit more on the process of single particle excitations.

2.4.1 Single particle excitation spectrum

The effect of the quasiparticle operator $\gamma^\dagger_{\mathbf{k},\uparrow}$ on the BCS ground state given in equation (2.38) is

$$\gamma^\dagger_{\mathbf{k},\uparrow} |\Psi_{\mathrm{BCS}}\rangle = c^\dagger_{\mathbf{k},\uparrow} \prod_{\mathbf{l}\neq\mathbf{k}} \left(u_{\mathbf{l}} + v_{\mathbf{l}} c^\dagger_{\mathbf{l},\uparrow} c^\dagger_{-\mathbf{l},\downarrow} \right) |0\rangle \,. \tag{2.47}$$

This constitutes the removal of a fermion pair at $\mathbf{k}, -\mathbf{k}$ and the addition of a single fermion with momentum $\mathbf{k}$ (and spin $\uparrow$). Since the Hamiltonian is diagonal with respect to the quasiparticle-operators, we can see from equation (2.36) that the corresponding energy is

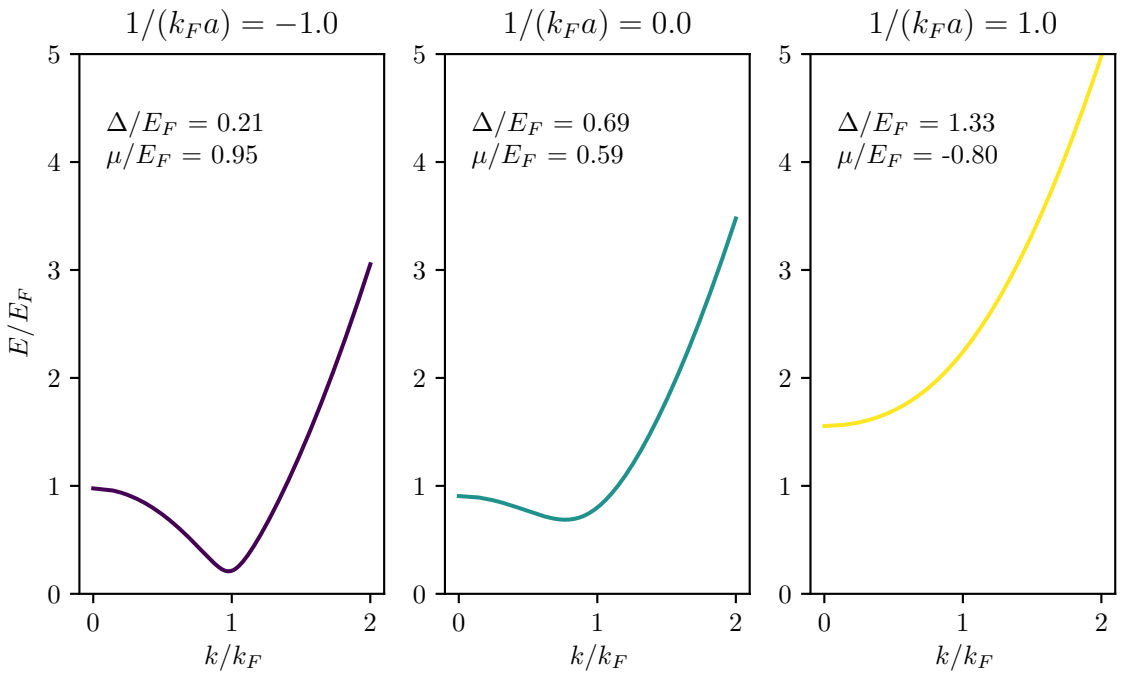

Figure 2.5: Single particle excitation. Energy of single particle excitation versus k/k_F for different interaction parameters $1/(k_F a)$. Indicated gap Δ and chemical potential μ are calculated from the mean field solution in figure 2.4.

$E_{\mathbf{k}} = \sqrt{\epsilon_{\mathbf{k}}^2 + \Delta^2}$. For a Fermi gas in the BEC regime with $\mu < 0$ the minimum energy to remove a fermion is therefore $\sqrt{\mu^2 + \Delta^2}$ and occurs at $k = 0$. On the BCS-side ($\mu > 0$), the minimum energy occurs at the Fermi surface $k = k_F$ and is Δ. The excitation spectrum across the crossover is shown in figure 2.5. Interestingly, for the BCS superfluid the energy to remove a particle with momentum $-\mathbf{k}$ close to the Fermi surface is equal to adding a particle of momentum $\mathbf{k}$ and costs energy. This is known as particle-hole symmetry. A consequence of this fact is phase separation in polarised Fermi gases, as the majority fermions need to *pay* energy in order to enter the superfluid.

2.4.2 The gap and critical temperature

At finite temperature, single particle excitations are present. The pairing field $\langle c_{\mathbf{k},\uparrow} c_{-\mathbf{k},\downarrow} \rangle$ then modifies to take the number of quasi particles into account

$$\langle c_{\mathbf{k},\uparrow} c_{\text{-}\mathbf{k},\downarrow} \rangle \to -u_{\mathbf{k}} v_{\mathbf{k}} \left(1 - \langle \gamma^\dagger_{\mathbf{k},\uparrow} \gamma_{\mathbf{k},\uparrow} \rangle - \langle \gamma^\dagger_{\mathbf{k},\downarrow} \gamma_{\mathbf{k},\downarrow} \rangle \right). \tag{2.48}$$

The gap equation (2.40) has then to be rewritten to yield [11, 51, 74]

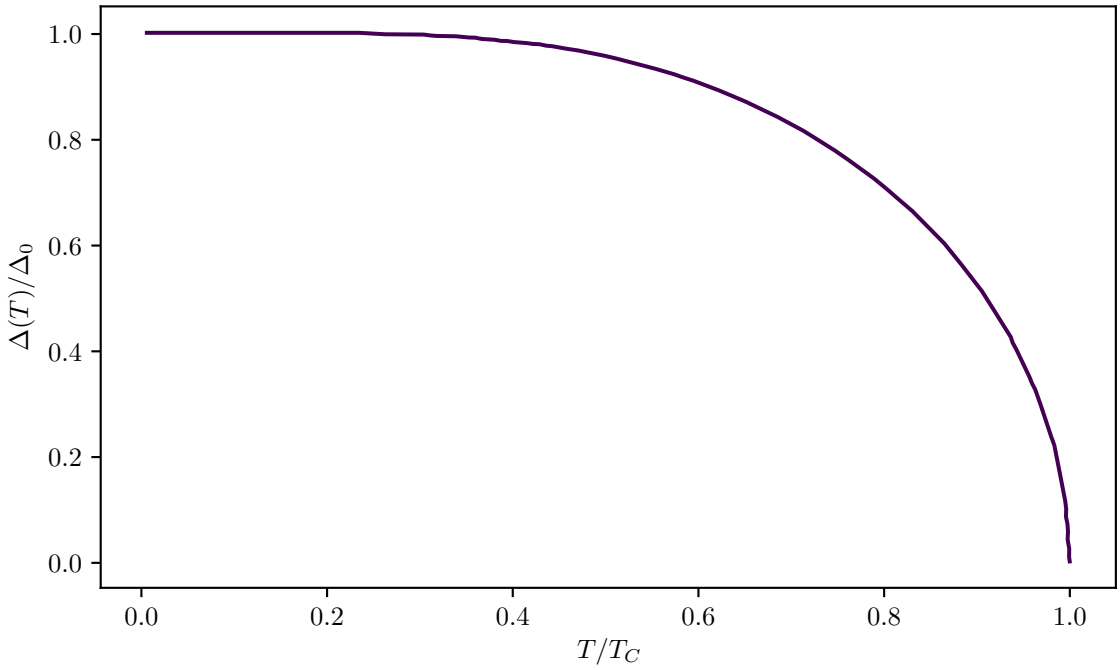

Figure 2.6: Temperature dependent gap. Temperature dependence of the gap $\Delta(T)$ in units of zero-temperature gap Δ_0 versus T/T_C. Data adapted from ref [51].

$$-\frac{m}{4\pi\hbar^2 a} = \int \frac{d^3k}{(2\pi)^3}\left(\frac{1}{2E_k}\tanh\left(\frac{E_k}{2k_BT}\right) - \frac{1}{2\epsilon_k}\right). \tag{2.49}$$

From this equation it is possible to calculate a temperature dependent gap, as shown in fig 2.6 for the BCS regime [51]. For $T \ll T_C$ the gap approaches $\Delta(T) \approx \Delta_0 - \sqrt{2\pi\Delta_0 k_B T}e^{-\Delta_0/(k_BT)}$. Close to T_C, where $T_C - T \ll T_C$, it becomes $\Delta(T) \approx \sqrt{\frac{8\pi^2}{7\zeta(3)}}k_B T_C\sqrt{1-\frac{T}{T_C}}$. By setting $\Delta = 0$ and using that far in the BCS regime the energy to add a particle is close to the Fermi energy $\mu \approx E_F$, we can calculate the characteristic temperature T^* at which the gap vanishes. Far on the BCS side, this represents the critical temperature at which pairing occurs,

$$T^*_{\mathrm{BCS}} = T_{\mathrm{C,\,BCS}} = \frac{e^\gamma}{\pi}\Delta_0 = 0.613\,T_F\,e^{-\pi/(2k_Fa)}. \tag{2.50}$$

Here, Δ_0 is the zero-temperature gap value and $e^\gamma = 1.78$. Since Δ_0 decreases exponentially with increasing $1/(k_Fa)$, so does the critical temperature. Beyond mean-field, this critical temperature is corrected by a multiplicative factor due to screening via particle-hole fluctuations. This is known as Gorkovs effect and the critical temperature then becomes [72]

$$T_C \approx 0.277\, T_F\, e^{-\pi/(2k_F a)}. \tag{2.51}$$

In order to get the critical temperature across the crossover, we have to include the uncondensed pairs into the calculation [11, 74, 75]. Far on the BEC side, the critical temperature becomes

$$T_{\mathrm{C,\ BEC}} = \frac{\pi\hbar^2}{m}\left(\frac{n}{2\zeta(\frac{3}{2})}\right)^{2/3} = 0.22\, T_F, \tag{2.52}$$

with $\zeta\left(\frac{3}{2}\right) \approx 2.612$. This is exactly the critical temperature for Bose-Einstein condensation of bosons with mass $m_B = 2\,m$ at a density of $n_B = n/2$ [76]. This is a consequence of the fact that far in the BEC-regime the binding energy becomes large and the bound atoms act like one composite boson across all other energy scales. In the weakly attractively interacting case, the critical temperature has a small positive correction [77, 78]

$$\frac{\Delta T_C}{T_{\mathrm{C,BEC}}} \approx 1.31 a_{ad} n^{1/3}.$$

The crossover should now smoothly connect the two regimes, meaning there must be a maximum in between [38]. The critical temperature in the crossover has been calculated in several ways, such as functional renormalisaton theory [20] and Monte Carlo methods [79, 80], but also analytically [11, 75]. T_C and T^* across the BEC-BCS-Crossover for a homogeneous system are shown in fig 2.7. Although being explored theoretically, an accurate measurement of the critical temperature with an independent thermometry is still missing.

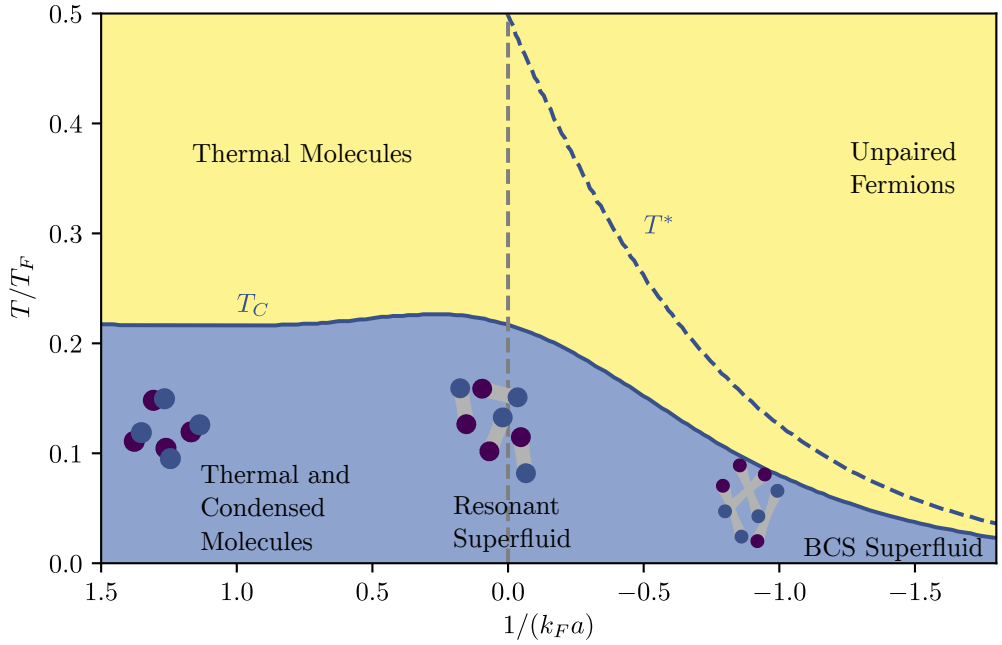

Figure 2.7: BEC-BCS phase diagram. The critical temperature T_C and pairing temperature T^* across the BEC-BCS crossover. Data adapted from [10, 11]. More recent theoretical approaches will be discussed in chapter 5.

2.4.3 Condensate fraction and superfluid fraction

There is an important distinction between the superfluid density and condensate fraction in the BEC-BCS Crossover. The superfluid density n_s describes the number of atoms that partake in the superfluid behaviour of the sample, namely the frictionless flow. The condensate fraction N_c on the other hand describes the fraction of actually Bose-condensed particles. At zero temperature, independent of interaction, the superfluid density always becomes the total density. The condensate fraction at $T = 0$ can be calculated by evaluating the number of condensed particles via [82]

$$N_0(T=0) = \lim_{T\to 0} \sum_{\mathbf{k}} u_{\mathbf{k}}^2 v_{\mathbf{k}}^2 \tanh\left(\frac{\beta E_{\mathbf{k}}}{2}\right) = \sum_{\mathbf{k}} u_{\mathbf{k}}^2 v_{\mathbf{k}}^2. \tag{2.53}$$

Inserting the expressions for $u_{\mathbf{k}}$ and $v_{\mathbf{k}}$, we obtain an expression for the condensate density at zero temperature [81]

$$n_0 = \frac{N_0}{V} = \left(m^{3/2}/8\pi\hbar^3\right) \Delta^{3/2} \sqrt{\mu/\Delta \sqrt{1+\mu^2/\Delta^2}}. \tag{2.54}$$

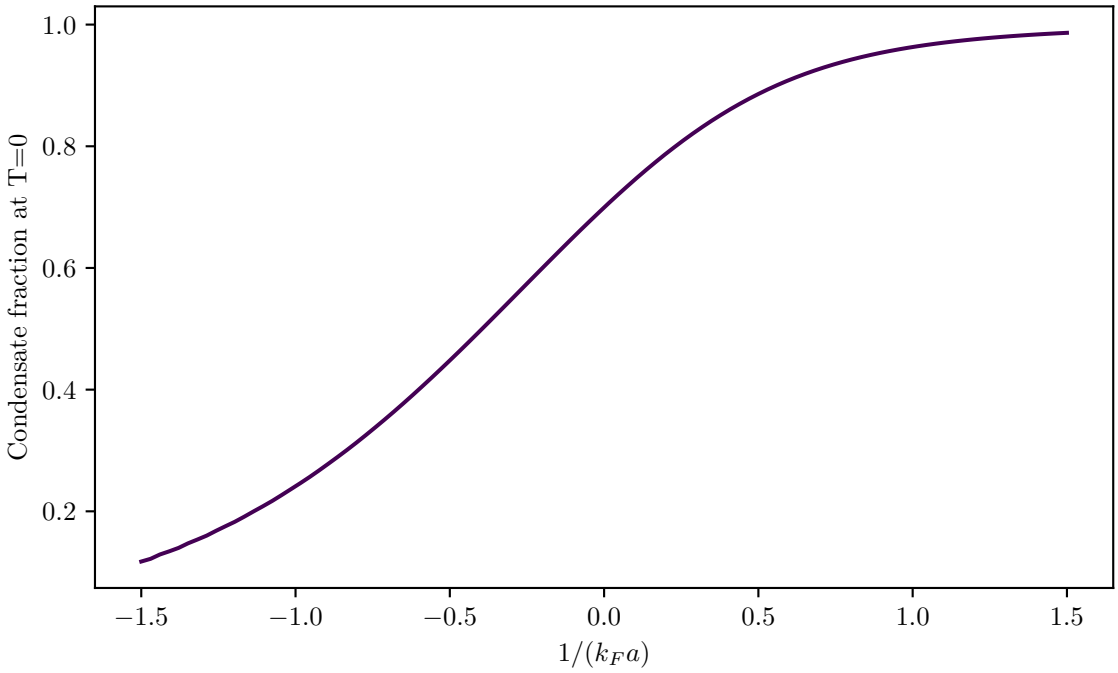

Figure 2.8: Zero-temperature condensate fraction. Condensate fraction at $T = 0$ across the BEC-BCS crossover within mean-field approximation [81]. In the BEC limit the condensate fraction tends to 1, whereas in the BCS limit it depletes exponentially.

The result for the condensate fraction $N_0/2N$ within mean field theory is plotted in figure 2.8. In the BEC-regime, where $1/(k_F a) \gg 1$ and $\mu/\Delta \ll -1$, the number of Bose condensed pairs approaches the number of atoms divided by two, which means all atoms are condensed. The condensate density then is given by half the total density $n_0 \approx n/2$. On the BCS side with $\mu/\Delta \gg -1$ and $1/(k_F a) \ll -1$, the density is proportional to the gap and therefore exponentially vanishes $n_0 \approx (m k_F/8\pi\hbar^2)\Delta$. As a consequence, the condensate fraction exponentially approaches zero as well. At unitarity, with $\mu \approx 0.59 E_F$ and $\Delta \approx 0.69 E_F$ (see figure 2.4), the condensate fraction at zero temperature is given by $2N_0/N = 0.6994$. From equation (2.53) without the limit $T \to 0$, the temperature dependence of the condensate fraction can be calculated. However, this has to be taken with a grain of salt, as the mean-field solution overestimates T_C at unitarity and in the BEC regime [11, 75, 83, 84]. For $1/(k_F a) = -1$, the temperature dependence of the condensate fraction is plotted in figure 2.9. The superfluid fraction can be obtained by investigating the role of fermionic and bosonic excitations. It can be shown, that the superfluid density always approaches unity, when $T \to 0$. The temperature dependence of the superfluid density is also shown in figure 2.9.

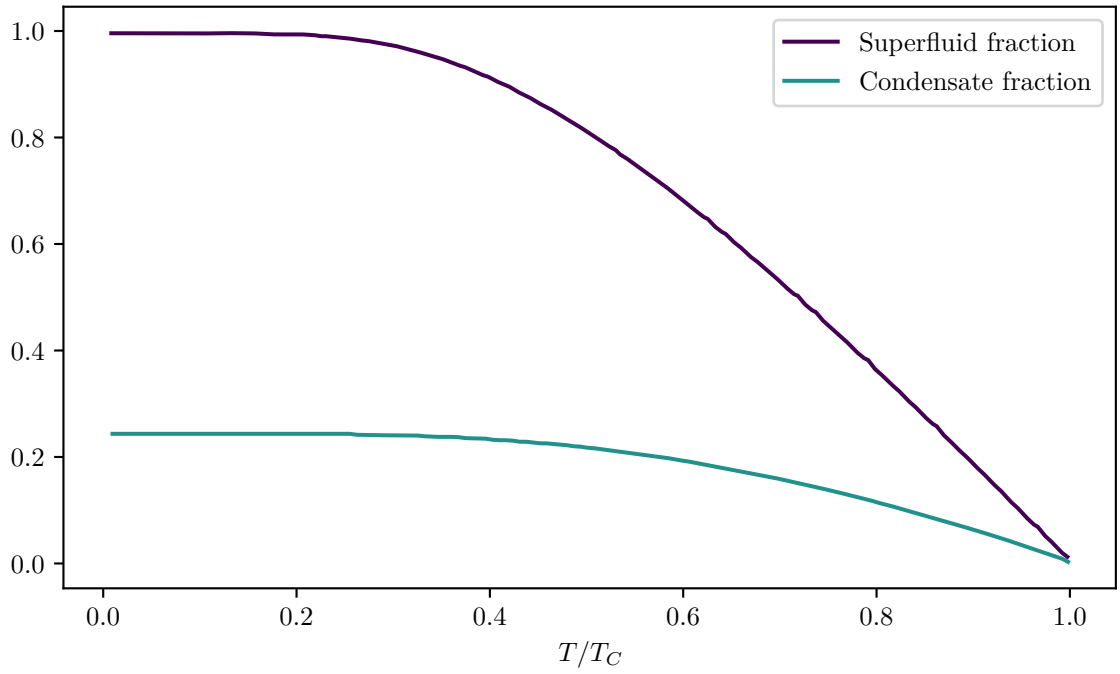

Figure 2.9: Condensate fraction and superfluid fraction at $1/(k_F a) = -1$ versus temperature. Data adapted from ref [51].

CHAPTER 3

Machine Learning and Neural Networks

This chapter aims to provide a concise introduction to neural networks within the context of this thesis. As many techniques are rather novel to the field of physics, the goal is to establish all machine learning related concepts used in later chapters on a level accessible to a physicist without prior experience. The structure of the chapter is to first give a brief overview of the history of neural networks, followed by an overview with a very simple but instructive example. This is followed by a quantitative introduction to neurons and neural networks in general. The process of supervised learning with such a network is explained in the subsequent section. Two techniques from unsupervised learning – principal component analysis and autoencoders – are explained in the last section.

3.1 A very brief history of key achievements and recent progress

Although the interest in artificial neural networks sparked exponentially in the last years, the idea of artificial neurons itself was already developed in 1943 by Warren McCulloch and Walter Pitts, who proposed a simple linear threshold model which has become famous as the McCulloch-Pitts neuron [85]. This neuron constitutes a simple function, which outputs 1 if the sum of all inputs is equal or larger than a threshold value and zero otherwise. Additionally, the neuron has inhibitory inputs that set the output to 0 when activated. In 1949 Donald Hebb developed a neuroscientific theory of learning, which would become famous as hebbian learning, named after its discoverer [86]. The theory is often shortly formulated as *what wires together, fires together*, which essentially describes the interplay between neurons in the brain: If neurons fire together repeatedly, they become associated. By modelling the connections between neurons in this way, learning should become possible. Combining both concepts in 1957, Frank Rosenblatt showed that shallow networks of a certain kind of neuron with threshold behavior called perceptron were able to learn linearly separable models [87]. Although promising, the interest in neural networks froze, when in 1969 Marvin Minsky and Seymour Papert proved that the perceptron is severely limited in its capabilities and that Rosenblatts method would not work with several layers of perceptrons that would be needed to learn more complicated problems [88]. This halted research for a period often referred to as *AI winter*. The interest in neural networks returned when the idea of backpropagation was developed in the 1970s [89, 90] and popularised through a series of important articles

in 1986 [91, 92]. Backpropagation, as described in section 3.4.2, provides the framework to train deep, multi-layer networks of neurons that act as universal function approximators [93, 94]. From this point on, neural networks became widely used to solve real world problems, starting, most prominently, with handwritten digit recognition with convolutional networks [95, 96]. Since then, many important discoveries were made. Finally, with the development of the internet and the accessibility of large datasets, neural networks have a arrived as an essential tool in many fields, including physics [23]. Sparked by recent breakthroughs in machine learning, covering image recognition [97], reinforcement learning [98, 99] and natural language processing [100, 101], neural networks have been used for quantum state tomography [102], the detection of phase transitions [31, 103], optimisation of experimental systems [47–50] and modeling of wave functions [104], among others.

3.2 Overview

The idea behind neural networks is remarkably simple and can in many cases be reduced to an optimisation problem or fitting routine. A neural network can be understood as a function that maps a given input to an output using a set of network-specific parameters. The problem then is finding the right parameters for this function, such that the mapping reliably reproduces the desired output given a specific input. Inputs can for example be pictures of handwritten digits, the audio wave of a speech recording or physics (image) data. The output could then be the digit which is seen on the picture, a transcript of the recording or certain physical quantities determined from the input data, respectively. It is easy to imagine that such a function generally is very complicated. Taking as an example a network which processes black-and-white pictures, the dimension of the input is the number of pixels contained in the pictures. For a 100×100 pixel picture, this constitutes to a 10 000 dimensional input on which the network-function operates. The number of parameters describing the function that is modelled by the network is potentially even orders of magnitude larger, as the number of parameters can grow quadratically with the complexity of the network. Consequently, such a function cannot be proposed ad hoc, but has to be generated by *learning* the parameters describing the desired mapping. This can be done by using example inputs with known outputs, a process called *supervised learning*. In simple terms, the idea of supervised learning is to define a loss function that is minimised for a correct prediction and change the parameters of the function such that the loss is minimised. This constitutes to finding a minimum in a high dimensional space.

3.2.1 A simple example

Before formalising neural networks it is useful to illustrate an example to understand the general interplay between the different components of a neural network. This will introduce the terminology and the basic concept. The example is inspired by ref [105], which also presents a much more thorough and hands-on introduction to deep neural networks than is possible in the scope of this thesis.

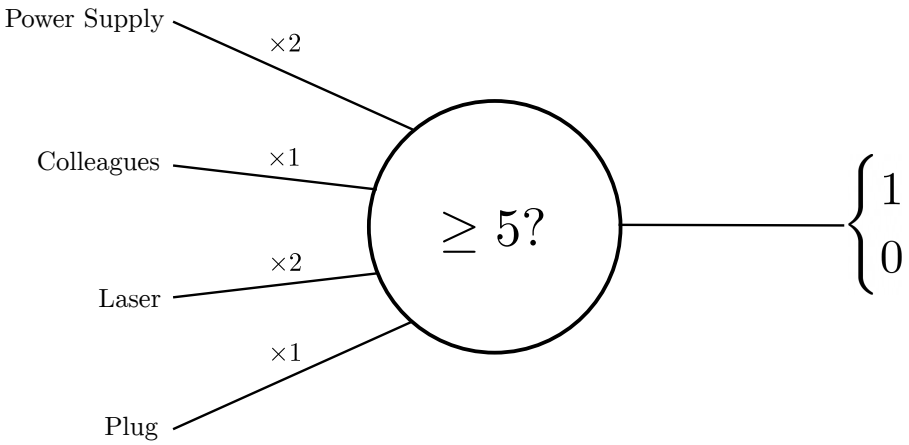

Figure 3.1: A single neuron as a decision machine. Positive input conditions are set to 1, negative conditions are set to 0. In order to determine the neuron output a weighted sum is calculated. If the value of this sum is above the threshold, the neuron fires (output is 1).

How to make a decision

For this example we look at a simple model of an artificial neuron with threshold behaviour: the perceptron. The perceptron has one or more weighted inputs and a single output. If the sum of its entries is larger or equal than some threshold value it will output 1 (fire), if it is smaller it will output 0. We can use this behaviour to make very simple decisions, for example, whether the experiment is running well. For this, we assume an output of 1 means the experiment is running well, and an output of 0 means it is not. We furthermore (quite unrealistically) assume, there are only four important factors:

1. The power supplies are working properly,
2. There are two or more people present in the lab,
3. The lasers are working properly,
4. The plug beam is aligned well.

Those conditions are the four inputs to the neuron and if they are met, the corresponding input is set to 1, and 0 if it is otherwise. In order to make an informed decision about if the experiment is running well, we need to weight these inputs. It is for example probably less important that there are enough people present, or the plug beam is aligned well, than it is that the power supplies or lasers are working properly. Since they are more important, we weight the latter two inputs with a factor of 2 and choose the threshold value (usually referred to as bias) to be 5, see figure 3.1. This represents a situation, where the power supplies as well as the lasers have to work well and at least one of the other two conditions has to be fulfilled. In the case, where point 1, 2 and 3 are fulfilled, the weighted sum becomes

$$2 \cdot 1_{\text{PowerSupplies}} + 1 \cdot 1_{\text{Colleagues}} + 2 \cdot 1_{\text{Laser}} + 1 \cdot 0_{\text{Plug}} = 5 \geq 5 \rightarrow \text{Neuron fires!}$$

In the case, where the lasers are not working properly but all other factors are met, the output reads

$$2 \cdot 1_{\text{PowerSupplies}} + 1 \cdot 1_{\text{Colleagues}} + 2 \cdot 0_{\text{Laser}} + 1 \cdot 1_{\text{Plug}} = 4 < 5 \rightarrow \text{Neuron does not fire!}$$

This shows, that the neuron has the wanted behaviour and is able to predict if the experiment is running well.

A more complex decision

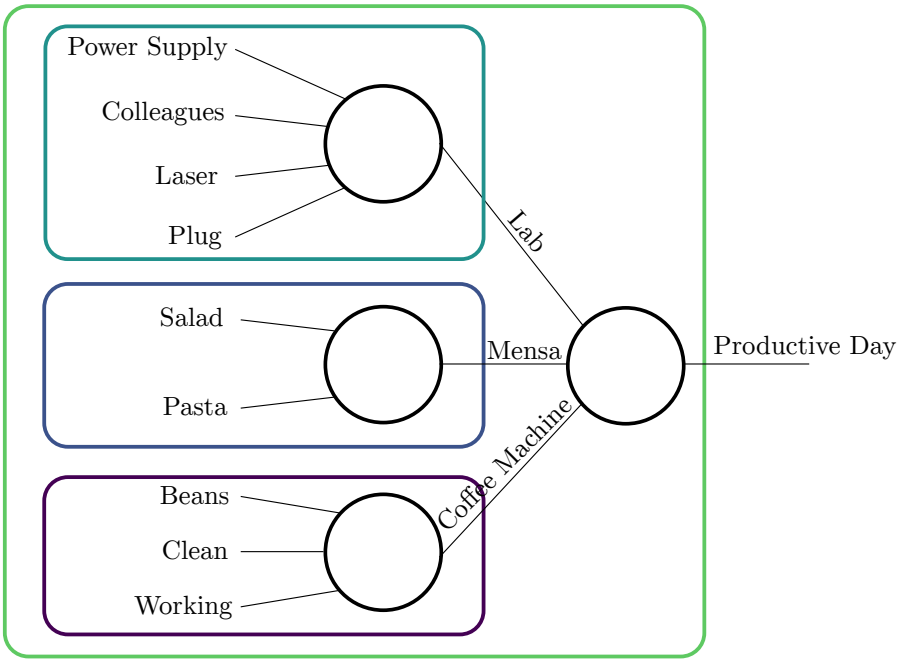

Figure 3.2: A neural network for a more complex decision. The inputs of the neuron that decides whether the day was productive or not are outputs of neurons themselves. Every box represents a separate decision process.

It might be that we want to make a more complex decision which is hard to represent by a single neuron, for example because some pre-condition has to be met, which in itself has condition it depends on. We can represent such a more complex decision by a *network* of neurons, where the output of one neuron can form the input of another neuron. For example, we could decide whether it was a productive day in the lab. For this, say, three requirements have to be met:

1. The experiment worked properly, which depends on the above points.
2. The food in the Mensa was good, which happens if

a) The salad bar was opened,

b) The vegetarian food offer contained pasta.

3. The coffee supply was good, which might depend on

a) There were enough coffee beans,

b) The coffee machine was clean,

c) The coffee machine worked as intended.

Each of the above points constitutes a neuron (= a decision) by itself, and the output of those neurons forms the input for the neuron that decides whether the day was productive or not, see figure 3.2. This kind of decision process is the key to how artificial neural networks work. An even more complex decision than the above for example might be to decide, whether the time-of-flight picture of an atomic cloud shows a condensate or not. In this case, the values of the pictures pixel represent the inputs. However, in this case it is impossible to choose the weights by hand, as we did above. Not only because the networks tend to get very complicated with many intermediate decisions and contain a lot of weights, but also because it is not inherently clear what the weights should look like, i.e. which pixel is how important for the detection of a condensate? The solution for this problem is to employ a learning process. By using known input-output pairs, it is possible to adjust the weights of the network incrementally, such that the mapping between the input-output pairs can be represented by the network. The assumption is then that the network is able to generalise the observations from the training data to unknown data. This process is known as supervised learning and is essential to train neural networks. In the following chapters, all of the above considerations will be formalised and the technique for tuning the weights will be introduced.

3.3 Neural networks

Neural networks are collections of artificial neurons. This section will first introduce the building blocks of these networks: artificial neurons. In the next step, we will assemble collections of these neurons to achieve neural networks. Finally, we will explain convolutional and max pooling layers which are often used for image processing tasks.

3.3.1 An artificial neuron

The basic building blocks of neural network are artificial neurons. Artificial neurons generate an output y by first calculating the vector product between an input vector $\mathbf{x}$ and a weight vector $\mathbf{w}$ and adding a bias b. Mathematically, the bias can be understood as the weight for an additional input neuron that is always set to 1. Using the resulting weighted sum as an argument to an activation function $\Phi(z)$ gives the neurons output

$$y = \Phi\left(\sum_i w_i x_i + b\right). \tag{3.1}$$

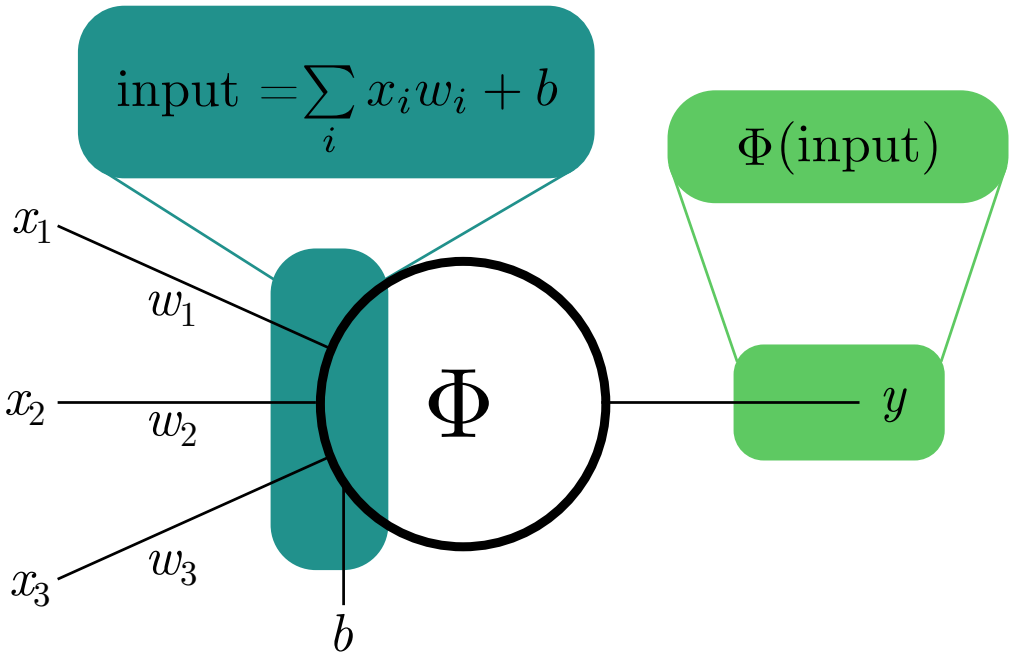

Figure 3.3: Structure of an artificial neuron. An artificial neuron with inputs x_i, weights w_i and bias b. The output of the neuron is given by evaluating the activation function Φ with the weighted sum of inputs as argument.

A sketch of a general neuron is shown in figure 3.3. In principle, the activation function can be arbitrary as long at is non-linear and differentiable. In the example from section 3.2.1, the activation function is the Heaviside function[1] – in this case, the neuron is often called perceptron [87]. Another popular choice is to use a sigmoid function $f_\sigma(x) = \frac{1}{1+e^{-x}}$. However, it has been shown that deep neural networks can be trained faster if the activation function is the rectified linear unit[2] [106]:

$$\mathrm{ReLU(x)} = \max(x, 0).$$

3.3.2 Fully connected neural networks

Neural networks are collections of artificial neurons arranged in layers, as sketched in figure 3.4. Every layer represents a number of neurons arranged as a vector, such that the output of any given layer is a vector of neuron outputs, which can serve as the input for the next layer. Every neuron of a layer receives every entry from the output vector of the previous layer as a weighted input. As a consequence, between a layer with n neurons and a layer with m neurons, there are $n \times m$ individual weights. Since the input to a neurons activation

[1] The Heaviside function is not differentiable at the position of the step, but its derivative is set to zero by convention at this point.
[2] The derivative at $x = 0$ is set to zero by convention.

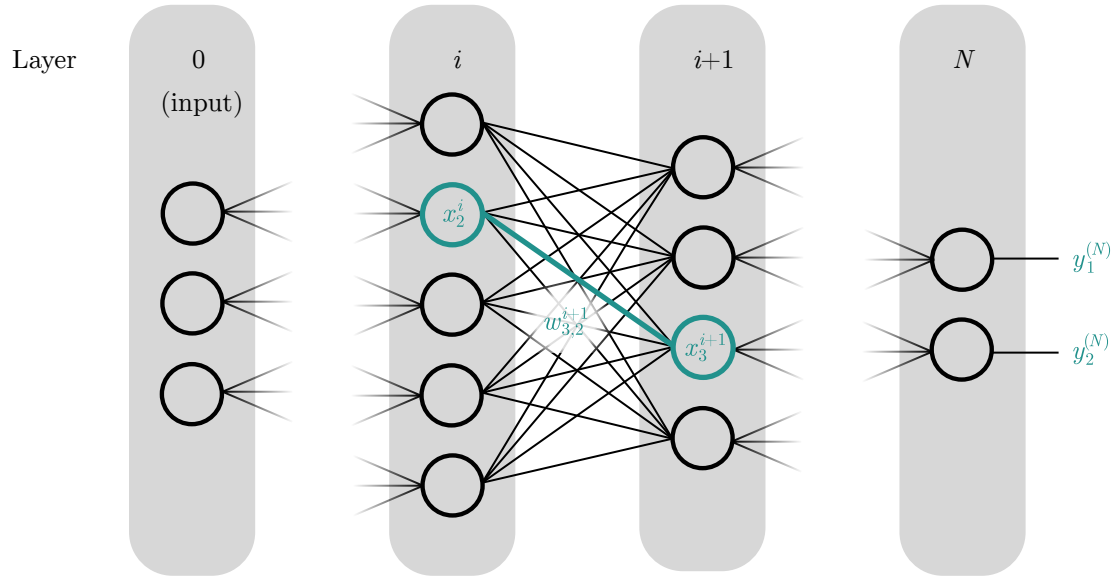

Figure 3.4: An artificial neural network. The output of each neuron in the ith layer is connected with the input of every neuron in the (i+1)th layer. The strength of the connection is characterised by the weights. For clarity, the bias is not displayed here. It is represented by a neuron with subscript zero without any connections to the preceding layer.

function can be interpreted as a scalar product (see equation (3.1)), it is straightforward to formulate the weights connecting two layers $i-1$ and i as matrix. The output of neuron j in layer i can then be rewritten in terms of the weight matrix and preceding layer vector as

$$y_j^i = \Phi\left(\sum_k w_{jk}^i x_k^{i-1} + b_j^i\right) \equiv \Phi(z_j^i), \tag{3.2}$$

where w_{jk}^i is the k-th entry in the j-th row of the weight matrix $\hat{\mathbf{w}}^i$ connecting layers $i-1$ and i whose rows are the weight vectors $\mathbf{w}$ of the neuron inputs. Note, that in order to change the offset of the weighted sum an additional *bias neuron* is inserted in every layer. This neuron has no connections to the preceding layer and is always set to 1. The value of the bias is adjusted through the corresponding weights b_i^j connecting it to the next layer.

One can easily imagine that with enough layers and neurons per layer, the output of the total network – which is the output of the n-th layer for an n-layer network – can be become very sensitive to its inputs. In fact, it can be shown that artificial neural networks can approximate any continuous function [93, 94]. The task at hand is to find a set of weight-matrices $\hat{\mathbf{w}}^i$ that represent a function which maps the given inputs to their desired output. This means we need to find a set of parameters that describe the functional connection

between inputs and outputs for a function of a shape given by the neural network architecture.

3.3.3 Convolutional neural networks

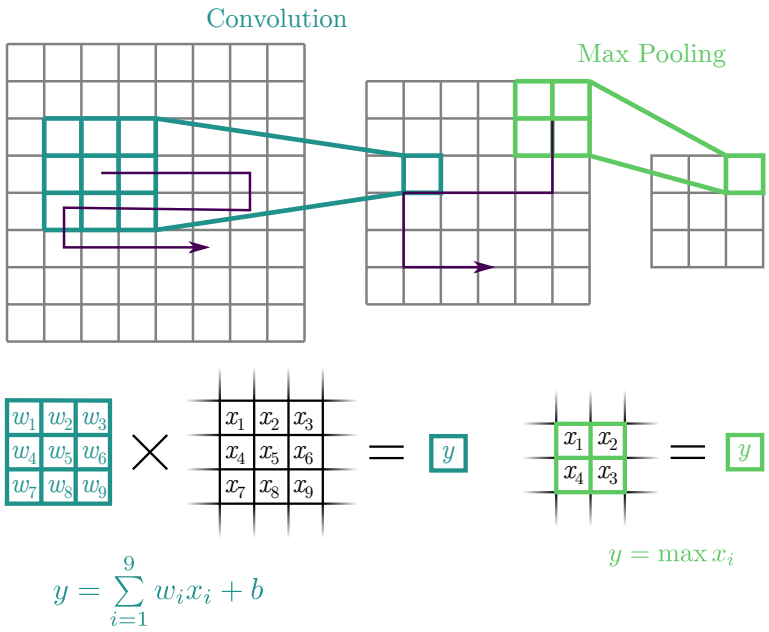

Figure 3.5: Concept of convolutional neural networks. Convolutional neural networks usually combine convolutional layers and max pooling layers. Convolution is indicated in turquoise. The turquoise window slides across the input picture (grey array, with stride=1) and produces the values of the output by calculating the convolution between filter mask and input entries. For max pooling (green) only the maximum value of the current window remains as output (here: stride=2).

Convolutional neural networks (CNNs) are a subclass of neural networks often used in image recognition. The idea behind CNNs is to model how the visual field interacts with the neurons in the visual cortex. The idea is based on the work of Hubel and Wiesel, who showed that in cats and monkeys certain neurons in the visual cortex are activated by certain windows in the visual field called the receptive field of a neuron [107]. A first realisation of this behaviour was realised with the Neocognitron [108], however, at the time there was no gradient-based learning mechanism. This was only developed later [95, 96] and meant a breakthrough for the application of CNNs [109]. Although inspired by biological processes, it has not been shown that the actual biological process can be represented by CNNs.

Generally, CNNs mimic the receptive field through small filters that are applied to the input. The exact shapes of these filters are learned during training. However, in order to mimic the receptive field properly, they are chosen small as compared to the input. The connection to the following neurons is then given by the convolution[3] between the current

[3] Actually, the word convolution, though name-giving for CNNs, is used inaccurately here. Most of the time

window on the input array (=the visual field) and the filter. The number of calculated convolutions for a given input is dependent on the number of filters, filter size, the padding of the input and the stride, see figure 3.5. Each convolution results in a weighted sum that acts as an input for an activation function which – analogously to section 3.3.1 – is often chosen to be the rectified linear unit (ReLU). Since the convolution operation is a weighted sum with the filter entries being the weights, it is easy to see that any learning algorithm that is applicable to fully connected layers is applicable to convolutional layers as well. As compared to fully connected layers, convolutional layers realise three important concepts for machine learning systems [110]: **Sparse connectivity** between layers, realised by a small filter and therefore potentially very few weights as compared to a fully connected layer. This does not only reduce memory and computing requirements, but also increases statistical efficiency. The second concept is **weight sharing**. As the filter is moved across the input array, the weighted sum is always calculated from the same weights defined by the filter. This means that in order to detect features (which need to be smaller than the filter size), it is not necessary to train a separate set of weights for every position. The third concept is **equivariance**, which means that if a feature on the input moves, the respective output feature will move accordingly.

Convolutional layers are usually combined with max pooling layers, which represent non-linear discretisation and downsampling of the inputs [109–111]. This means, a group of pixels on the input is represented on the output by just the pixel with the maximum value in the sample. This brings several advantages. For one, an obvious reduction of the layer dimensionality. Secondly, a natural increase of the receptive field, without the need to change the size of the convolutional filters. Lastly, max pooling layers reduce the memory requirements and amount of computation needed and, hence, reduce overfitting. The functionality of convolutional and pooling layers is sketched in figure 3.5.

3.4 Supervised learning

Supervised learning in the context of neural networks means optimising the weights of the connections, such that the neural network can represent some labelled dataset (inputs with known outputs). For this, two things have to be taken care of: Firstly, it has to be characterised how accurate the network output is as compared to the ground truth, which is represented by the labels. Secondly, an update rule for the weights is needed, that adjusts the total network output towards higher accuracy.

3.4.1 The loss function

The goal of the loss or cost function is to characterise how far away the current output of the network is from the label (=desired output). Consequently, its value should be small if the output is close to the ground truth and large if it is not. Throughout this thesis, the used loss function is simply the mean squared distance between the network output and the label

not an actual mathematical convolution is used, but a cross-correlation between filter and input array, however, this does only change the ordering of the weights in the filter.

$$E\left(\mathbf{y}, \mathbf{y}'\right) = \frac{1}{2}\|\mathbf{y} - \mathbf{y}'\|^2, \tag{3.3}$$

but generally the choice of the loss function is arbitrary, as long as it quantifies the distance between desired and actual output and can be written as an average over many examples. In case of classification tasks oftentimes the cross entropy is used as the loss function [110]. Here, $\mathbf{y}$ is the network output given an input $\mathbf{x}$ and $\mathbf{y}'$ is the label associated with $\mathbf{x}$. Note, that in general network output and label will never be equal, but the goal is to make them as close as possible. $\|\mathbf{a}\|$ denotes the euclidean norm of vector $\mathbf{a}$. That is because, in general, the output of the network can be multidimensional, corresponding to the number of neurons in the output layer. Since the network output $\mathbf{y}$ is dependent on all weights w^i_{jk}, the loss function can be seen as a very complicated function in a parameter space spanned by all weights. Since the function is positive definite and small losses mean good performance, the task is to find a (optimally global) minimum for the averaged loss function in weight-space.

3.4.2 Backpropagation

Although independently developed much earlier [110], the general idea of gradient based optimisation has been popularised in context of deep learning in the 1980's [91, 92, 112]. The general idea is to calculate the gradient of the loss function in weight space and update the weights with small increments in direction of the steepest descent. This will eventually change the weights in such a way that the average error is minimised. Unfortunately this technique will not necessarily find global minima, however, in many practical problems this does not play a role, as different local minima seem to be very close in performance [109]. The key to updating the weights with respect to the average loss function is to calculate the partial derivative of the loss function and use it to update the weights towards lower loss-values

$$w^i_{jk} \to w^i_{jk} - \eta \frac{\partial E}{\partial w^i_{jk}}, \qquad b^i_j \to b^i_j - \eta \frac{\partial E}{\partial b^i_j}, \tag{3.4}$$

where η is called the learning rate, determining the proportionality factor between the value of the gradient and the step-size towards the minimum. A small learning rate means small steps and therefore slower convergence and higher risk to get stuck in weak minima, but also better sampling of the minimum-position in parameter space.
The name backpropagation stems from the fact, that in order to calculate the above partial derivatives, we have to iteratively move backwards through the neural network applying the chain rule until the desired weight is reached. As the partial derivative for weights that are in lower layers contains the derivatives of higher weights, it makes sense to start with the output layer of the N-layer network $y = y^N$. In order to get $\partial E/\partial w^N_{jk}$, one can apply the chain rule and arrive at

$$\frac{\partial E}{\partial w_{jk}^N} = \frac{\partial E}{\partial y_j^N} \times \frac{\partial y_j^N}{\partial w_{jk}^N} = \frac{\partial E}{\partial y_j^N} \times \frac{\partial \Phi(z_j^N)}{\partial z_j^N} \times \frac{\partial z_j^N}{\partial w_{jk}^N}, \tag{3.5}$$

where $y_j^i = \Phi(z_j^i)$ was used in the last step. Since $z_j^i \propto w_{jk}^i y_k^{i-1}$ is just a weighted sum, the last term is simply y_k^{i-1}. In case of the biases, the above expression is similar with $w_{jk}^N \leftrightarrow b_j^N$, however, the last term always evaluates to 1, as the bias values are not dependent on the preceding layer. In the case of calculating the gradient with respect to weights which are not in the last layer, the last term again has to be expanded with respect to the preceding neurons output. However, the first two terms still show up in the total expression and therefore only have to be calculated once. This justifies the term *backpropagation*, as partial derivatives of layers with low i have to be calculated from the partial derivatives already calculated from layers with larger i. In order to write down an iterative form for backpropagation, it is useful to define the derivative of the loss function with respect to the inputs of each layer [105]

$$\delta_j^i = \frac{\partial E}{\partial z_j^i}. \tag{3.6}$$

The first two terms of equation (3.5) then simply become

$$\delta_j^N = \frac{\partial E}{\partial y_j^N} \times \frac{\partial \Phi(z_j^N)}{\partial z_j^N}, \tag{3.7}$$

and by expanding the partial derivatives for neurons in preceding layers, it is easy to show that

$$\delta_j^i = \left(\left(\hat{w}^{i+1} \right)^{\mathrm{T}} \delta^{i+1} \right)_j \times \frac{\partial \Phi(z_j^i)}{z_j^i}. \tag{3.8}$$

This is a recursive formula for the δ-terms. In order to get the full partial derivative of the loss function, the δ-terms have to be multiplied with the derivative of the neuron input with respect to the bias or weight. In case of the bias, as argued above, this term is always equal to 1 and hence

$$\frac{\partial E}{\partial b_j^i} = \delta_j^i. \tag{3.9}$$

In case of the weights the output of the preceding neuron has to be taken into account:

$$\frac{\partial E}{\partial w_{jk}^{i}} = y_k^{i-1}\delta_j^i. \tag{3.10}$$

Equations 3.7, 3.8, 3.9 and 3.10, together with the update rule 3.4 provide means to iteratively change the weights, such that a minimum in the loss function – and therefore a good approximation of the function mapping the inputs to the desired outputs – can be found.

Stochastic gradient descent

Backpropagation as mentioned above updates the weights based on any single labelled datapoint. The loss function in parameter space can however look different for different labelled examples. It is therefore necessary to find a minimum in an *average* loss function, as this will mean *on average* the network performs well. One technique to improve the sampling is to use small batches of data (called mini batches) and calculate the average gradient over these batches, as this will give an approximation of the averaged loss function [105]

$$\bar{E} = \frac{1}{N}\sum_i E(\mathbf{y}_i, \mathbf{y}_i'). \tag{3.11}$$

It is not necessary to calculate the average loss for all labelled datapoints for every update, but instead use mini batches of sufficient size to approximate the average loss to save space and time. The size of the mini batches can be tuned as a meta parameter of the network.

Adam

One of the main reasons for the success of gradient-based back propagation is its simplicity: For many activation functions, calculating the gradient of the activation function is equally efficient to evaluating the function itself. The backpropagation process is therefore as efficient as the feed forward process. To optimise the process of finding a minimum in weight space, sometimes it might be more efficient to scale the learning rate appropriately to the (past) gradient values to make larger steps in weight-space, where no high resolution is needed. Adam (short for adaptive momentum estimation) [113] is an adaptive learning rate algorithm specifically designed for deep learning applications. Adam uses estimations of the first and second moment of the gradient to scale the learning rate appropriately. It does so by calculating exponentially moving averages on the gradient. The update rule for the weights then becomes

$$t \leftarrow t + 1 \tag{3.12}$$
$$g_t \leftarrow \nabla_w E_t(w_{t-1}) \tag{3.13}$$
$$m_t \leftarrow \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{3.14}$$
$$v_t \leftarrow \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{3.15}$$
$$\hat{m}_t \leftarrow m_t / (1 - \beta_1^t) \tag{3.16}$$
$$\hat{v}_t \leftarrow v_t / (1 - \beta_2^t) \tag{3.17}$$
$$w_t \leftarrow w_{t-1} - \eta \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon) \tag{3.18}$$

Here, t is the time step, $\hat{m}_t$ and $\hat{v}_t$ are the bias-corrected estimations of the first and second moment estimation m_t and v_t with $m_0 = 0$ and $v_0 = 0$. $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$ are constants that empirically work well and oftentimes remain unchanged. η is the step size parameter which is usually tuned as a meta-parameter depending on the application or dataset. The gradient is stochastically calculated over mini-batches. More details on the performance and explanations can be found in refs [113, 114]. Note that in order to use the Adam update rule, still only the partial derivatives with respect to the loss function have to be calculated. This means, the efficiency as compared to regular gradient descent does not change.

3.4.3 Overfitting and regularization

Overfitting happens when a network not only learns the information it needs to generalise to unseen examples, but also learns specifics of the training dataset that are not necessary for generalisation, such as noise. For the training set the error rate will then still decrease, however, for unseen data the error rate might actually rise. This can be detected by monitoring the error rate on a randomly chosen subsample of the training data that is not used for training. There are several methods to prevent overfitting in the first place and increase the ability of networks to generalise. Collectively, these methods are called regularisation methods [110]. One way of regularisation is to average the output of many different networks trained on the same data. However, with big datasets and long training times this is oftentimes unfeasible. Regularisation throughout this thesis is performed by removing randomly chosen neurons and their connections from the network, so for each training case in a mini-batch only a thinned out neural network is sampled. It can be shown, that sampling a network trained this way is similar to sampling many networks and averaging the output. This method is called dropout regularisation and has been shown to be a very efficient and resource saving way of regularisation [115, 116].

3.4.4 The networks in this thesis

The networks employed in chapter 5 are used to analyse data from pictures. As argued above, it is quite natural to use convolutional neural networks in this case, as the detection of local features, independent of their position and the use of correlation between close pixels as opposed to far away pixels might prove useful. For these reasons, the networks

are all combinations of several convolutional plus pooling layers with few fully connected layers before the output. The exact filter-size, number and architecture is mostly designed empirically and will be mentioned in the respective section of the thesis and appendix A.1.

In section 9 the situation is slightly different. Here, the input dimension is comparably low and therefore convolutional layers are not needed. We use the smallest possible fully connected network which manages to represent the data reasonably well. Again, the exact architecture will be mentioned in the according section.

3.5 Unsupervised learning

As opposed to the supervised learning methods presented in section 3.4, where the learning effect comes from optimising a system with pre-labelled data, the goal of unsupervised learning is to infer information from the unlabelled data itself. Goals of unsupervised learning tasks can be to infer the probability distribution that generated a dataset, cluster the data according to its similarity, generate new data based on the inferred distribution, or reduce the dimensionality and find suitable data representations [110, 117]. Chapter 6 will make use of an autoencoder to look at the underlying structure and similarity of a dataset and infer the critical point of a phase transition. Chapter 9 will be concerned with finding new datapoints that can be explained with an underlying structure in the learning data found by principal component analysis. Both methods will be introduced in the following sections.

3.5.1 Principal component analysis

Principal component analysis (PCA) [118, 119] is an unsupervised learning technique based on a linear coordinate transformation that changes the axes to represent the highest variance in a given dataset. Intuitively, this transformation can be thought of as centering the data to the center of a coordinate system and drawing a multi-dimensional ellipse around it. The principal axes of this ellipse, starting from the largest one, represent the new axes and are called principal components. By representing the data with a number of principal components smaller than the original dimensionality, it is possible to reduce the dimension of the dataset, but still retain most of the information.

More accurately, we look for a coordinate system, in which the covariance matrix $\hat{C}$ is diagonal. More information on how to calculate PCA can be found in refs [120, 121]. We assume k n-dimensional datapoints, arranged in vectors $\mathbf{x}_i$, or, written as a matrix $\hat{X}$, where the $\mathbf{x}_i$ represent the columns. For every datapoint the mean has to be subtracted - this resembles moving the datapoints to the center of the coordinate system in order to draw the ellipse from the intuitive picture above.

$$\tilde{\mathbf{x}}_i = \mathbf{x}_i - \bar{\mathbf{x}} \tag{3.19}$$

With this, the next step is to calculate the covariance matrix $\hat{C}$, with its entries

$$C_{ij} = \frac{\sum_{k}^{n}(x_{i,k} - \bar{x_i})(x_{j,k} - \bar{x_j})}{(n-1)}, \tag{3.20}$$

where $x_{i,k}$ is the k-th element of the i-th datapoint and $\bar{x}_i$ is the corresponding mean. In order to diagonalise the covariance matrix, one has to calculate the unit eigenvalues λ_i and corresponding eigenvectors $\mathbf{v}_i$. The matrix of eigenvectors $\hat{V} = (\mathbf{v}_1 \dots \mathbf{v}_n)$ can then be used to transform $\hat{C}$ into diagonal form

$$\hat{V}^{-1}\hat{C}\hat{V} = \hat{D}, \tag{3.21}$$

where $\hat{D}$ is a diagonal matrix with the λ_i as diagonal entries. The eigenvectors should be ordered in descending order of the eigenvalues, as the vector with the highest λ_i corresponds to the dimension of highest variance. In order to reduce the dimension of the dataset, $\hat{V}$ can be built with a subset of the v_i: $\hat{\tilde{V}} = \big(\mathbf{v}_1 \dots \mathbf{v}_p\big)$. This will only keep the $p < n$ principal components with the highest variance. The magnitude of λ_i gives the amount of information in the corresponding dimension, and a dimensional reduction that cuts away eigenvectors with low λ_i should represent the data well. The original data transformed into the new coordinate system reads

$$\hat{Y}' = \hat{\tilde{V}}^{\mathrm{T}}\hat{X}^{\mathrm{T}}. \tag{3.22}$$

$\hat{Y}'$ now contains n p-dimensional feature vectors, which represent the original feature vectors in the transformed space after dimensionality reduction. Oftentimes for illustrative purposes it is useful to transform the reduced data back into the original coordinates to check if the important information was indeed retained. In this case the new feature vectors can be transformed into the original n-dimensional coordinate system by the transformation

$$\hat{X}' = \hat{\tilde{V}}\hat{Y}' + \hat{\bar{X}}. \tag{3.23}$$

Here, $\hat{\bar{X}}$ is the matrix containing the mean values, that were initially subtracted from the dataset. In the case where all eigenvectors where in $\hat{\tilde{V}}$, $\hat{X}'$ is equal to $\hat{X}$, however, in the case where only p Eigenvectors where used, the information from the dimensions with low variance is lost. If the data still resembles the original data reasonably well, it can be assumed that the dimensionality reduction does only remove non-essential information from the data representation. Oftentimes, the first few principal components of a dataset can be used to reconstruct new feature vectors or find trends in the data that might be obstructed by the high dimensionality of the datapoints. Generally, the first principal components contain features, which are (maybe scaled by a factor) shared by most of the datapoints.

3.5.2 Autoencoder

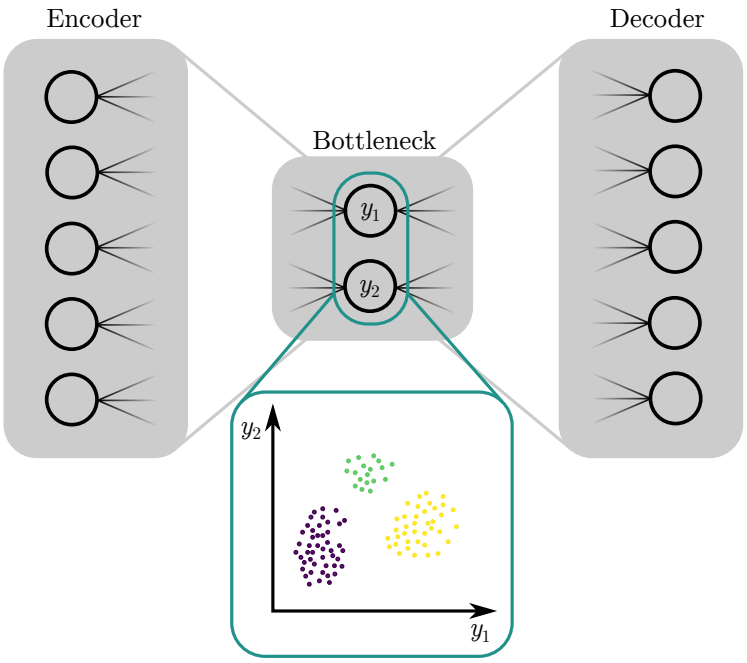

Figure 3.6: Concept of an autoencoder. The basic building blocks are the encoder and decoder networks, which are connected through a bottleneck. Input and output dimension have to be equal. The outputs of the bottleneck layer can be used to reduce the dimension of the input dataset.

Autoencoders are a type of neural network used for unsupervised learning, utilising the techniques from supervised learning. As the name suggests, they aim to to encode typically high dimensional input data into an efficient lower dimensional representation [110, 122]. Autoencoders consist of three general parts as shown in figure 3.6: Firstly, an encoder-network containing at least the input layer. Secondly, a decoder network containing at least the output layer that has the same shape as the input layer, and lastly a bottleneck with less neurons than the input/output layers. The shape of the encoder and decoder networks is in principle arbitrary, as long as the input and output shapes are equal. The network is *supervisedly* trained to reproduce the input. That means the input itself is the label the network is trained with. Since no separate label apart from the input is needed, autoencoders themselves are unsupervised learners.

Due to the lower dimensionality of the bottleneck, not all information can propagate from input to output. Instead, in order to achieve good performance, the network has to *encode* the information from the input into a representation that can *fit* through the bottleneck and can then be reconstructed efficiently by the decoder network. An autoencoder with a two-neuron bottleneck can be used to reduce the dimensionality of a dataset down to a two dimensional representation where the axes represent the neurons outputs [123]. Generally, classification tasks for new inputs can then be achieved with clustering methods. This capturing of essential

information through the network is comparable to how PCA captures important information. Indeed, it can be shown that autoencoders are a more general approach to dimensionality reduction. The effect of PCA can be captured by using linear activations on all neurons, or a single sigmoid layer [124–126]. However, autoencoders are much more powerful since they are not restricted to linear transformations. Another interesting and instructive application is to use autoencoders to denoise data [127]. Since in the best case only relevant information can propagate through the bottleneck, random noise will not show up on the output of the network. This can be used to denoise data or pictures by sending them through the network.

Figure 3.7 shows the comparison between PCA and autoencoder for the MNIST dataset of handwritten digits [128] for a dimensionality reduction down to two dimensions. The comparison shows that the autoencoder separates some of the classes more clearly than the PCA does. However, it sometimes might be desirable to use PCA, as the principal components allow direct analysis of the features that are extracted and provide more insight into how the algorithm does the splitting. More information on dimensionality reduction with autoencoders can be found in ref [123].

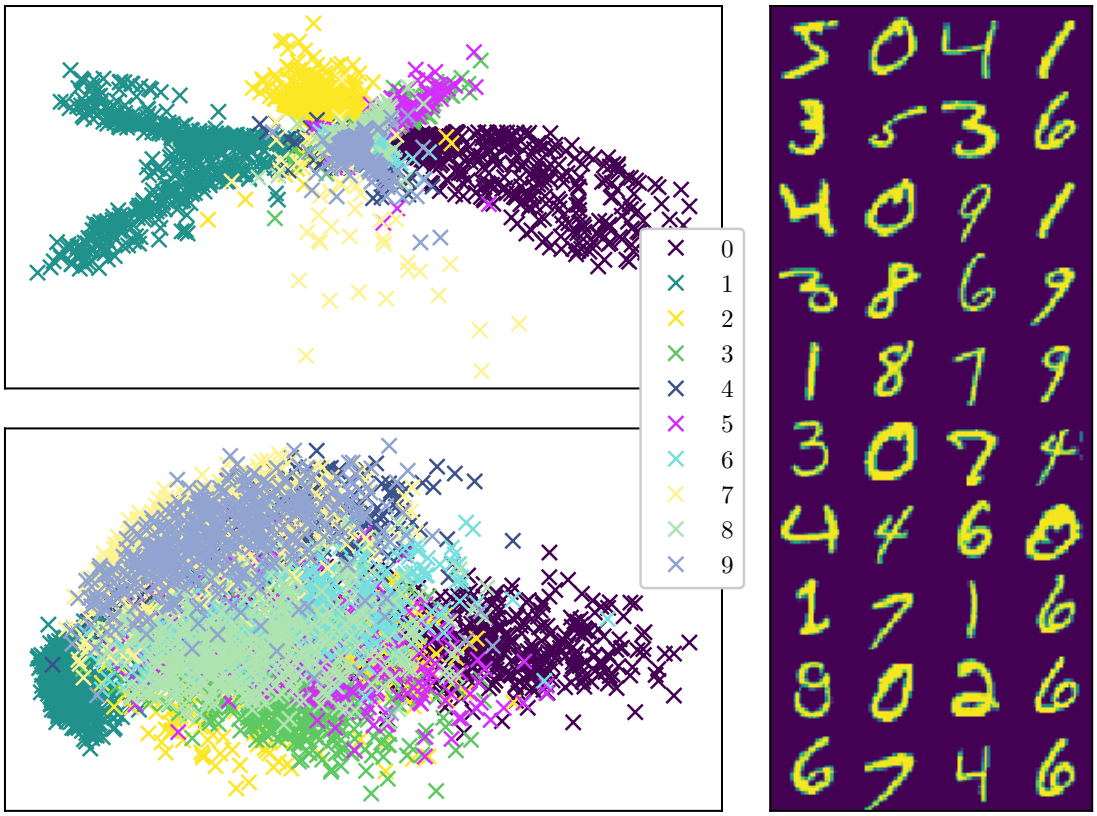

Figure 3.7: Comparison of PCA and autoencoder on the MNIST dataset. The autoencoder approach (upper left) and the PCA approach (lower left) are used for dimensionality reduction of the MNIST dataset down to two dimensions. The color of the datapoint represents the label of the picture that produced the datapoint. The MNIST dataset contains 70000 labelled 28×28 pixel pictures of handwritten digits such as the pictures in the right panel [128]. This plot is reproduced from ref [123].

CHAPTER 4

Experimental Setup and Techniques

This chapter serves as an introduction to the experimental system and techniques that are used to generate an ultracold Fermi gas and observe it. It will provide all information needed in order to understand the shared general experimental details needed for the investigations later on in this thesis. The chapter starts with a brief review of the experimental system that is used in order to generate a degenerate sample of ^{6}Li atoms. It then proceeds to describe the data acquisition via absorption imaging, as well as the rapid ramp technique - a method for detecting the condensate fraction in a cloud. Experimental details that are only important in the context of single measurements will be reviewed in the corresponding section of the thesis.

4.1 The experimental system

Construction of the Na-Li machine that was used during this thesis started in 2013, when the research group moved from Cambridge to Bonn. The main construction was done by Dr. Alexandra Behrle, Dr. Timothy Harrison and Dr. Kuiyi Gao. I joined the project in 2015, when the experiment was, apart from a few details, already running on a daily basis. The following introduction serves as a brief review of the steps that are taken in the system in order to achieve a degenerate sample of ^{6}Li atoms. Far more information on (the construction and testing of) the experimental system can be found in the theses of Dr. Alexandra Behrle [129] and Dr. Timothy Harrison [130].

4.1.1 The vacuum system

The vacuum system is divided into two main parts that are connected by the variable pitch Zeeman slower, which also contains a differential pumping tube, see figure 4.1. On one side is the dual species oven with sources of ^{6}Li and ^{23}Na, which generates collimated atom beams that travel through the Zeeman slower towards the main chamber. The oven design was inspired by ref [131] and consists of an oven tube with heated reservoirs for lithium (400 °C) and sodium (330 °C), separated by a mixing nozzle that is held at a higher temperature (475 °C) to prevent clogging. The oven tube is connected to the oven chamber containing a water cooled copper cold-cup through a second nozzle (490 °C). The cold cup together

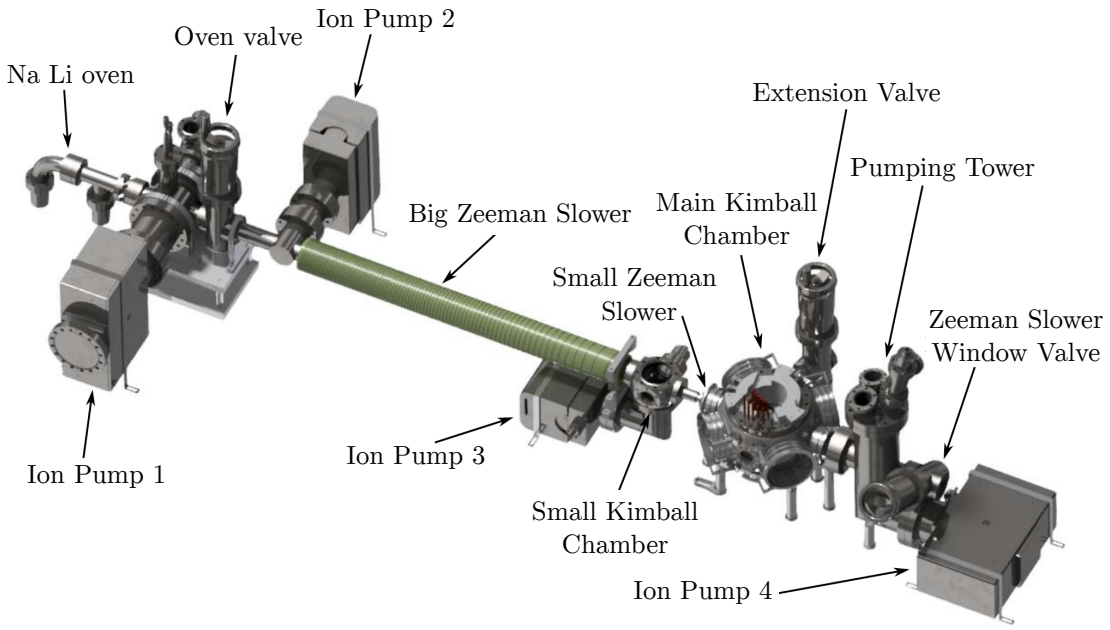

Figure 4.1: Vacuum system of the Li-Na machine. The picture shows all essential components of the vacuum system without optics or breadboards.

with a skimmer is used to collimate and steer the atom beam and capture atoms that are not ending up in the beam. The oven chamber is connected to a large ion getter pump[1], which is protected against coating from the atom sources via a large baffle. Towards the differential Zeeman slower tube is a valve to isolate the experiment chamber from the oven chamber in case it has to be opened for refilling of the atom sources. There is a second ion getter pump[2] before the Zeeman slower. The Zeeman slower itself is divided into two stages, separated by a small Kimball chamber that was initially used to align the Zeeman slower laser beams while monitoring the fluorescence. The Zeeman slower stage contains a differential pumping section that serves as pressure isolation. It is connected to a third ion getter pump[3]. The smaller second Zeeman slower stage is connected to the main experiment chamber, where all subsequent stages of the experiment are realised. The main chamber has six large viewports, one in every spatial direction and is designed to have good optical access. In order to achieve a high number of atoms in the magneto-optical trap (MOT), the MOT position is close to the end of the Zeeman slower stage (4.5 cm) to optimise incoming atom flux. After subsequent magnetic transport, the final position of the cloud is 3 mm below the upper fused silica viewport. The distance is chosen to work with a high NA objective that will be used for future experiments [132]. All viewports surrounding the main

[1] Gamma Vacuum, $100\,\mathrm{l\,s^{-1}}$
[2] Gamma Vacuum, $75\,\mathrm{l\,s^{-1}}$
[3] Gamma Vacuum, $45\,\mathrm{l\,s^{-1}}$

chamber are non-magnetic and homemade, following a procedure introduced in ref [133]. Connected to the main chamber is a pumping tower containing a fourth ion getter pump[4] and a Titanium-Sublimation pump to reach ultra-high vacuum. Thanks to the differential pumping, the pressure in the main chamber ($\approx 1 \times 10^{-11}$ mbar) is much lower than in the heated oven chamber ($\approx 5 \times 10^{-9}$ mbar). On the far side of the oven is the Zeeman slower window, which is the entry point for the Zeeman slower laser beam propagating towards the oven. The window is made from sapphire and is heated to 210 °C in order to prevent coating from the atom beam.

4.1.2 From hot to MOT: Zeeman slower, Magneto-Optical Trap and molasses

In the following subsection the optical stages of the experiment will be explained. Firstly, the Zeeman slower which decelerates the atoms coming from the oven so they can be captured. Secondly, the dual species magneto-optical trap (MOT), which captures the atoms in the experiment chamber and cools them down to Doppler temperature. Before loading into the magnetic trap, we use molasses to further cool down the sodium atoms to achieve lower temperature and better capturing in the initial magnetic trap. In order to get an overview of the required laser wavelengths, the level structure of lithium and sodium is given in figure 4.2.

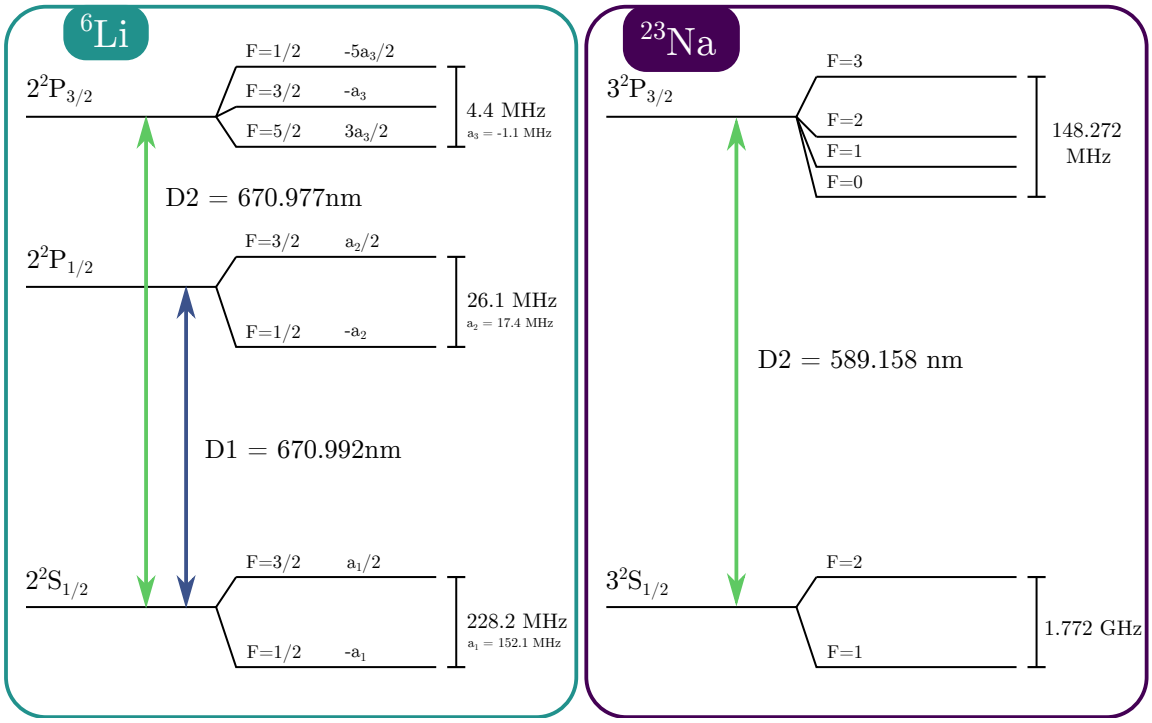

Figure 4.2: Hyperfine structure overview. Hyperfine structure for ^{6}Li [134] (left) and ^{23}Na [135] (right). For both species, we work on the D2 line.

[4] Gamma Vacuum, 100 l s^{-1}

Zeeman slower

The first stage of cooling down the atoms coming from the oven is the Zeeman slower stage. In the experiment, we employ a variable pitch dual-species spin flip Zeeman slower. The design is based on ref [136]. Construction and testing are detailed in the theses of Dr. Alexandra Behrle [129] and Dr. Timothy Harrison [130]. The Zeeman slower is optimised for sodium, as high sodium number simplifies effective cooling of the Li cloud down to the dipole trap stage. There are two stages in the Zeeman slower: the big Zeeman slower stage, which provides the initial offset field and has a decreasing magnetic field, and another small Zeeman slower directly in front of the experiment chamber with a two-layer variable pitch coil. The two stages are separated by a small Kimball chamber which contains the zero crossing of the magnetic field. In order to cool sodium, we employ laser light to address the $3^2S_{1/2}\,|F=2, m_F=2\rangle \rightarrow 3^2P_{3/2}\,|F'=3, m_{F'}=3\rangle$ transition. For lithium, we address the $2^2S_{1/2}\,|F=3/2, m_F=3/2\rangle \rightarrow 2^2P_{3/2}\,|F'=5/2, m_{F'}=5/2\rangle$ transition. To account for the Doppler shift, the light is detuned by 500 MHz with respect to the atomic frequency. In principle, both transitions are closed, but we found that adding small amounts of repumper light improves the performance [130]. The capture velocities for ^{23}Na and ^{6}Li are $1\,250\,\mathrm{m\,s^{-1}}$ and $1\,800\,\mathrm{m\,s^{-1}}$, respectively. The flux of sodium at the end of the slower stage was estimated to be $\approx 1 \times 10^{12}\,\mathrm{s^{-1}}$. Since the magnetic field of the small Zeeman slower extends to the MOT position and turning off the Zeeman slower coil would move the position of the MOT, another coil on the opposite site of the main chamber from the entry point is used to compensate the Zeeman slower field at the MOT position.

Magneto-optical trap

In order to cool the atoms down to the Doppler limit after the Zeeman slower stage, we capture both species in a magneto-optical trap in standard configuration [137]. The dual-species MOT consists of two counterpropagating beams with opposite circular polarisation in all three spatial directions. The required optics is arranged in towers around the experimental system which contain the optics to shape and combine the beams at 589 nm for the $3^2S_{1/2}\,|F=2, m_F=2\rangle \rightarrow 3^2P_{3/2}\,|F'=3, m_{F'}=3\rangle$ transition of sodium and 671 nm for the $2^2S_{1/2}\,|F=3/2, m_F=3/2\rangle \rightarrow 2^2P_{3/2}\,|F'=5/2, m_{F'}=5/2\rangle$ transition of lithium. One of the goals during construction was to achieve high atom numbers, which is why the size of the MOT beams is comparably large at 19.8 mm waist radius for sodium and 11.1 mm waist radius for lithium. The quadrupole field required for trapping is generated by two separate MOT coils above and below the chamber, see figure 4.4. These coils are displaced with respect to the center of the chamber in order to keep the distance between the end of the Zeeman slower and the MOT low and the atom flux high. In order to achieve a high density of Na atoms without large losses through light-assisted collisions, we use a Dark-SPOT MOT [138] for Na. For this, a hollow beam of repumper light ($3^2S_{1/2}\,|F=2, m_F=2\rangle \rightarrow 3^2P_{3/2}\,|F'=2, m_{F'}=2\rangle$) is generated in the upper tower, such that in the center of the MOT no repumper light is present. More details on the optics and generation of this beam can be found in ref [129]. Before transporting the sodium atoms into the magnetic trap, we have a short 2 ms time of molasses with a rapid turnoff of the MOT coils, bringing the temperature of the sodium down to 200 µK as compared to the Doppler temperature of 240 µK. In order to match the

cloud size of Li to the size of the sodium cloud during loading into the magnetic trap and further reduce the temperature, for lithium, we use a compressed MOT (cMOT) [139, 140]. For this, we ramp the power and detuning of the lithium MOT beams at the end of the MOT phase. The final temperature of the Li cloud was estimated to be $\approx 750\,\mu K$. Using the cMOT yielded a factor of 3 increase in atom number in the initial magnetic trap [130].

4.1.3 Magnetic trap and evaporation

The following section briefly reviews the experimental steps taken from the end of the molasses up until loading of the atoms into the dipole trap. It will describe the loading into the magnetic trap, namely the optical pumping and magnetic transport. Afterwards, the magnetic trap and evaporation procedure will be reviewed.

Optical Pumping

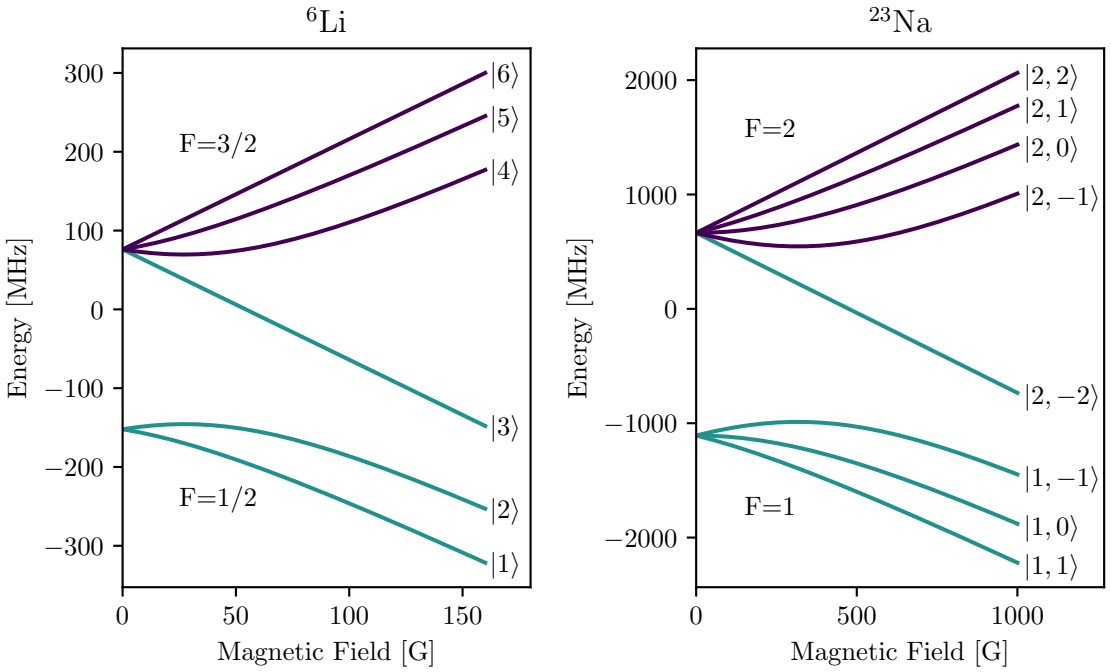

Figure 4.3: Breit Rabi diagram for the ground states of ^{6}Li (left) and ^{23}Na (right). Relevant data taken from [134, 135]. Curves are colored according to their high-field manifold. Labelling of the lithium states is the pragmatic labeling that is used throughout this thesis.

In order to trap both kinds of atoms in the magnetic trap, they have to be in low-field seeking states, meaning the magnetic moment has to be positive. The Breit-Rabi diagrams for lithium and sodium are shown in figure 4.3. Additionally, in order to prevent spin-exchange collisions, which preserve total angular momentum but change the hyperfine state to a possibly non-trappable state, the atoms have to be in maximally stretched states. In our case, the

chosen states are the $|F = 2, m_F = 2\rangle$ state for sodium and the $|F = 3/2, m_F = 3/2\rangle \equiv |6\rangle$ state for lithium. In order to pump the atoms into these states after the MOT phase, a bias field of 2.6 G is employed with help of a MOSFET and the magic layer of the Feshbach coils (see below). This lifts the degeneracy of the m_F-levels. By applying pump (Na: $|F = 2\rangle \rightarrow |F' = 2\rangle$, Li: $|F = 3/2\rangle \rightarrow |F' = 3/2\rangle$) and repump (Na: $|F = 1\rangle \rightarrow |F' = 2\rangle$, Li: $|F = 1/2\rangle \rightarrow |F' = 3/2\rangle$) light through the lower MOT tower, the atoms are pumped to the above states with high efficiency and can be captured by the magnetic trap.

Magnetic transport and Feshbach coils

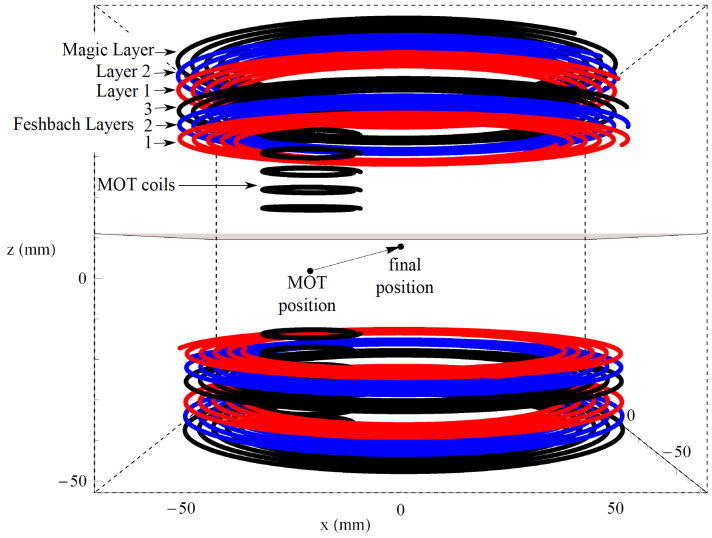

Figure 4.4: Feshbach and MOT coils. For the Feshbach coils the names of the different layers are indicated on the upper coils. In the experiment, we use Feshbach Layer 1, 2 and 3 for creation of a homogeneous magnetic field. The other layers are used for different tasks, i.e. compensation of stray fields and gravity compensation. Also indicated is the magnetic transport route from the MOT position to the final cloud position close to the UKEA viewport (indicated in grey). Picture adapted from ref [130].

The quadrupole trap is formed by two large coils in anti-Helmholtz configuration. The very same coils will later also provide the homogeneous offset field in the dipole trap, after switching the current direction on one side to achieve approximate Helmholtz-configuration with a small remaining field curvature. The coils are wound onto a Tekadur body that places them such that the center position is ≈ 3 mm below the upper UKEA viewport, which also marks the final atom position. The coils are arranged in six parallel layer pairs each (see figure 4.4), and each layer is separately water cooled. Water cooling is achieved by running water through the 4 mm × 4 mm wire with a hollow core of radius $r = 1.25$ mm. The inner and outer radius of each layer is 51 mm and 72.5 mm, respectively. Magnetic transport into this magnetic trap is achieved by initially trapping the atoms by running high current through the MOT coils and then simultaneously ramping the MOT coils down and Feshbach coils up. The exact, optimised ramp can be found in ref [130]. The position of both coil pairs and the

initial and final atom position are indicated in figure 4.4. Since the Feshbach coils provide a quadrupole field which is prone to Majorana losses in the center region, we send a high power 532 nm laser beam[5] through the center of the trap. This ejects the atoms from the center and prevents losses (see section 4.1.4 for a description of the underlying mechanism). The trap depth of the magnetic trap at the final position is $\approx 2\,\mathrm{mK}$.

Evaporation in the plugged quadrupole trap

Once loaded, sodium is selectively evaporated from the trap. Since the number of sodium atoms is much higher than lithium, this will sympathetically cool the lithium sample without any major loss of lithium atoms. The evaporation is done by applying radio-frequency radiation to drive the sodium from the trapped $|F = 2, m_F = 2\rangle$ state into the untrappable high-field seeking $|F = 1, m_F = 1\rangle$ state. To selectively address only the atoms that are on the edge of the trap (and therefore carry the highest energy), the frequency is ramped from 1.9 GHz towards the resonance at 1.772 GHz to a final value of 1.774 GHz within 5 s. At 1.777 5 GHz, we decompress the magnetic trap by decreasing the current through the coil. This is to decrease the density of the sodium cloud and suppress three-body collisions. At the end of the evaporation all sodium is evaporated. The exact characteristics of the ramp are optimised to retrieve the highest possible atom number in the molecular BEC after dipole evaporation.

4.1.4 Dipole trap

This section describes the final trapping stage of the lithium and the creation of an ultracold sample of degenerate ^{6}Li atoms with a temperature of just a few percent of T_F. Firstly, a small review of the idea behind Dipole traps will be given. Secondly, the characteristics of the trap will be discussed. This will be done in a bit more detail, as the exact shape of the trap will become important for the thermometry introduced in chapter 5. The laser system for the trap was built by Justas Andrijauskas during his masters thesis [141] and implemented into the experiment by Timothy Harrison, Alexandra Behrle and Kuiyi Gao [129, 130].

Working principle

The dipole force used for trapping atoms in dipole traps arises through interactions of the atoms with the dispersive part of a light field [142]. Essentially, the light induces a dipole in the atom and the interaction of this dipole with the light field creates a force that acts on the atom. The potential created by the interaction of an atom with a semi-classical light-field is given by the action of the dipole-operator $\hat{\mathbf{p}}$ on the electric field $\mathbf{E}$ [142]

$$U_{\mathrm{dipole}} = -\frac{1}{2}\left\langle \hat{\mathbf{p}}\mathbf{E}\right\rangle . \tag{4.1}$$

Here, $\langle \ldots \rangle$ denotes the average over fast oscillating terms. The electric field in complex notation is given by $\mathbf{E}(\mathbf{r}, t) = \hat{\mathbf{e}}\tilde{E}(\mathbf{r})\exp(-i\omega t) + \mathrm{c.c.}$. The dipole moment induced by this light

[5] generated by a Sprout-G laser with 15 W output power.

field is given by $\mathbf{p}(\mathbf{r},t) = \hat{\mathbf{e}}\tilde{p}(\mathbf{r})\exp(-i\omega t) + \text{c.c.}$. The amplitude of the dipole moment induced by the electric field is related to the amplitude of the electric field itself via $\tilde{p} = \alpha\tilde{E}$. Here, α is the complex polarisability, which depends on the driving frequency ω. The polarisability describes the strength and phase relation between driving field and induced dipole. Putting this into equation (4.1) yields

$$U_{\text{dipole}} = -\frac{1}{2\epsilon_0 c}\text{Re}(\alpha)I, \tag{4.2}$$

where I denotes the intensity distribution of the light field. The potential created by the interplay between electric field and dipole only depends on the real part of the polarisability. The absorption of power by the atom and subsequent emission of dipole radiation is described by $P_{\text{abs}} \propto \text{Im}(\alpha)$ and is only dependent on the imaginary part. This is known as the scattering force and is the essential mechanism behind laser cooling used e.g. during the Zeeman lower and MOT stage. The sign of $\text{Re}(\alpha)$ in the above equation determines the sign of the potential. The strength of the potential for different atomic species is only determined by the polarisability. In general, the polarisability dependends on the dipole matrix element of all possible transitions from the ground state $|g\rangle$ to some excited state $|e\rangle$ with energies E_g and E_e [143],

$$\alpha(\omega) = \sum_e |\langle e|\hat{\mathbf{p}}\mathbf{E}|g\rangle|^2 \left(\frac{1}{E_e - i\hbar\Gamma_e/2 - E_g - \hbar\omega} + \frac{1}{E_e - i\hbar\Gamma_e/2 - E_g + \hbar\omega}\right). \tag{4.3}$$

Here, $1/\Gamma_e$ describes the lifetime of the excited state. For large detuning it is sufficient to model the system as an effective two-level system. Using the Lorentz approximation, we can approximate α by

$$\alpha(\omega) = \frac{e^2}{m_e}\frac{1}{\omega_0^2 - \omega^2 - i\omega\Gamma} \tag{4.4}$$

$$\text{with } \Gamma = \frac{e^2\omega^2}{6\pi\epsilon_0 m_e c^3}. \tag{4.5}$$

Here, ω_0 is the angular frequency associated with the level splitting $(E_e - E_g)/\hbar$. Putting this into equation (4.2) yields

$$U_{\text{dipole}} = -\frac{3\pi c^2}{2\omega_0^3}\left(\frac{\Gamma}{\omega_0 - \omega} + \frac{\Gamma}{\omega_0 + \omega}\right) I(\mathbf{r}). \tag{4.6}$$

Defining the detuning $\Delta = \omega - \omega_0$ and applying the rotating wave approximation $\omega \approx \omega_0$, such that the second term becomes small as compared to the first term yields the final result

$$U_{\text{dipole}} = \frac{3\pi c^2}{2\hbar\omega_0^3}\frac{\Gamma}{\Delta}I(\mathbf{r}). \tag{4.7}$$

This shows that the sign of Δ determines the sign of the potential. A *blue detuned* light field with $\Delta > 0$ will act repulsive on the atom, such as the plug beam used to eject atoms from the center of the magnetic trap. For a *red detuned* field with $\Delta < 0$, the atoms will move towards higher intensity. To trap a cold sample of atoms, the scattering rate between light field and atoms should remain low. Luckily, since the scattering rate due to absorption scales with Γ/Δ^2 and the potential with Γ/Δ, large detunings as used in practice reduce the heating rate but keep a reasonable deep potential. The exact shape of U_{dipole} is determined through the intensity distribution of the generating light field $I(\mathbf{r})$. When using high power laser beams as a source for the trapping potential, the shape of the beams is usually gaussian. For cold samples that will reside close to the center of the beam it is usually sufficient to make a harmonic approximation and approximate the Gaussian potential as a harmonic oscillator potential. The dipole potential can then be written in the form of a harmonic oscillator potential with

$$U_{\text{dipole}} = \frac{1}{2}m(\omega_x^2x^2 + \omega_y^2y^2 + \omega_z^2z^2), \tag{4.8}$$

where $\omega_i \propto \sqrt{P}$. However, in some cases, the effect of the anharmonicity might become important, such as for the thermometry introduced in chapter 5. In principle, arbitrary shapes can be used for $I(\mathbf{r})$ to design potentials, such as hollow beams, light-sheets or higher order modes (see chapter 10 and refs [144, 145]).

Trap geometry

The trap used throughout this thesis is formed by two separate laser beams originating from a high power diode-pumped continuous wave ytterbium fiber laser[6] at 50 W. The beams are split and prepared and brought to the experiment table through high power fibers[7]. Each beam initially had a maximum power of 13 W before entering the experiment chamber, however, in the last year we noticed a decrease in maximum power of the laser from 50 W down to 35 W, which changes the initial trap depth. Evaporation is achieved by gradually lowering the power in the beams with an exponential ramp. For a sequence that results in a molecular BEC, the final power is 160 mW for the horizontal beam and 0 mW for the vertical beam, subject to change depending on the desired trap geometry. The trap is formed by two gaussian beams: For one, a horizontal beam with a waist of 110 µm × 53 µm (the smaller value being along gravity (z-)direction). Secondly, a vertical beam which is retro-reflected by the upper UKEA viewport. This beam is circular with beam waist 120 µm × 120 µm. Additionally to the potential created by the dipole beams, the non-vanishing curvature of the

[6] IPG Photonics, model YLR-50-Y12

[7] Horizontal beam: NKT Photonics, Photonic Crystal Fiber Cable LMA-PM-15, 10m, Vertical beam: Thorlabs, P3-1064PM-FC-10

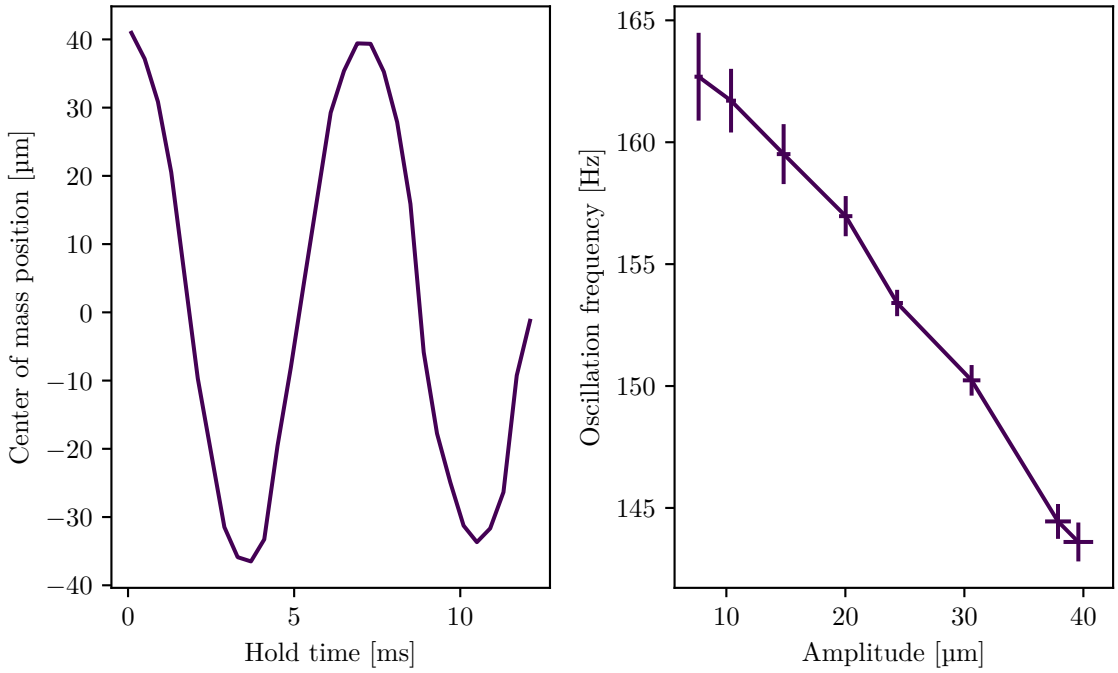

Figure 4.5: Center of mass oscillations in the dipole trap. (Left) Center of mass oscillations of an atomic cloud after rapidly changing the gradient. **(Right)** Power dependence of the oscillation frequency. For a perfectly harmonic potential, the oscillation frequency would not depend on the amplitude. Since the shape of our trap is gaussian, when exciting motion in the trap, the cloud probes the anharmonic parts of the potential.

Feshbach coils due to imperfections in the Helmholtz configuration acts as a potential. This additional potential is confining in the x-y-plane with $\approx 2\pi \cdot (16\,\mathrm{Hz}, 16\,\mathrm{Hz})$ depending on the field value and anti-confining in z-direction with $\approx 2\pi \cdot 32\,\mathrm{Hz}$. The total trap frequencies in the dipole trap can be measured by exciting center-of-mass oscillations within the trapping potential by giving the cloud a slight *kick*, for example by adding an additional gradient with a separate coil and suddenly switch the coil off. For weak kicks in the harmonic limit, the trap frequency will stay constant. For stronger kicks, however, the non-harmonic parts of the trap will be probed, effectively reducing the measured trap frequency, see figure 4.5.

Loading, state preparation and evaporation

We achieve loading into the dipole trap by simultaneous linear down-ramping of the magnetic trap and plug beam and up-ramping of the dipole beams within 150 ms. This was found empirically to be the best way to maximise the number of atoms in the dipole trap [130]. For simplicity, from here on, the lowest six hyperfine states of ^{6}Li will be labeled $|1\rangle$ to $|6\rangle$, where $|1\rangle$ represents the lowest state, see figure 4.3. After loading, the trap is filled with a pure sample of lithium in the maximally stretched $|6\rangle$ state. Since thermalisation in the dipole trap for cold temperatures can only take place for distinguishable particles

due to the Pauli principle, it is necessary to create a state-mixture. This is done by two Landau-Zener sweeps: Firstly, all atoms are transferred from state $|6\rangle$ to state $|1\rangle$. This is done by applying rf-radiation at 239 MHz and ramping the magnetic field from 5.6 G to 7.1 G. The transfer has an efficiency of $> 96\,\%$. The second Landau-Zener sweep creates an equal mixture of $|1\rangle$ and $|2\rangle$, driving $|1\rangle \rightarrow 50\,\%\,|1\rangle + 50\,\%\,|2\rangle$. This transfer is done at a magnetic field of 102 G by applying a linear rf sweep from 60.375 MHz to 60.625 MHz. The rf power, frequency and magnetic field can be tuned to achieve an equal mixture. This transfer creates a pure state $\frac{1}{\sqrt{2}}(|1\rangle + |2\rangle)$, which means thermalisation is still not possible. However, by adding a decoherence time of 100 ms during the ramp time to the evaporation field, the sample collapses into a mixed state of $50\,\%\,|1\rangle$ and $50\,\%\,|2\rangle$. After the mixture has been achieved, the magnetic offset field is ramped to 795 G, close to the Feshbach resonance at 834 G. There, interactions are strong and the sample thermalises fast, which enables efficient evaporation. Evaporation is achieved by gradually lowering the power in the dipole beams with an exponential ramp within 0.9 s. This changes the trap depth from $\approx 65\,\mu$K to $\approx 800\,$nK. Sometimes, due to shape of the Feshbach resonances (see section 2.2.2) it might be useful to have a condensate in a mixture of $|1\rangle$ and $|3\rangle$, instead of $|1\rangle$ and $|2\rangle$. In case of the magnetic field quenches introduced in chapter 8 for example, it is desirable to use the narrowest Feshbach resonance. In this case, before evaporation we do another Landau-Zener sweep to transfer $|2\rangle \rightarrow |3\rangle$ and subsequently evaporate close to the $|1\rangle + |3\rangle$ Feshbach resonance.

The ultracold Fermi gas

At the end of the evaporation, we create a molecular BEC of ^{6}Li in a mixture of $|1\rangle$ and $|2\rangle$. The atom number is $\approx 2 \cdot 10^6$ per state and the temperature is $\approx 0.07\,T_F$. From here on, the magnetic offset field can be adiabatically sweeped to the desired value to investigate the BEC-BCS crossover.

4.2 Imaging of cold atomic clouds

Raw data throughout this thesis consists nearly exclusively of pictures of atom clouds, captured with CCD cameras. For this, a laser pulse close to resonance of the imaging transition is directed at the atoms and during the duration of the pulse, the CCD chip is illuminated. Together with a carefully chosen telescope with a well calibrated magnification, this will yield a picture of the shadow of the atomic cloud. From this, it is possible to deduce the optical density which is proportional to the atom density. This section introduces the imaging system and technique and the data preparation to extract meaningful quantities from the shadow of an atomic cloud.

4.2.1 Imaging systems

In our experiment we use four different imaging systems for different purposes. For usage, we name them by their experiment control handle, namely ANDOR0, ANDOR1, ANDOR2 and ALTA0. Two cameras image along the vertical direction: ANDOR0 images the atoms from below, with the imaging beam coming from above the chamber, whereas ANDOR1 images

from above. ANDOR2 images from the side, perpendicular to the horizontal dipole beam and plug beam. ALTA0 also images from the side, but along the horizontal dipole beam direction, perpendicular to the imaging axis of ANDOR2. The following subsections will give a brief overview over the ANDOR imaging systems, as the ALTA0 system is mainly used for maintenance reasons and does not play a role in the remainder of this thesis. All ANDOR camera systems use fast kinetics mode to take two pictures with the size of the active pixel area in quick succession.

ANDOR0 imaging system

The ANDOR0 imaging system is designed to image the momentum distribution of the atom cloud. This is done by imaging the atom cloud after $\approx 15\,\mathrm{ms}$ time of flight, which is a quarter period of the potential created by the field curvature of the Feshbach coils and therefore shows the momentum distribution, see also ref [146]. The camera in use is an iXon-ultra 897 with $16\,\mu\mathrm{m} \times 16\,\mu\mathrm{m}$ pixel size and an active area of 512×512 pixels. The cloud is imaged from below, which means the imaging beam comes from above the chamber. The magnification of the imaging optics is $M = 4.25$ and the Numerical Aperture of the imaging lens is $\mathrm{NA}_{\mathrm{ANDOR0}} = 0.1$ with a depth of field of $d_{\mathrm{dof,\ ANDOR0}} \approx 70\,\mu\mathrm{m}$.

ANDOR1 imaging system

The ANDOR1 imaging system is used for in-situ pictures of atom clouds. The imaging axis is along the axes of the Feshbach coils, with imaging from above the chamber as compared to ANDOR0. The camera model in use is the same as ANDOR0, namely an iXon Ultra 897 with an active pixel area of 512×512 pixels and a pixel size of $16\,\mu\mathrm{m} \times 16\,\mu\mathrm{m}$. The magnification of the imaging optics is $M = 6.55$, the numerical aperture is $\mathrm{NA}_{\mathrm{ANDOR1}} = 0.25$ with a depth of field of $d_{\mathrm{dof,\ ANDOR1}} \approx 10\,\mu\mathrm{m}$.

ANDOR2 imaging system

The ANDOR2 imaging system images the atom cloud from the side, perpendicular to the direction of the horizontal dipole and plug beams. It consists of an iXon Ultra 888 camera with 1024×1024 pixels active area and a pixel size of $13\,\mu\mathrm{m} \times 13\,\mu\mathrm{m}$. The magnification of the imaging system is $M = 6.0$. This camera is mainly used to image the cloud in trap, e.g. to extract the trap frequency in z-direction.

4.2.2 High field imaging of lithium

In the experiment, we image ultracold clouds of lithium atoms at fields around the Feshbach resonance ($834\,\mathrm{G}$ for $|1\rangle + |2\rangle$ or $690\,\mathrm{G}$ for $|1\rangle + |3\rangle$). As the exact energy splitting for the imaging transition $2^2S_{1/2}\,|1, 2 \text{ or } 3\rangle \rightarrow 2^2P_{3/2}\,|m_j = -3/2\rangle$ is dependent on the magnetic field through the Zeeman shift, the imaging frequency has to be adjusted depending on the chosen interaction. This is done by calculating the energy shift between the transitions through the Breit-Rabi problem and adjusting the offset lock of the imaging laser appropriately. As the exact value of the field at the atom position is only approximately known, usually we scan

an additional small offset frequency around the predicted resonance to determine the exact resonance frequency more accurately.

4.2.3 High intensity absorption imaging

The following technique to image atoms at high intensities was first introduced in ref [147], and later adapted by our group [129, 130, 148, 149] in order to quantify the atomic density in pictures of atomic clouds. It was used during this thesis for the ANDOR0 and ANDOR1 imaging system, both of which where used to image atom clouds.

Limits of imaging

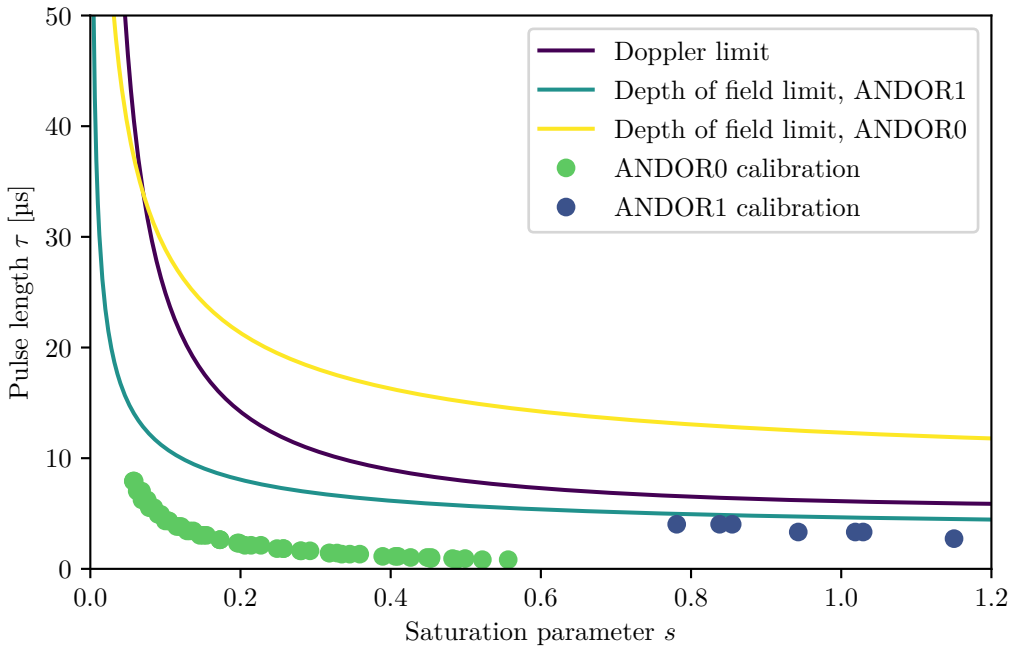

Figure 4.6: Doppler and depth-of-field limit for ^{6}Li. The depth-of-field limit is shown for the ANDOR0 imaging system and the ANDOR1 imaging system separately. As d_{dof} is a factor of 7 smaller for ANDOR1, this imposes the stricter limit. During imaging we make sure to stay below the limits. The green and blue datapoints show the configurations that we used for the imaging calibration.

In order to image the atomic cloud, we shine in resonant laser light onto the cloud. This resonant light exerts a force on the atomic cloud and accelerates it. In order to quantify the number of atoms present in the sample, we have to stay resonant and keep the cloud stationary during the imaging process. The average acceleration of the cloud can be calculated through the scattering force exhibited by the resonant light beam

$$\bar{a} = \frac{\hbar k_L}{m}\frac{\Gamma}{2}\frac{I}{I + I_{\mathrm{sat}}}, \tag{4.9}$$

where I and I_{sat} are the (saturation) intensity, Γ is the linewidth of the imaging transition and k_L is the wave vector associated with the imaging light. By using an imaging pulse of length τ, the atoms will on average acquire a velocity of v_τ and travel a distance of d_τ given by

$$v_\tau = \bar{a}\tau \qquad \text{and} \qquad d_\tau = \bar{a}\tau^2/2. \tag{4.10}$$

In order to stay resonant with the cloud during the imaging pulse, the Doppler shift introduced by the velocity v_τ has to stay small as compared to the power broadened linewidth of the imaging transition, requiring

$$\underbrace{k_L v_\tau = \frac{\hbar k_L^2 \Gamma}{2m}\frac{s}{1+s}\tau}_{\text{Doppler shift}} \ll \Gamma\sqrt{1+s}. \tag{4.11}$$

Here, we defined the saturation parameter $s = I/I_{\mathrm{sat}}$. Rearranging equation (4.11) gives a limit for the pulse duration τ given a saturation parameter s

$$\tau \ll \frac{2m}{\hbar k_L^2}\frac{(1+s)^{3/2}}{s} \overset{^6\mathrm{Li}}{=} 2.16\,\mu\mathrm{s}\frac{(1+s)^{3/2}}{s}. \tag{4.12}$$

Even at a comparably high value of $s = 1$, this gives a limit of $\approx 6\,\mu$s. During the experiment, we usually stay below $6\,\mu$s imaging pulse duration. In addition to this Doppler limit, a second limit is imposed by the fact that the cloud should be approximately stationary during imaging. This means, the cloud should not move out of focus of the imaging system, as this would change s during the imaging process. Consequently, we require the travelled distance of the cloud to be smaller than the depth of field of the imaging system

$$d_\tau \ll d_{\mathrm{dof}}, \tag{4.13}$$

where the depth of field is given by $d_{\mathrm{dof}} = \lambda_L/\mathrm{NA}^2$. Inserting equation (4.10) and rearranging for τ gives

$$\tau \ll 2\sqrt{\frac{m d_{\mathrm{dof}}}{\hbar k_L \Gamma}\left(\frac{1}{s}+1\right)} \overset{^6\mathrm{Li}}{=} 1.04\times10^{-3}\,\mathrm{s/m}^{1/2}\sqrt{d_{\mathrm{dof}}\left(\frac{1}{s}+1\right)}. \tag{4.14}$$

The numerical apertures of the imaging systems are $\mathrm{NA}_{\mathrm{ANDOR1}} = 0.25$ and $\mathrm{NA}_{\mathrm{ANDOR0}} =$

0.1, which gives $d_{\text{dof, ANDOR0}} \approx 70\,\mu\text{m}$ and $d_{\text{dof, ANDOR1}} \approx 10\,\mu\text{m}$ for ANDOR0 and ANDOR1, respectively. The maximum pulse time for varying s are plotted in figure 4.6 together with the experimental values for the calibration in the next section.

High intensity imaging calibration

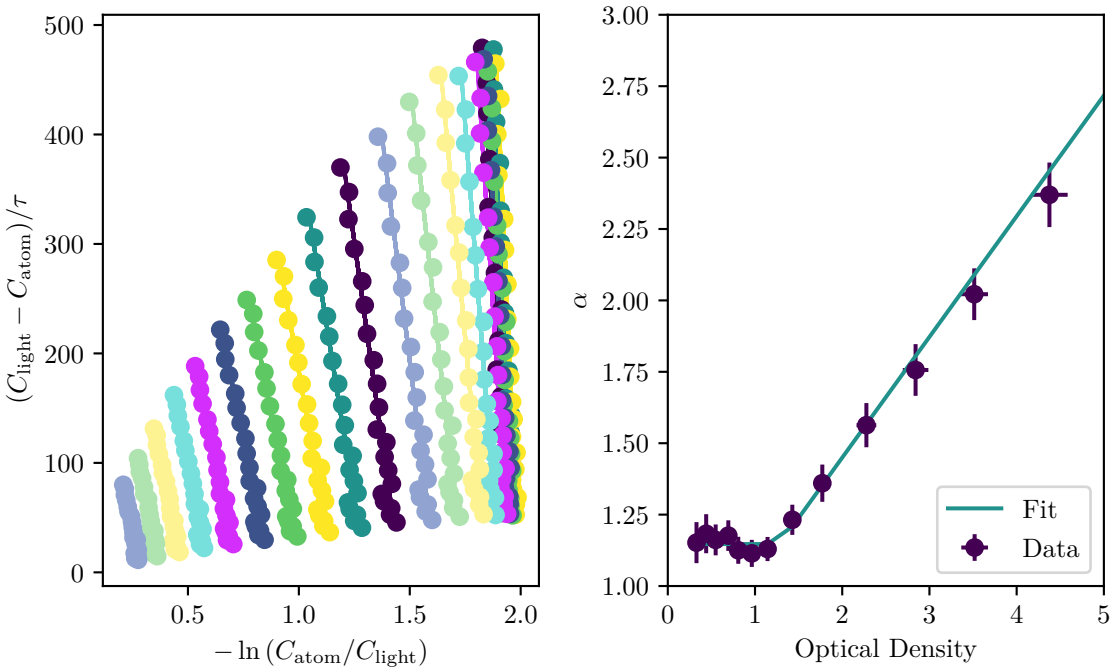

Figure 4.7: Results of the imaging calibration. (Left) Scatter plot that is used to extract α. Each colored line of datapoints represents a density bin, with each datapoint at different imaging counts (intensities). A linear fit to these lines yields α. **(Right)** Values of α for different values of the measured optical density. The turquoise line is a piecewise linear fit that is used to extract α at intermediate values of the optical density.

In order to analyse the density of the atomic cloud, we take two pictures per sequence. One picture with the atom cloud and one picture without the atom cloud at the same intensity and pulse length. We additionally use an averaged dark picture that is shot without any light to account for systematic errors. We subtract this dark picture from both the atom and the light picture. The modified Beer-Lambert-Law for high intensities is given by [147]

$$n\sigma_0 = -\alpha \ln\left(\frac{I_{\text{atom}}}{I_{\text{light}}}\right) + \frac{I_{\text{light}} - I_{\text{atom}}}{I_{\text{sat}}}, \tag{4.15}$$

where n is the atomic density, $\sigma_0 = 3\lambda_L^2/(2\pi)$ is the bare atomic cross section and I_{atom} and I_{light} are the intensities of the atom and light pictures, respectively. α is the value of interest, as it links the bare saturation intensity I_{sat} to the effective saturation intensity

$I_{\mathrm{sat}}^{\mathrm{eff}} = \alpha I_{\mathrm{sat}}$ which accounts for imperfections due to misalignment in the polarisation of the light or saturation effects. Knowing α in the above equation gives a direct connection between the measured intensity distribution and atom density distribution. In the experiment, we measure the intensity distribution of the pictures with a CCD camera. The intensity is proportional to the number of counts via

$$I_{\mathrm{atom/light}} = \frac{\hbar\omega_L}{\eta\tau A_{\mathrm{px/obj}}} C_{\mathrm{atom/light}}, \tag{4.16}$$

where $A_{\mathrm{px/obj}} = A_{\mathrm{px}}/M^2$ is the effective pixel size with physical pixel size A_{px} and magnification M. $C_{\mathrm{atom/light}}$ are the corrected counts of the light and atom picture, respectively. We calculate the corrected counts by subtracting the averaged dark picture from the recorded picture $C_{\mathrm{atom/light}} = C'_{\mathrm{atom/light}} - C_{\mathrm{dark}}$. Inserting this into equation (4.15) and assuming pulse length τ yields

$$n\sigma_0 = -\alpha \underbrace{\ln\left(\frac{C_{\mathrm{atom}}}{C_{\mathrm{light}}}\right)}_{\text{Optical Density}} + \frac{C_{\mathrm{light}} - C_{\mathrm{atom}}}{C_{\mathrm{sat}}\frac{\tau}{\tau_0}}. \tag{4.17}$$

Here, C_{sat} is the number of counts that corresponds to illumination with I_{sat} for a duration τ_0. In order to get a functional connection between counts and intensity, we varied the intensity $s = I/I_{\mathrm{sat}}$ by changing the RF power of the control AOM. We measured the power in the beam with a powermeter and and aperture to restrict the beam to the size of the CCD chip. Given the ^{6}Li saturation intensity $I_{\mathrm{sat}} = 2.54\,\mathrm{W/cm^2}$, the corresponding bare counts at the atom position were found to be 936(46) counts for ANDOR0 and 323(25) counts for ANDOR1 for a pulse length of 1 µs. From a rewritten equation (4.17),

$$\frac{C_{\mathrm{atom}} - C_{\mathrm{light}}}{\frac{\tau}{\tau_0}} = \alpha C_{\mathrm{sat}}\mathrm{OD} + C_{\mathrm{sat}} n\sigma_0, \tag{4.18}$$

it is now possible to extract α. We do this by taking several pictures with different counts (intensities) and plotting the left hand term against optical density. Since the second term on the right hand side is proportional to the atomic density, it is important to define equidensity bins for each picture. Grouping together corresponding bins from the different intensity-pictures, we can then do a linear fit to the data and extract $C_{\mathrm{sat}}^{\mathrm{eff}} = \alpha C_{\mathrm{sat}}$ as the slope. With this, it is possible to relate the optical density directly to the atom density and get an accurate measure of the atom number, see figure 4.7. We use this calibration method for all measurements that require an accurately measured atom number.

4.3 The rapid ramp technique

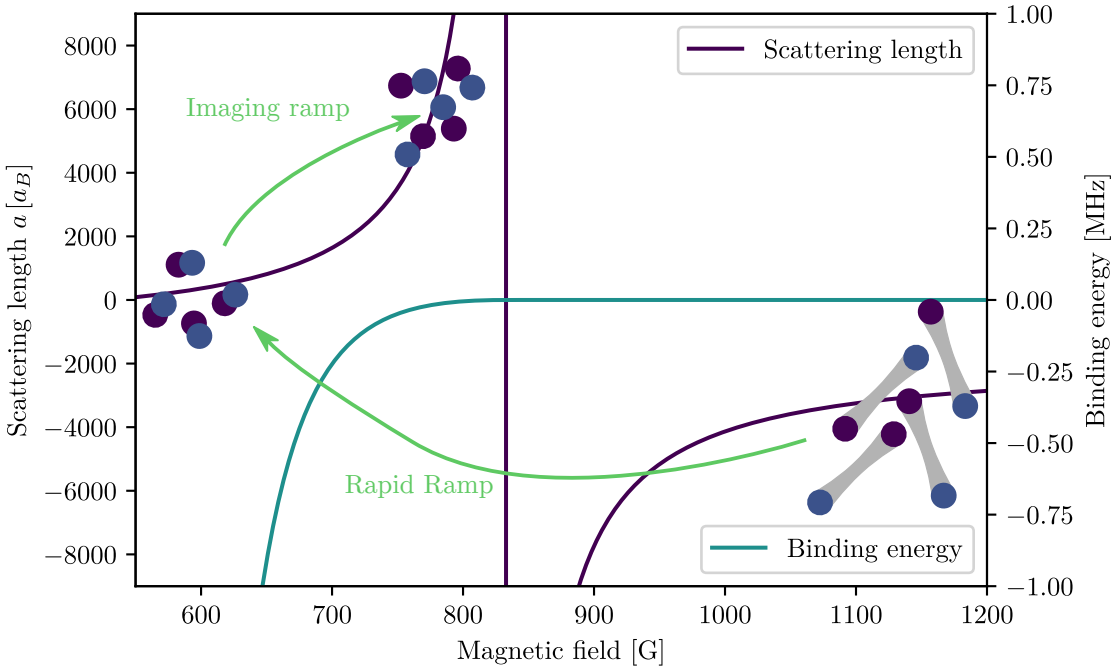

Figure 4.8: Visualisation of the rapid ramp and imaging ramp. By rapidly sweeping the magnetic field across the Feshbach resonance (purple curve), the weakly bound cooper pairs (right) are converted to tightly bound molecules with high binding energy (blue curve). For imaging, the field is ramped adiabatically closer to resonance, where molecules can be imaged at the same time as atoms but are still tightly bound. [51]

Accessing the momentum distribution of pairs on the BCS-side of the Feshbach resonance is complicated. As seen before, the binding energy of the Cooper pairs vanishes exponentially for weak interactions. For such fragile pairs, the observation of the momentum distribution in time-of-flight is not straightforward when the thermal energy $k_B T$ is comparable to the binding energy. In this case, the pairs break during time of flight and the measured density distribution does not reflect the momentum distribution before time of flight. One way to circumvent the breaking of pairs is to rapidly project across the Feshbach resonance and *convert* the weakly-bound Cooper pairs to tightly bound Feshbach molecules. These molecules will retain the momentum distribution of the Cooper pairs, however, they will not break during time of flight. This idea was initially developed at JILA [17] and later adapted for ^{6}Li at MIT [18]. Initially thought to be a sudden projection, the real process of transforming the Cooper pairs to tightly bound molecules is more complicated and connected to the different time-scales of the system. In particular, one does not want the momentum distribution to change due to the ramp dynamics or to have molecules form or break during the ramp. Theoretical investigations have looked into the interpretation of the ramp [150–152], especially for finite ramp times. These investigations showed a non-monotonic behaviour of the final

condensed molecule fraction versus the initial magnetic field. The formation of the condensate during the ramp was ruled out by investigating the duration of condensate formation by modulation of the magnetic field, see ref [153]. It was shown that the formation of the condensate takes much longer than the time to cross the Feshbach resonance. In conclusion, the rapid ramp technique has manifested itself as one of the most important tools to study the momentum distribution of a pair condensate on the BCS side of the resonance. Although the exact conversion efficiency is not known, fitting a bimodal distribution to the resulting momentum distribution gives a measure for the condensate fraction of the pair condensate.

4.3.1 Experimental implementation

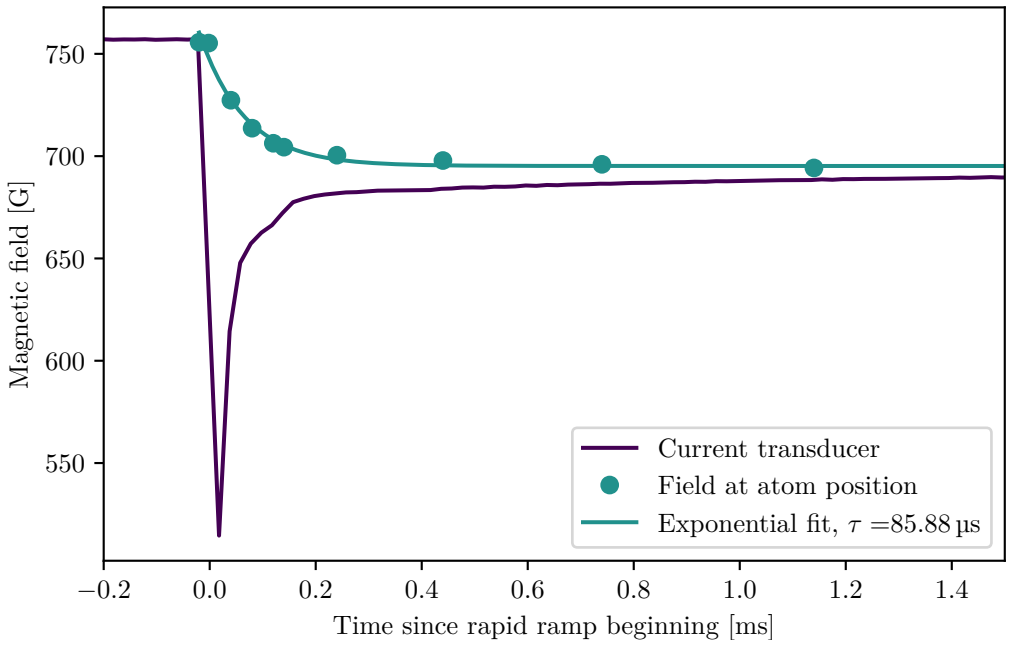

Figure 4.9: Experimental rapid ramp. The purple line shows the current through the Feshbach coils as measured by a current transducer. The turquoise datapoints show the real magnetic field at the atom position, measured by the Zeeman shift of the imaging transition. Data is fitted by an exponential function (turquoise line) with a characteristic time of 86 µs.

The goal of the rapid ramp is to quickly change the interaction, such that fragile cooper pairs are converted to tightly bound Feshbach molecules. Experimentally, this is done by ramping the magnetic field from its initial value to a final value close to the zero crossing of the Feshbach resonance on the BEC side, where the binding energy is high, see figure 4.8. Subsequently, the field can be ramped adiabatically towards resonance again, in order to image molecules and atoms at the same time. The initial fast change is realised by opening the H-Bridge connected to the Feshbach coils and to a snubber circuit. Additional

control is gained through the use of an additional IGBT[8] that opens a short-circuit across the power supply, allowing it to discharge more quickly. The current through the Feshbach coils measured by a current transducer[9] is shown in figure 4.9 (purple line). However, the actual magnetic field at the atom position changes much slower than the current in the coils suggests due to eddy currents in the chamber and surroundings. The actual field at the atom position was measured by observing the frequency shift in the imaging transition of the atoms. The resulting magnetic field is shown in figure 4.9 (turquoise dots). The initial change in magnetic field is roughly $3.7\,\mathrm{G\,\mu s^{-1}}$, allowing the transition from the BCS side to the BEC side in about 30 µs, depending on the initial field.

[8] 1MBI600U4B-120 FUJI
[9] LEM, TN 600-S ULTRASTAB

CHAPTER 5

Neural Network assisted detection of a Fermionic Condensate

This chapter describes how we use neural networks to explore the BEC-BCS crossover for an ultracold Fermi gas. We train a network to accurately predict the condensate fraction from time-of-flight pictures across the BEC-BCS crossover. So far, to extract the condensate fraction, the rapid ramp technique (described in section 4.3) had to be used in order to convert Cooper pairs on the BCS side to tightly bound molecules. However, this might induce problems in experiments where a change in the scattering length can obscure underlying physics such as for the quench experiments introduced in chapter 8. By training the network on the easily distinguishable regimes of high condensate fraction and thermal gas and predicting the transition point, it is possible to improve the resolution of measurements of the transition temperature across the whole crossover. Surprisingly, we found that the network can not only distinguish between a thermal and condensed sample, but also between Cooper pairs and tightly bound molecules, although no additional labelling was provided.

5.1 Introduction: The BEC-BCS crossover

The goal of this section is to give an overview of the general problem: The characterisation of the condensation behaviour across the BEC-BCS crossover and the problems connected to detecting it. It will also give a general purpose argument as to why neural networks are a good choice to tackle these problems.

5.1.1 Condensation in the BEC-BCS crossover

Following the theoretical picture introduced in section 2.4.2, this section introduces the experimental approach to the phase diagram of the BEC-BCS crossover. The crossover was made accessible in the early 2000's by the creation of samples of ultracold molecules of fermions [15, 154–156]. Shortly after, condensation in degenerate samples of these molecules was observed [16, 65, 157–160]. Far on the BEC side of the resonance, a bimodal distribution can be observed in time of flight signaling the presence of a condensate. The central part of this distribution is relatively broad as compared to a condensate of weakly interacting

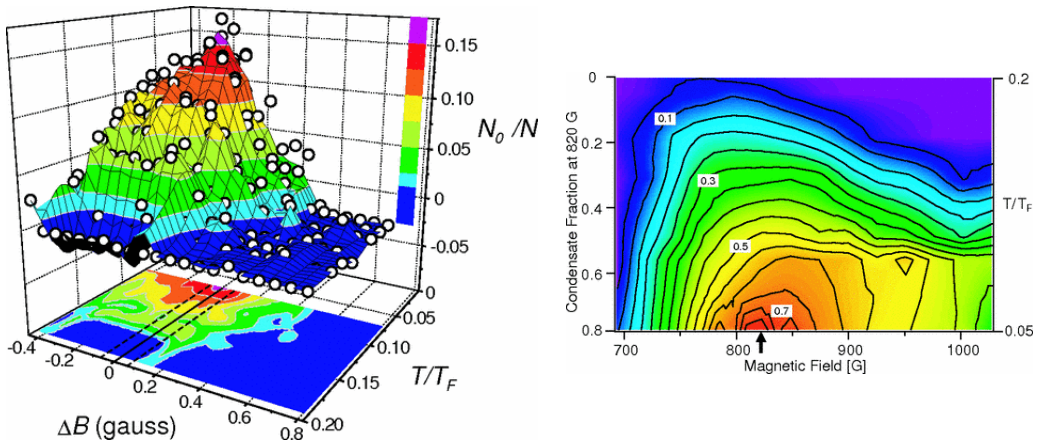

Figure 5.1: Early measurements of the condensate fraction in the BEC-BCS crossover. Regal et al. [17] (left) and Zwierlein et al. [18] (right) measured the temperature by fits to the cloud surface far above the resonance, where the gas is only weakly interacting.

bosons. This is a consequence of high mean-field energy in the condensate [51]. Towards the Feshbach resonance, the condensate part broadens and finally becomes indistinguishable from the thermal background, which prevents any meaningful fitting. This is the regime where the average distance between the molecules is comparable to the molecular size. The situation on the BCS side is made worse by an exponentially decaying pair binding energy which can fall below the energy scale of the local temperature, which will eventually break the pairs during time of flight. One technique to image the bimodal distribution in momentum space is the rapid ramp technique described in section 4.3. This technique was developed for ^{40}K at JILA [17] and later adapted to ^{6}Li at MIT [18] and allowed a more careful investigation of the phase diagram including the condensate fraction. The results are shown in figure 5.1. The onset of condensation has also been measured in imbalanced gases [161]. Theoretical investigations into these phase diagrams have generally confirmed the behaviour of the critical temperature and condensate fraction [162–164], however, the role of the rapid ramp on the measured quantities is hard to quantify. More recent theoretical work has investigated the critical temperature across the crossover with functional renormalisaton group [20], quantum Monte Carlo simulations [19] and diagrammatic approaches [21], see figure 5.2. Here, the mean-field limits for the BEC and BCS regime discussed in section 2.4.2 are recovered, however, the exact behaviour around unitarity is unclear. Recent measurements at unitarity, where the equation of state for a Fermi gas is known, have yielded an experimental value of $T_C/T_F = 0.167(13)$ [22]. In the following, we develop a scheme to improve the resolution of

T_C/T_F and generalise to regimes, where a rapid ramp might not be feasible.

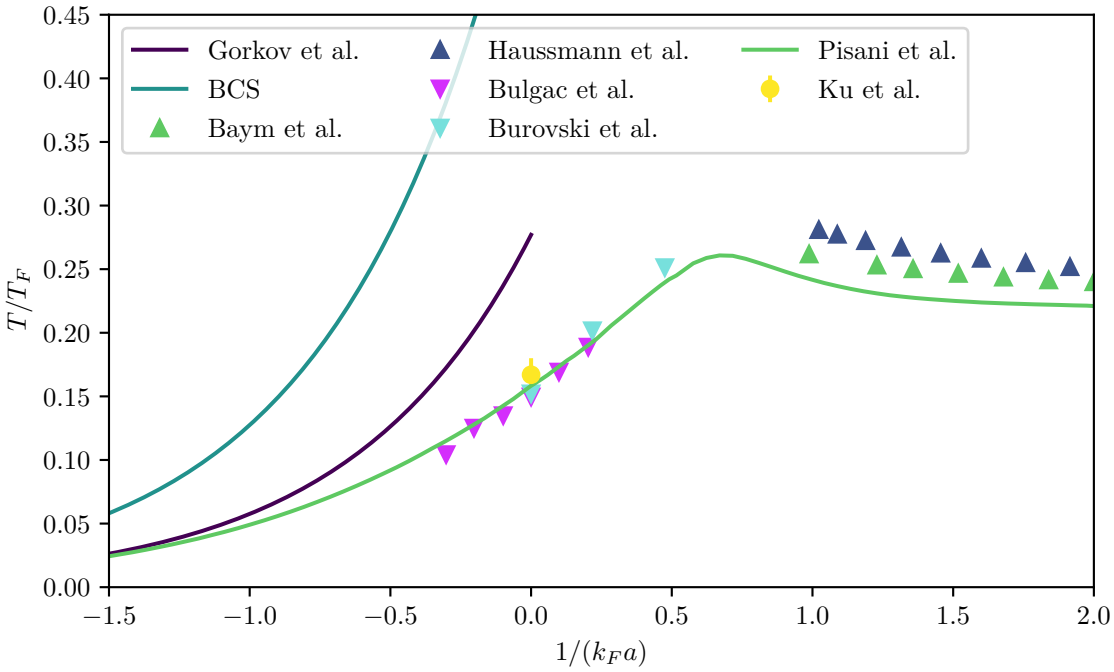

Figure 5.2: Theoretical and numerical studies of the critical temperature. We show the BCS limit (yellow line) and the Gorkov approximation [72] (turquoise line) as well as the theoretical approaches of Pisani et al. [21], Haussmann et al. [38], Baym et al. [165–167], Bulgac et al. [19] and Burovski et al. [168]. Also shown is the experimentally determined value from Ku et al. [22].

5.1.2 Neural networks for phase transitions

Machine learning and neural networks have manifested themselves as valuable tools in physics [23, 169]. The task of detecting a phase transition seems in particular suited to the strengths of the methods: In many cases, exact models only exists in limits far away from the transition boundary, but break down close to it. In these cases, machine learning methods are able to identify patterns in the data that might correspond to the states away from the transition point, and, more importantly, might be able to generalise the detected patterns across the transition point to yield an output that characterises the phase transition. Supervised learning with labels in the well-known region has been shown to be able to generalise well in different phase transition contexts on theoretical data [24–26, 28–30] and in the context of experimental data for topological phase transitions [31]. Phase transitions have also been characterised with adversarial networks [170], by confusion [103] and in an unsupervised manner [27, 171]. Apart from phase transitions, machine learning methods have also been applied to reconstruct quantum states [104, 172, 173] and the optimisation of experimental systems (see chapter 9 and references therein). In summary, machine learning provides a set of valuable tools to investigate phase transitions, due to the ability of generalisation of

the techniques through crossover territory. In the following section, we will apply a neural network to the BEC-BCS crossover to not only detect, but also quantify the presence of a condensate. This will make a detection of condensation possible without the additional rapid ramp procedure which, in some cases, might obscure underlying physics or might just not be feasible, e.g. in the limit of high fields and weak interactions [51]. The necessity for a rapid ramp, however, still remains, as it is used to generate labels for the training data. As such, it still dictates the accuracy of the magnitude of the condensate fraction.

5.2 A neural network for predicting the condensate fraction

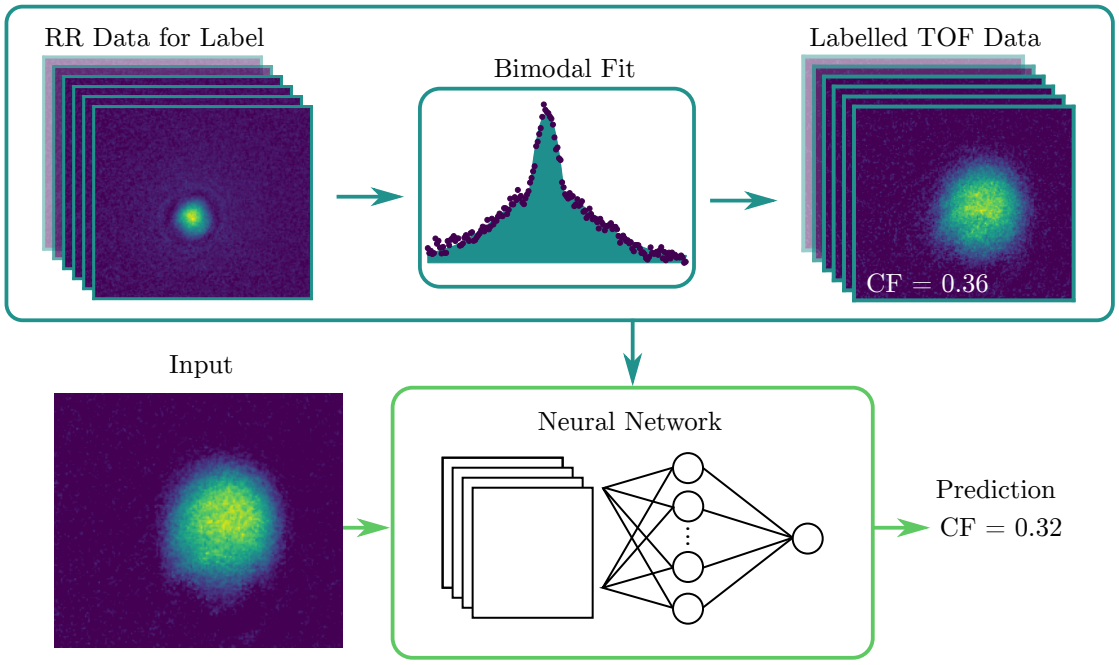

Figure 5.3: Introduction to neural network assisted detection of the condensate fraction. The process of labelling training data is shown in the turquoise frame. Two datapoints are needed for every datapoint of the training sample. For one, a rapid ramp picture which is fitted with a bimodal distribution to extract the condensate fraction. The condensate fraction then acts as a label for the time of flight pictures without rapid ramp that serve as an input to the neural network. Trained with this data, the network is able to predict a condensate fraction from an unlabelled input picture.

In order to circumvent the rapid ramp to detect the condensate fraction, the input to the network should be given by time of flight pictures at the corresponding interaction, where the condensate fraction should be determined, see figure 5.3. As mentioned in section 5.1, this means no bimodal distribution might be observed. However, a signature for feature detection should be present, although not detectable by current methods. This can be understood as follows: First, picture a weakly attractive gas in a homogeneous trap at zero temperature. As time of flight maps the momentum distribution onto the density distribution of the cloud, the

momentum distribution will show a clear edge at the Fermi surface, as compared to thermal distribution, which will remain gaussian. Going to finite temperatures, the Fermi edge flattens out and the signature in time of flight becomes less pronounced. This is further enhanced by the presence of an harmonic trap which further obscures the Fermi edge to a point where the difference between a thermal and Fermi distribution becomes indistinguishable from noise for fitting routines. However, the presence of a feature that allows the identification of a condensate can for example be found in the residuals for fits to the clouds [51]. This however, does make a quantification of the condensate fraction difficult and requires high precision and averaging. As the neural network learns the detection from just the experimental data it will also learn all experimental imperfections and day-to-day deviations, making it much more robust against changes in the experimental conditions. Furthermore, as the network is trained to predict from single shots, no averaging is needed. The problem with a direct detection of the condensate fraction without rapid ramp is highlighted in figure 5.4. The figure shows pictures taken at unitarity with rapid ramp (upper middle row) and without rapid ramp (lower middle row) and the corresponding averages along the y-direction. For the rapid ramp pictures, a clear change in the bimodal distribution can not only be seen in the averages, but even in the pictures of the cloud. The situation is different without rapid ramp. Apart from a slight atom loss due to heating of the cloud, no clear change in the density distribution is visible on the pictures or the averages. Using the time-of-flight pictures as the input to the network, the condensate fraction that acts as the label is extracted from pictures with rapid ramp at the same experimental conditions. The training process and the performance of the network will be discussed in the next section.

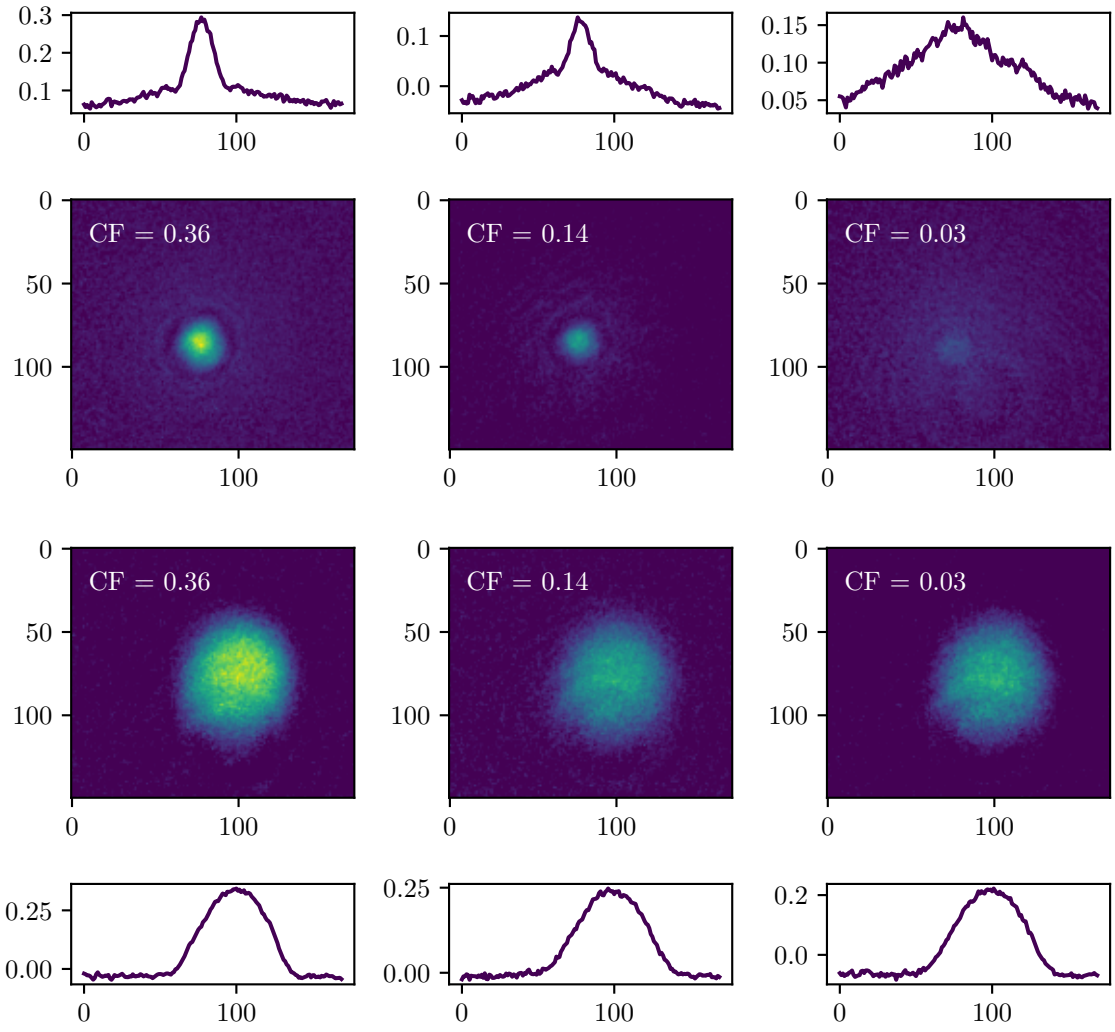

Figure 5.4: Example data for neural network assisted detection of the condensate fraction. Examples for pictures (two middle rows) with different condensate fractions indicated in the upper left corner. The pictures in the upper row are taken after rapid ramp and show a bimodal distribution. The curves above the pictures show the mean along the y-axes of the corresponding picture. For high condensate fraction, the central feature of the distribution is more pronounced. The bottom pictures show the same experimental conditions as the upper ones, but without rapid ramp. Apart from a decrease in optical density (which is accounted for by training at different fields), no obvious change of the imaged optical density takes place, which again is highlighted in the curves below the pictures that show the mean along the y-direction of the picture. The slight mismatch in center position of the clouds in the upper row as compared to the lower row is caused by the difference in time of flight for these pictures and a slight misalignment of the imaging axis with respect to gravity. All pictures within one row use the same color scale.

Magnetic field [G]	$1/(k_F a)$	Training datapoints
727	1.56	617
757	0.98	822
786	0.54	1040
835	-0.01	1731
894	-0.47	1884
914	-0.59	1801
Total	–	7895

Table 5.1: Amount of training data for different magnetic fields (interactions).

5.2.1 Generating data

Inspired by recent advances in image processing and the results reported in ref [31], we apply convolutional layers to the input time-of-flight pictures. Each convolutional layer is followed by max-pooling layers. The last layers of the network are fully connected layers. All layers use ReLU-activation. The final used network architecture is given in appendix A.1, however, several different architectures were tested with only minor differences in the performance. This indicates that the task is well-posed and neural networks with an architecture similar to ours are able to robustly learn to predict the condensate fraction. During training, the network is optimised via gradient descent with ADAM optimiser [113] (see section 3.4.2). All networks used in this section are implemented with TensorFlow [174]. For the training data, it is important to capture all regimes in the phase diagram: BEC, unitarity and BCS. Furthermore, to remove any accidental correlations between the pictures and the condensate fraction (e.g. cloud size $\leftrightarrow$ temperature $\leftrightarrow$ condensate fraction), several interactions in each regime are used for training. Decoupling of these accidental correlations is additionally achieved through a several months long period of data taking with breaks in between. This also guarantees the network seeing the experiment in different environmental conditions which helps generalisation. During the period of taking training data several changes in the experimental setup were taken that led to the cloud slightly moving and the imaging conditions slightly changing. We found that the network trained on newer data was still able to predict the condensate fraction of the old data thanks to equivariance of the convolutional layers. For the final training data we therefore use all training data taken. The amount of datapoints used for training at each interaction is shown in table 5.1. Each training data point $z_i = (x_i, y_i)$ is a tuple of a time of flight picture x_i that serves as the input and a label y_i extracted from a rapid ramp picture taken at the same experimental conditions. As x_i and y_i can not be measured in the same experimental cycle, each training data point consists of two independent experimental cycles. The process of training the network is sketched in figure 5.3.

For each experimental cycle we use samples of ^{6}Li in the lowest two hyperfine states $|1\rangle$ and $|2\rangle$ in a Gaussian trap with trap frequencies $(\omega_x, \omega_y, \omega_z) = 2\pi \times (168, 166, 238)\,\mathrm{Hz}$ in the harmonic limit. After evaporating to degeneracy close to the Feshbach resonance, we adiabatically ramp to the magnetic field corresponding to the desired interaction strength within 200 ms. After a variable heating time followed by an equilibration time of 50 ms we release the cloud by suddenly switching off the trap. For the input pictures x_i, the cloud

expands in time of flight for 5 ms which represents a good balance between signal and noise. For the labels y_i, we do a rapid ramp as described in section 4.3 and extract the condensate fraction through a bimodal fit.

Varying the temperature

In order to cover the whole phase diagram and train on different condensate fractions for a given interaction, the temperature of the sample has to be varied. In order to heat the atomic cloud we use a trap release heating procedure [175, 176]. For this, the confinement by the dipole trap at the final interaction value is suddenly reduced for a variable amount of time. For the horizontal dipole beam the reduced power is 50 % of the initial power and the vertical beam is turned off completely. Afterwards, the confinement is suddenly returned to its initial value for both beams. Varying the duration of the low confinement results in a variable heating of the cloud. The purple datapoints in figure 5.6 show the condensate fraction after variable heating time. In order for the cloud to thermalise after the heating, a hold time of 40 ms before the beginning of time of flight (and rapid ramp) is employed. The heating introduces an atom loss that is proportional to the heating time. This loss is accounted for when calculating the Fermi energy and Fermi temperature.

Training and performance of the network

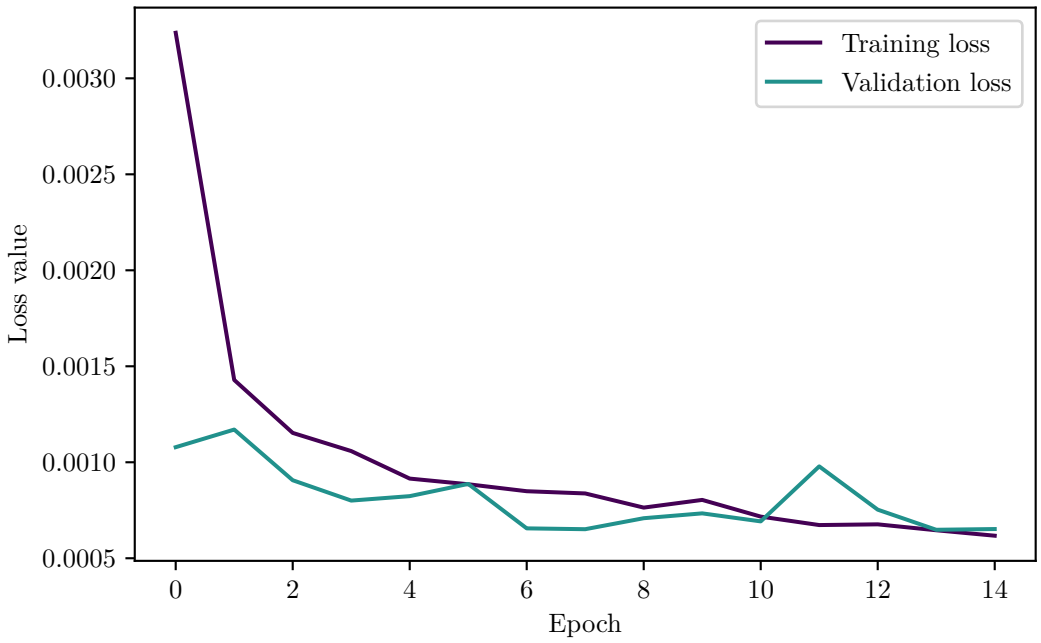

Figure 5.5: Loss on the training and validation datasets for all epochs of the training process. Apart from fluctuations, no relevant increase can be seen on the validation loss.

During training, the performance of the network can be characterised by the loss value (here: mean squared error), which the training network tries to minimise. However, when only using the training loss as a figure of merit, the network might overfit. For this reason, 10 % of the data is split and kept as a validation set. The network is trained for 15 epochs (full runs) on the training data. The datapoints are shuffled before every epoch and the loss is calculated. After each training epoch, additionally, the loss is calculated on the unseen validation dataset. Overfitting can then be detected by a decrease in loss on the training dataset but an increase in the validation dataset, signalling a loss of generalisation of the model on unseen data. Both training loss and validation loss are shown in figure 5.5. In order for the labels to be accurate, points with a condensate fraction lower than $y = 0.05$ were omitted during training due to the fitting procedure becoming inaccurate. Additionally, the condensate fraction of all points with a heating time obviously away from the transition point is set to 0. This takes care of the problem that a bimodal fit will *always* fit a parabola and therefore always get a non-vanishing condensate fraction. By setting the datapoints of the thermal samples to zero, a sudden increase in condensate fraction might be detectable more straightforwardly. The labels of exemplary training data and the neural network predictions for the pictures corresponding to these labels is shown in figure 5.6. Here, the grey shaded area marks the datapoints that are omitted during training. All predictions in this regime are generalisations from the neural network. We can also see that for long heating times the network predictions go to zero, indicating a purely thermal cloud.

In order to further check the generalisation ability of the trained network, we tested its predictions on labelled data at $1/(k_F a) \approx -0.93$ far away from the highest training dataset at $1/(k_F a) \approx -0.59$, shown in the left plot of figure 5.7. Here, the onset of the condensate fits very well with the onset predicted by the labels. The condensate fraction predicted by the network is systematically higher, however, it should be noted that the condensate fraction at this field is constantly below the training threshold and therefore in a regime, where the bimodal fit becomes inaccurate. Additionally, at these high fields, the rapid ramp might have a non-neglectable influence on the measured condensate fraction. This means, the difference in prediction does not necessarily stem from the increasing error in the network, but could also originate from the generalisation ability being more accurate than the bimodal fit plus rapid ramp effects in this regime. This, however, is hard to quantify without an in-depth understanding how the network calculates the condensate fraction. A second check was done by using pictures of a $|1\rangle + |3\rangle$ condensate instead of the $|1\rangle + |2\rangle$ condensate the network was trained on, shown in the right plot of figure 5.7. In principle, the physics of both condensates is equivalent, however, due to the additional $|2\rangle \rightarrow |3\rangle$ transfer with associated atom losses, the different center position of the Feshbach resonance and the difference in imaging, the general experimental situation is different. The interaction of the condensate at the measured field of 717 G is $1/(k_F a) \approx -0.41$. The network again is able to predict the onset of the condensate fraction. The numerical value, however, has a discrepancy as compared to the bimodal fit. Again, it is hard to quantify whether this is an actual error of the network or caused by the change in rapid ramp due to the width and position of the Feshbach resonance as compared to the generalisation ability of the network. Although difficult and subject of ongoing research in computer science, the next section will have a brief look on how to quantify whether the network learns physical features in order to determine the condensate fraction.

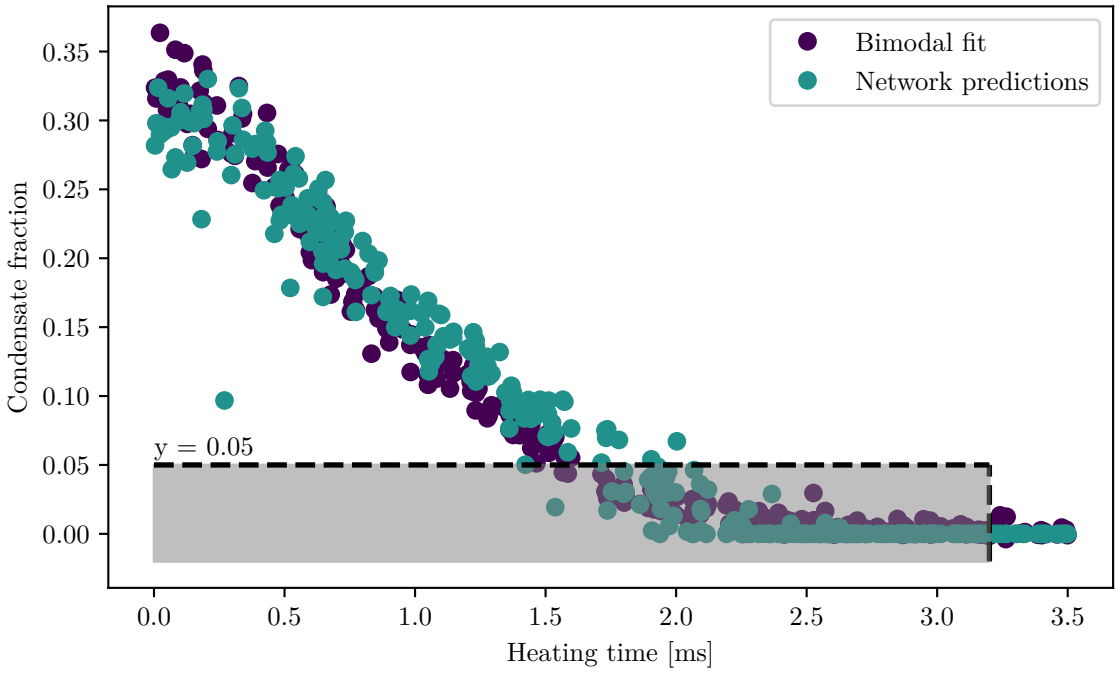

Figure 5.6: Example for the network predictions on labelled data. The purple datapoints represent values of the condensate fraction achieved by a bimodal fit after rapid ramp. The turquoise datapoints are predictions by the network from pictures at the corresponding heating time without rapid ramp. During training the datapoints in the grey-shaded area are masked out, as in this regime the bimodal fit is not reliable. All datapoints to the right of the shaded region are artificially set to zero condensate fraction, as no condensate is to be expected. Shown here is data close to unitarity with $1/(k_F a) \approx -0.01$.

5.2.2 What does the network learn?

It cannot be excluded that the network learns accidental correlations that hinder generalisation: To a reasonable good approximation within the experimental stability, the condensate fraction is not dependent on the atom number, however, because the heating introduces atom loss, there is an empirical functional connection between condensate fraction and atom number that could be learned by the network. In this work, we try to circumvent this kind of correlation by training at different interactions and throughout a prolonged period of time to capture different conditions within our experimental setup. Since neural networks of this high complexity still act like a black box, the argument might still hold with less obvious correlations. For these reasons, in the following section, we will try to show that a neural network is able to self-sufficiently extract important physical quantities from pictures, and also analyse which pixels on a an input picture change the condensate fraction and try to connect this to the physical picture.

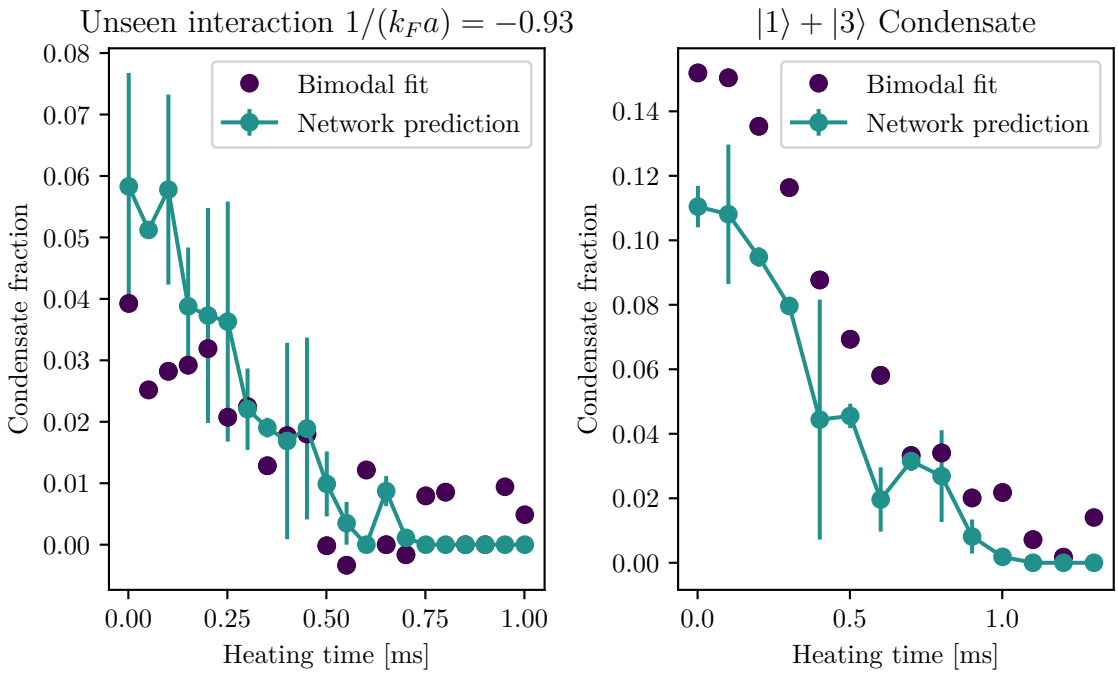

Figure 5.7: Verification of the generalisation abilities of the network. The left plot shows the network predictions and labels at $1/(k_F a) \approx -0.93$, which is far away from the highest training field at $1/(k_F a) \approx -0.59$. The right plot shows the results on a labelled dataset for a $|1\rangle + |3\rangle$ condensate at $1/(k_F a) \approx -0.41$. Although the physics does not change, the size of the cloud, the imaging and the atom number does.

Analysing the networks activations

Generally, neural networks act as a *black box*, that does not per se allow to grasp how input features map to the output. This comes from the fact that during training only input and output define the task of the network. How the network processes the input information, however, is elusive, especially when the network architecture is complex. Interpreting neural network processing is an ongoing and active field of research in computer science [178] and is important especially in cases where not only the solution of a task is important, but also understanding said solution. In the context of physics, the desired information is not only hidden in the final outcome, but also in the way the problem is solved. One ansatz to systematically investigate how the network processes information are approaches that try to estimate the importance of an input neuron to an output neuron. Here, we use DeepLIFT [177], a backpropagation-based approach that assigns importance scores to the inputs for a given output. This should yield a picture of *what the neural network looks at* when deciding on the condensate fraction. The DeepLIFT algorithm takes an input and returns an output in the same shape as the input. Visualising this output yields a picture, where the pixel values characterise the importance of each pixel to the output of the original network. The top row of figure 5.8 shows the different steps at a field close to unitarity. On the left is the average

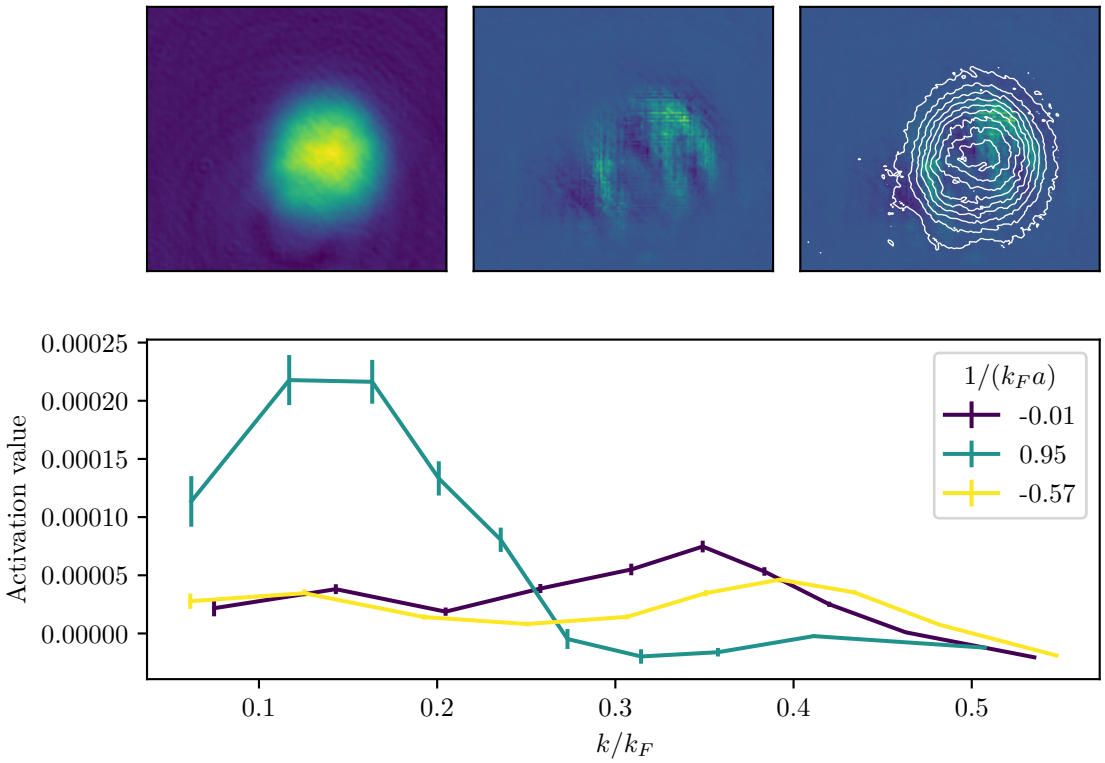

Figure 5.8: Analysis of the importance of activations in the initial layer of the neural network with DeepLIFT [177]. The pictures in the upper row show from left to right an averaged picture at unitarity, the averaged importance score per pixel in the input layer and the middle picture overlaid with density bins from the left picture. The lower graph shows the average importance score within momentum bins versus the momentum in units of k_F for three different interaction strengths.

of many pictures taken at different heating time. Shown in the middle is the average of the activation maps acquired from the DeepLIFT algorithm given all the inputs that go into the left picture. The picture on the right shows a density binning of the cloud superimposed with the activation picture. Analysing the mean activation value within each density bin allows for a conclusion which part of the cloud is most important for the determination of the condensate fraction. The lower plot in the figure shows the average activation value within the different OD bins (corresponding to different momenta in time-of-flight) for different interactions. On the BEC side, the network pays more attention to the cloud-center, whereas towards unitarity and the BCS side of the resonance, the activation per bin flattens, with a slightly elevated value on the edge of the cloud. Taking these curves for several fields in the BEC-BCS crossover and assembling them to a 2D diagram yields figure 5.9. Here, the dashed line marks unitarity. For low fields below unitarity, the network focuses on the center

of the cloud. This can originate from a hidden bimodal distribution. Towards the BCS side, the activations flatten and pay most attention to the low-density region of the cloud, which should contain the tail of the momentum distribution, effectively yielding a temperature measure.

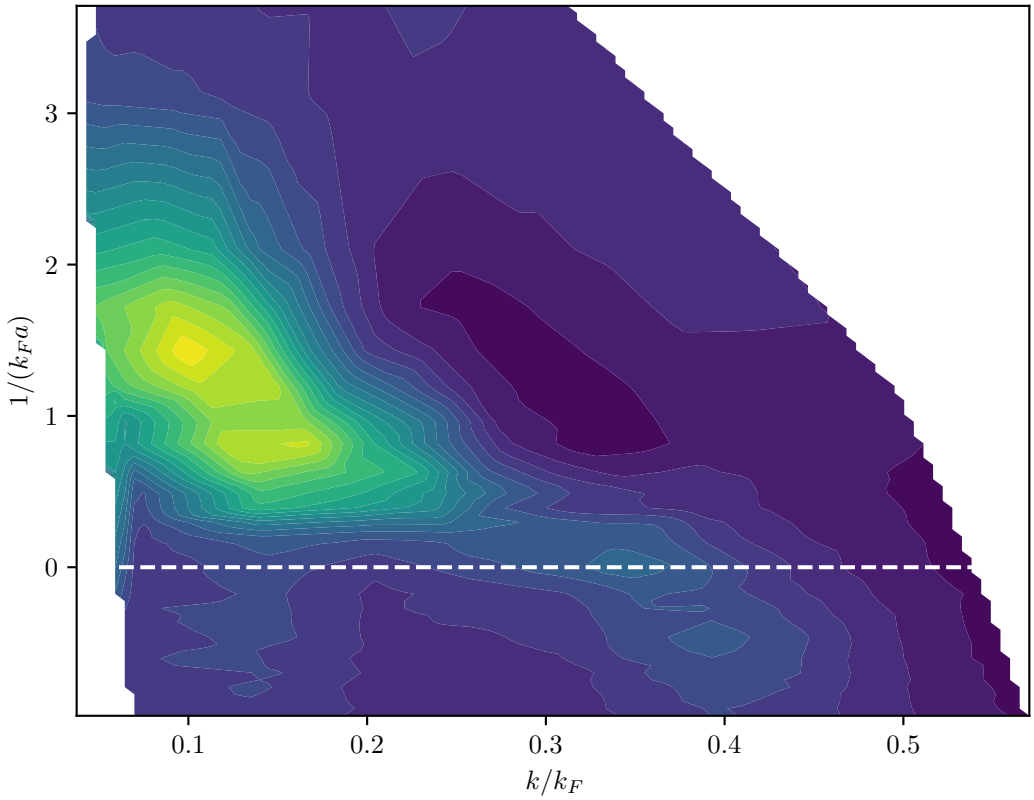

Figure 5.9: 2D diagram of the averaged importance score across the BEC-BCS Crossover. A 2D diagram of the averaged importance score per momentum bin for different interaction strengths across the BEC-BCS crossover.

5.3 Thermometry of an ultracold Fermi gas

With the neural network being able to determine the condensate fraction from time-of-flight pictures, the remaining missing information is the exact mapping between heating time and temperature T or T/T_F. We tested several methods for a reliable thermometry. Firstly, a preliminary thermometry that yields the temperature scaling by comparing the heating time to the heating time required to loose the condensate at unitarity, where T_C is known. Applying this method to other interaction strengths relies on the assumption that heating behaves the same across the crossover. Although this assumption is naive, we will see that it

is true within the accuracy of our measurements. Two thermometry methods were tested that do not rely on calibration at a fixed interaction strength. For one, we tested whether remaining sodium after evaporation in the magnetic trap could be used as a temperature probe. We found, however, that the lifetime of the sample is not sufficient for thermalisation between lithium and sodium after subsequent evaporation in the dipole trap. Consequently it was not possible to use this thermometry method. A second method that we tested relies on measuring the potential dependence of the in-trap density on the cloud surface. For this, the trapping potential has to be measured accurately and independent of the cloud density. The results of this thermometry method present the needed conversion from heating time to temperature throughout the crossover.

5.3.1 Preliminary thermometry

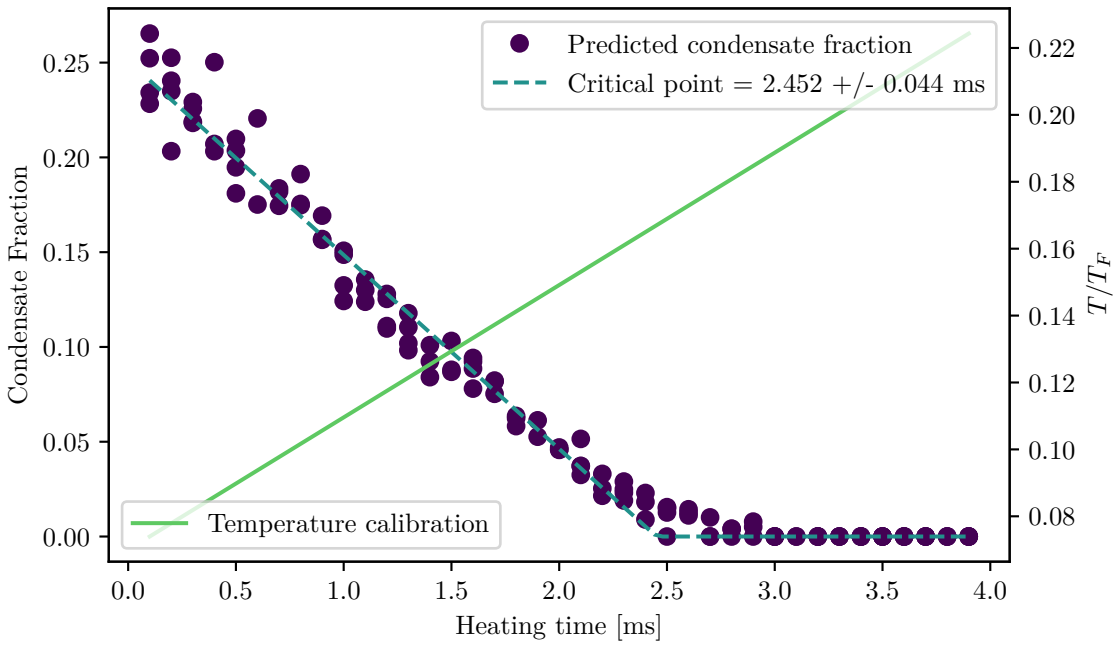

Figure 5.10: Preliminary temperature calibration. Purple dots show the predicted condensate fraction at unitarity. The turquoise dashed line shows a simple piecewise linear fit to estimate the heating time corresponding to the critical point. The green line shows a plot of equation 5.1 and is used to convert trap heating time to temperature.

We measure the initial temperature of the cloud after evaporation by adiabatically ramping the magnetic field to its value on the BCS side, where the condensate fraction – as measured by the rapid ramp – vanishes. For a cold enough sample this will be at the edge of the crossover regime with $1/(k_F a) \approx -1$ where the Gorkov approximation from equation (2.51) holds to reasonable accuracy and the critical temperature decays exponentially with $1/(k_F a)$. Measuring the interaction at which the condensate fraction vanishes then yields the

temperature of the gas. We measure the initial temperature of the gas to be $T_{\text{initial}}/T_F = 0.06(2)$. To get an estimate of the heating rate, it is possible to use the measured value $(T_C/T_F)_{\text{ref}} = 0.167(13)$ at unitarity from ref [22]. The temperature of the sample T can then be estimated by

$$T/T_F(\tau) = \frac{(T_C/T_F)_{\text{ref}} - T_{\text{initial}}/T_F}{\tau_C} \cdot \tau + T_{\text{initial}}/T_F = \underbrace{\frac{0.167 - 0.07}{2.452}}_{=0.039}\,\text{ms}^{-1} \cdot \tau + 0.07, \quad (5.1)$$

where τ_C is the heating time corresponding to the critical point at unitarity and τ is the current heating time. The determination of τ_C from a plot of the condensate fraction versus heating time and a plot of equation (5.1) is shown in figure 5.10. Although this method might work at unitarity and serve as a cross-check at different interactions, there is no immediate reason why the temperature scaling should not depend on the interaction. For this reason, it would be much more desirable to have a thermometry at every interaction to calibrate the temperature curve.

5.3.2 Sodium thermometry

One idea to achieve a reliable thermometry was to use some remainder of the sodium from the evaporation in the magnetic trap as a temperature probe. This was inspired by earlier measurements that used ^{41}K as a probe for a ^{6}Li Fermi sea [179], although here, the goal was to use as little sodium as possible in order to not change the lithium cloud as compared to the phase diagram data. For this reason, instead of measuring a bimodal distribution of a condensed sodium sample as shown in ref [179], the goal was to extract the temperature from the thermal distribution of a sodium sample with an atom number below the critical number required for Bose-Einstein condensation. In order to create such a sodium probe and measure the temperature with it, we had to overcome some experimental problems: For one, in order to image ^{23}Na at the corresponding fields around the lithium Feshbach resonances a new scheme for imaging had to be developed. This is because the magnetic moment of the $|F = 2, m_F = 2\rangle$ state used in the magnetic trap detunes the imaging frequency far away from the laser frequency. We found, however, that at fields close to the Feshbach resonance, the Zeeman shift of the $|F = 1, m_F = 1\rangle$ state cancels with the initial detuning. For this reason, we implemented a Landau-Zener sweep to transfer the sodium from $|F = 2, m_F = 2\rangle$ to $|F = 1, m_F = 1\rangle$, which enabled imaging of the latter state at a laser frequency close to the original frequency. A second problem was to find a good balance between the sodium atom number and lithium atom number. During the normal experimental cycle, after evaporation in the magnetic trap and transfer of the lithium into the dipole trap, only non-detectable amounts of sodium are present in the dipole trap. For this reason, we had to control the amount of sodium transferred to the dipole trap by an adjustment of the final magnetic trap evaporation frequency. This in turn leads to a lower lithium atom number due to the increased temperature when loading into the dipole trap. The effect is highlighted in the upper panel of figure 5.11, showing both lithium and sodium atom number for different values of the final evaporation frequency. Having a strong lithium signal was made more difficult by a measured reduction of the lifetime by a factor of 10 from $\approx 13\,\text{s}$ to $\approx 1\,\text{s}$ due to the presence

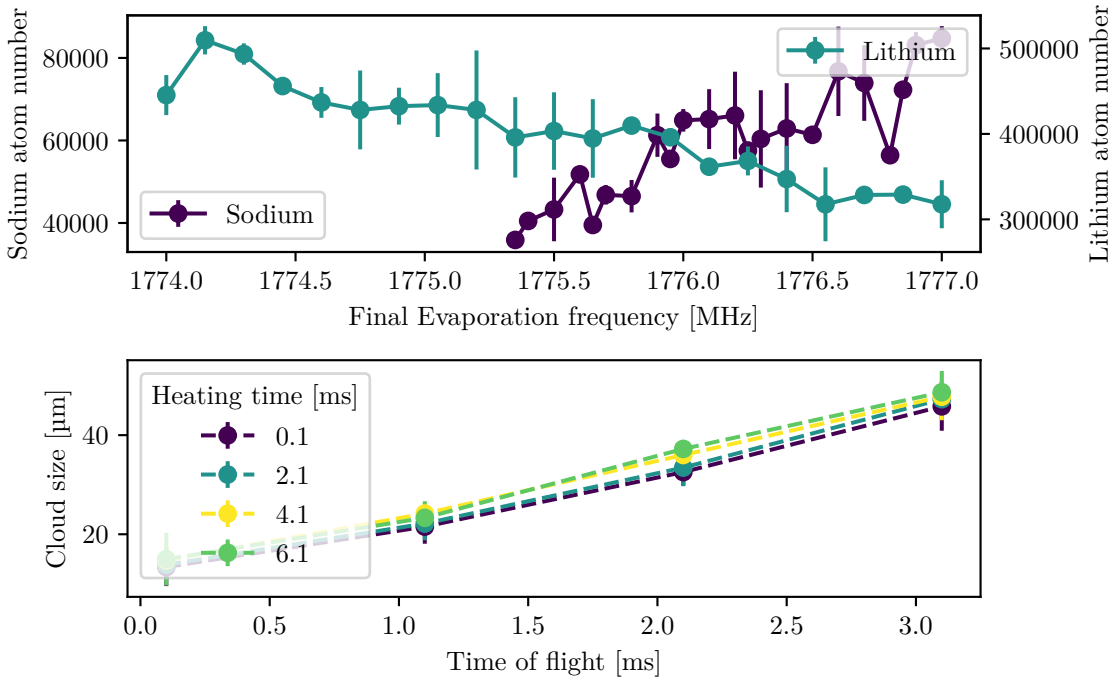

Figure 5.11: Illustration of the problems with sodium thermometry. (Top) Lithium and sodium atom number without additional waiting time versus final evaporation frequency. When carrying along sodium, the evaporation in the magnetic trap has to stop early. This increases the temperature of the lithium cloud and hence reduces the final atom number. **(Bottom)** Increasing the heating time does not change the expansion of the sodium cloud. At the same time, the condensate fraction of the lithium cloud decreases indicating heating on the lithium sample (not shown).

of sodium in the final dipole trap. Since heating, magnetic field ramps and thermalisation time together are in the order of magnitude of the lifetime, the number of lithium atoms at the time of the temperature measurement was strongly reduced. In the end, the number of Lithium atoms was not sufficient to thermalise well with the sodium sample. This lead to no measurable change in the expansion of the sodium probe in time of flight, as shown by the lower panel in figure 5.11. In the end, we found no set of parameters that would enable a reliable thermometry of a lithium cloud that resembles the cloud used in the phase diagram data. A new thermometer had to be found.

5.3.3 In-trap thermometry

Another way to measure the temperature of the lithium cloud at cold temperatures is to measure the potential dependence of the density in the wings of the distribution. Since the density is low on the cloud surface, the local k_F decreases, effectively reducing $1/(k_F a)$ down to the non-interacting case for low densities. In this regime, the density distribution can be treated as an ideal Fermi gas. The three-dimensional density distribution is then given by (see section 2.1.2)

$$n(\mathbf{r}) = \frac{1}{\lambda_{\mathrm{dB}}^3}\mathrm{Li}_{3/2}\left(-\exp\left(\frac{\mu - V(\mathbf{r})}{k_B T}\right)\right). \tag{5.2}$$

In the case of absorption imaging, the observed density distribution is integrated along the imaging direction, yielding the column density. However, the exponential function factors into the spatial directions, providing only a numerical prefactor that becomes irrelevant on the cloud surface [56]. For small arguments the polylog function can be approximated by the identity $\mathrm{Li}_y(x) \approx x$, which is valid in the low density region that is of interest. Using that for low densities the optical density is proportional to the atom density, we arrive at

$$\log(\text{Optical Density}) = A \underbrace{-\frac{1}{k_B T}}_{\text{Slope}} \cdot V(\mathbf{r}), \tag{5.3}$$

where A is a temperature independent offset that results from the prefactors of the exponential decay in the column density and the proportionality factors between optical density and column density. With the optical density being measured during imaging and the potential $V(\mathbf{r})$ known from a separate measurement, equation (5.3) provides the means of calculating the temperature. For this, a linear function is fitted to the logarithm of the optical density on the cloud surface versus the potential. The slope of this function then yields the temperature. In order to do so, two things have to be made sure. For one, the fitting region has to be estimated by estimating where the cloud becomes non-interacting. Secondly, the trapping potential has to be measured accurately.

Estimation of the non-interacting density regime

In order to estimate where in the trap the trapped gas becomes approximately non-interacting, the interaction term in the BCS Hamiltonian is treated perturbatively. The expectation value of the interaction energy per particle is then given by

$$\langle H_{\mathrm{int}} \rangle = \int d\Omega \frac{g n^2(\mathbf{r}')}{4}. \tag{5.4}$$

Assuming all densities to stay constant within this small volume, integrating over the volume of one particle $n^{-1}(\mathbf{r})$ yields

$$\frac{E_{\mathrm{int}}(\mathbf{r})}{E_F} = \frac{g n(\mathbf{r})}{4 E_F}. \tag{5.5}$$

Inserting the density distribution of a harmonically trapped gas at zero temperature $n(r) = n_0\left(1 - \left(\frac{r}{R_{\mathrm{TF}}}\right)^2\right)^{3/2}$ with R_{TF} being the Thomas-Fermi-radius for trap frequency

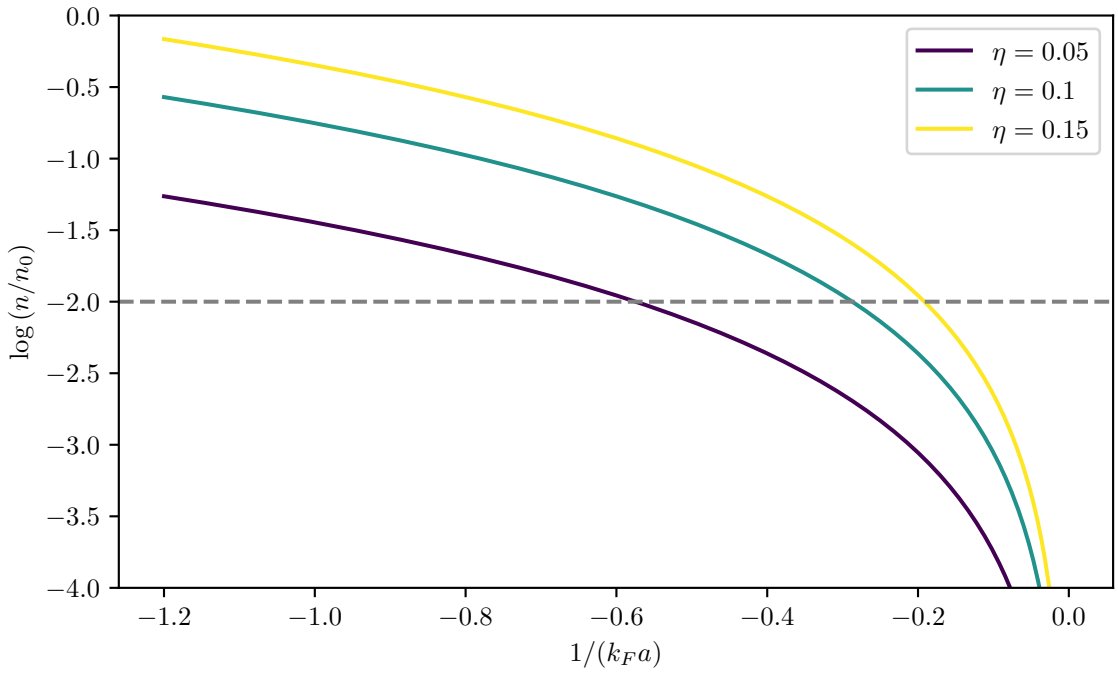

Figure 5.12: Theoretical estimation of the non-interacting density regime. Given an interaction $1/(k_F a)$, the curves show up to which density the requirement $E_{\text{int}} = \eta E_F$ is fulfilled. The grey dashed line indicates a density decrease to a fraction of $1/e^2$ of the initial density.

ω, $R_{\text{TF}} = \sqrt{2E_F/(m\omega)}$ and $n_0 = k_F/(3\pi^2)$ and the interaction given by $g = -4\pi\hbar^2 a/m$ yields an expression for the interaction energy in units of the Fermi energy dependent on the dimensionless radius $\tilde{r} = r/R_{\text{TF}}$.

$$\frac{E_{\text{int}}(\tilde{r})}{E_F} = \frac{2k_F a}{3\pi}\left(1-\tilde{r}^2\right)^{3/2}. \tag{5.6}$$

Assuming that the gas becomes non-interacting, when the interaction energy has decreased to $E_{\text{int}} = \eta E_F$, it is possible to find a radius (or density), from which on the interaction energy becomes neglectable. An exemplary plot of this criterion is shown in figure 5.12 for different values of η. Each curve shows, which density value corresponds to the non-interacting regime given interaction $1/(k_F a)$. In the measurement shown in figure 5.15, the chosen fitting region starts at a corresponding value of -4 (note the offset in the measurement), which makes sure that the fitted region is non-interacting except when getting very close to unitarity. Towards unitarity, we can use the arguments provided in ref [180], that even the unitary gas enters the Boltzmann regime when $\exp((\mu_0 - V(\mathbf{r}))/(k_B T)) \ll 1$. Here, μ_0 is the chemical potential in the center of the Trap.

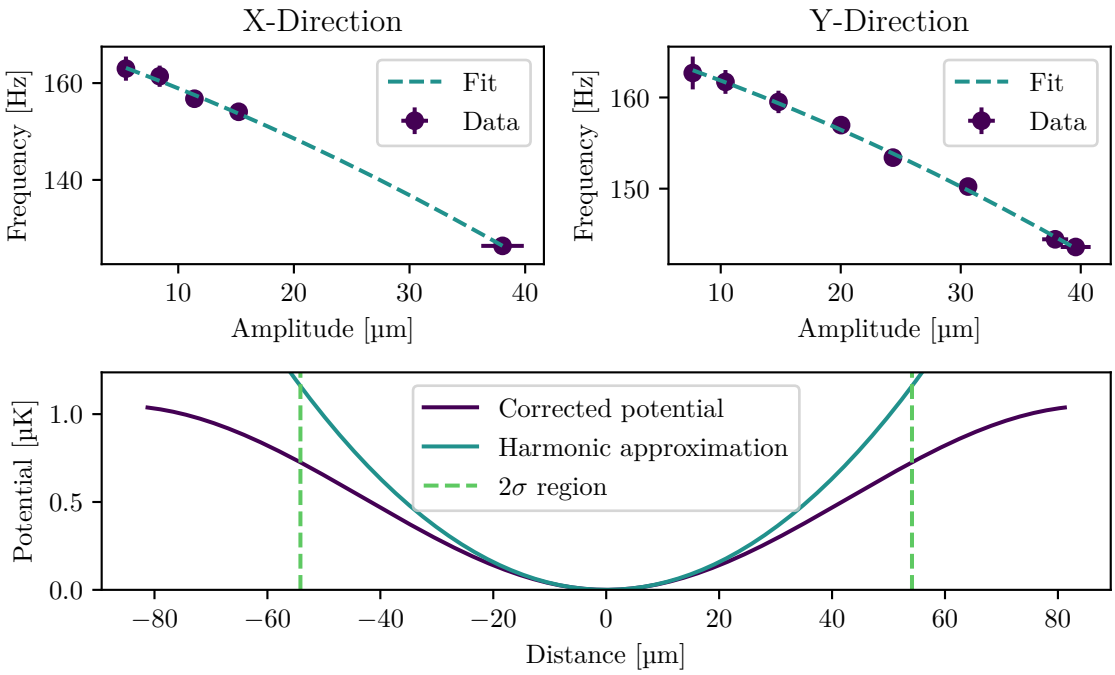

Figure 5.13: Mapping the trapping potential along the principal axes from trap frequency measurements. (Top) The plots show the measured trap frequency in the principal directions for different excitation amplitudes of the monopole mode. A drop in the frequency is visible due to the anharmonicity of the potential. A quadratic fit (turquoise dashed line) yields a position dependent frequency. **(Bottom)** Corrected potential (purple) and harmonic approximation (turquoise). A clear deviation from the harmonic approximation can be seen in the outer regions of the potential. The green dashed line shows the typical 2σ value of a Gaussian fit to the cloud density.

Mapping of the trapping potential

The dipole trap potential is formed by two crossed Gaussian beams. Usually, the trap is well characterised by a harmonic approximation as the atoms reside close to the center of the potential. In this case however, the region of interest is the low density region on the cloud surface. Outside of the center, non-harmonic effects might play a role for an accurate thermometry. For this reason, the anharmonic region of the potential has to be characterised. One way to do this is to measure the trap frequency for different excitation amplitudes within the trap. In case of a harmonic potential the measured trap frequency does not depend on the excitation amplitude of the oscillation. In the real case however, a large excitation amplitudes will probe the wings of the Gaussian potential, effectively slowing down the frequency due to the anharmonic motion. The upper plots of figure 5.13 show the frequency behaviour for changing amplitude together with a quadratic fit for both principal axes when imaging along z-direction. With the resulting amplitude-dependent frequency, the potential can be approximated by a corrected harmonic potential

$$V(x,y) = \frac{1}{2}m\left(\omega_x(x)^2x^2 + \omega_y(y)^2y^2\right). \tag{5.7}$$

Note, that the trap frequencies $\omega_i(i) = a_0 + a_1 \cdot i + a_2 \cdot i^2$, now depend on the corresponding coordinate $i \in [x, y]$. The lower panel in figure 5.13 shows a comparison between the corrected potential and the harmonic approximation. To access the low density region of the cloud with reasonable signal-to-noise ratio requires azimutal averaging of the optical density. The problem is, that equation (5.7) only provides an accurate description along the principal axes of the cloud. This is because all cross terms between x and y that are required for a series expansion of a 2D gaussian are not regarded. For this reason, a 2D gaussian is fitted to the potential only along the principal axes. With the original potential being 2D gaussian, this takes care of all cross-terms in the correct way. Figure 5.14 shows a comparison between the profile gained from measuring the trap frequencies and the density distribution of the atoms. The upper left picture shows the average of 30 density distributions and equi-density lines. The upper right picture shows the same equi-density lines (dashed) and equi-potential lines (solid) from the reconstructed potential. The general shape of the cloud is recovered, however some deviations are visible. When determining the temperature, these deviations will introduce additional errors, which will limit the accuracy of the measured temperature. The pictures in the lower row show the corresponding mean values when averaging along the principal axes.

Extracting the temperature

With a model for the potential, the next step is to extract the functional dependence of the optical density. For this, the potential is separated into bins of size dB. The potential value and the logarithm of the optical density is then averaged along these equi-potential bins. Fitting a linear function to the low density region, where the shape of the curves becomes linear then should yield the temperature, as expected from equation (5.3). As an example, the resulting curve generated from the cloud in figure 5.14 is shown in figure 5.15. Each thermometry measurement consists of ≈ 30 repetitions. This was repeated for several heating times at different interaction values in order to get a reliable mapping from heating time to temperature.

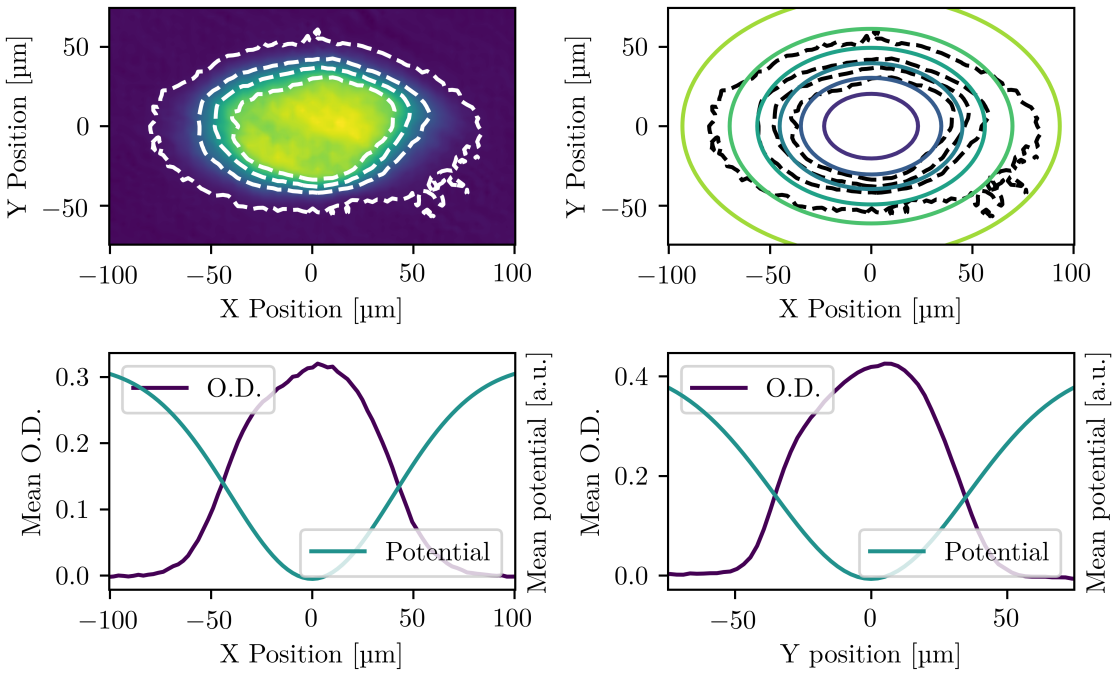

Figure 5.14: Comparison of reconstructed potential and density distribution of the atom cloud. (Upper left) Averaged optical density picture of an atom cloud with equi-density lines (white dashed lines). **(Top right)** Same equi-density lines (black dashed lines) as in the left picture. Colored lines indicate equi-potential lines as determined by the potential reconstruction method. **(Lower row)** Mean of the density distribution and the reconstructed potential along x and y direction. The picture of the cloud is taken at an interaction strength of $1/(k_F a) \approx -0.5$ without additional heating.

5.3.4 Thermometry results

The results for the thermometry are shown in figure 5.16. Each curve represents a different interaction throughout the BEC-BCS crossover, with each datapoint at a given heating time consisting of 30 repetitions that are averaged and analysed as described in the previous section. Since all curves are overlapped to a good degree, we fit a linear function to all datapoints, shown as the dashed line. For long heating times $> 4\,\mathrm{ms}$ the curves seem to flatten. We exclude these datapoints from the fit, since the condensate at all interaction values has vanished by then and we are only interested in temperatures up to the critical temperature. For values larger than $1/(k_F a) = 0.3$ fitting becomes unreliable due to atom loss. The final conversion function from heating time to temperature then becomes

$$T(\tau) = 29.00(199)\,\mu\mathrm{K\,s^{-1}} \cdot \tau + 0.075(5)\,\mu\mathrm{K}. \tag{5.8}$$

Note, that for typical values used in these experiments the initial temperature of $0.075(5)\,\mu\mathrm{K}$ is consistent to the temperature measured by the vanishing condensate on the BCS side from section 5.3.1 of $0.064(18)\,\mu\mathrm{K}$. The slope from equation (5.1) is $\approx 36\,\mu\mathrm{K\,s^{-1}}$,

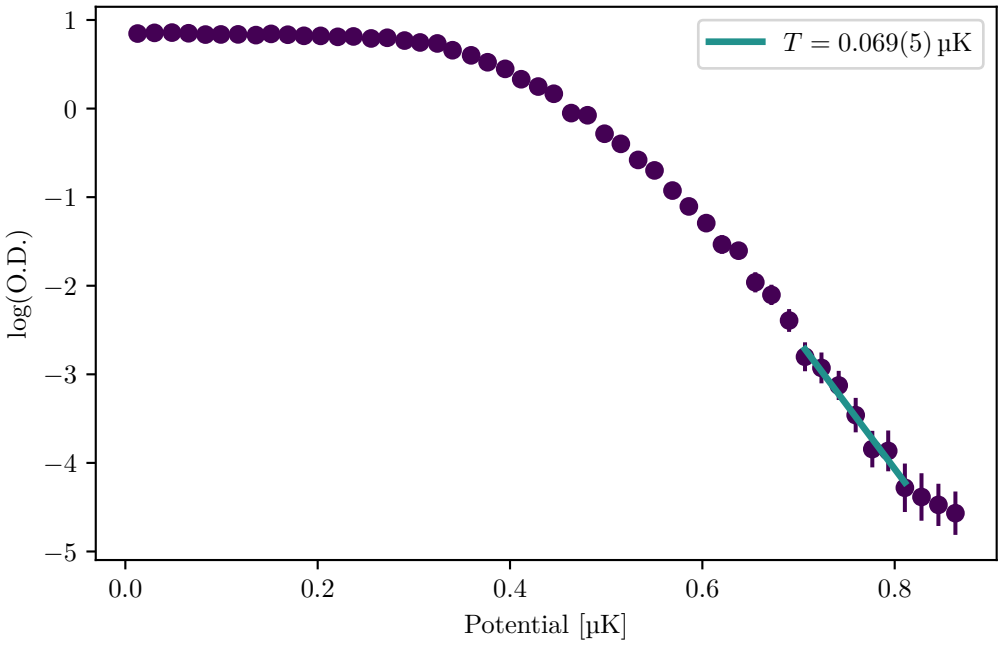

Figure 5.15: Determination of the temperature by a fit to the cloud surface. The purple datapoints show the logarithm of the mean optical density within each bin versus the mean potential value within each bin. The bins are created with respect to the equi-potential lines. The turquoise line shows a linear fit to the cloud surface. From the slope of the curve it is possible to determine the temperature. Error bars are given by the error of the mean within each bin. The underlying cloud is the same as in figure 5.14 at an interaction strength $1/(k_F a) \approx -0.5$ and without additional heating. The last points show a change in slope due to the sampling of noise outside of the cloud.

which is close to the accurate value determined by our thermometry. Note, that equation (5.1) is given in units of T_F which changes throughout the measurements and introduces an error to this estimation.

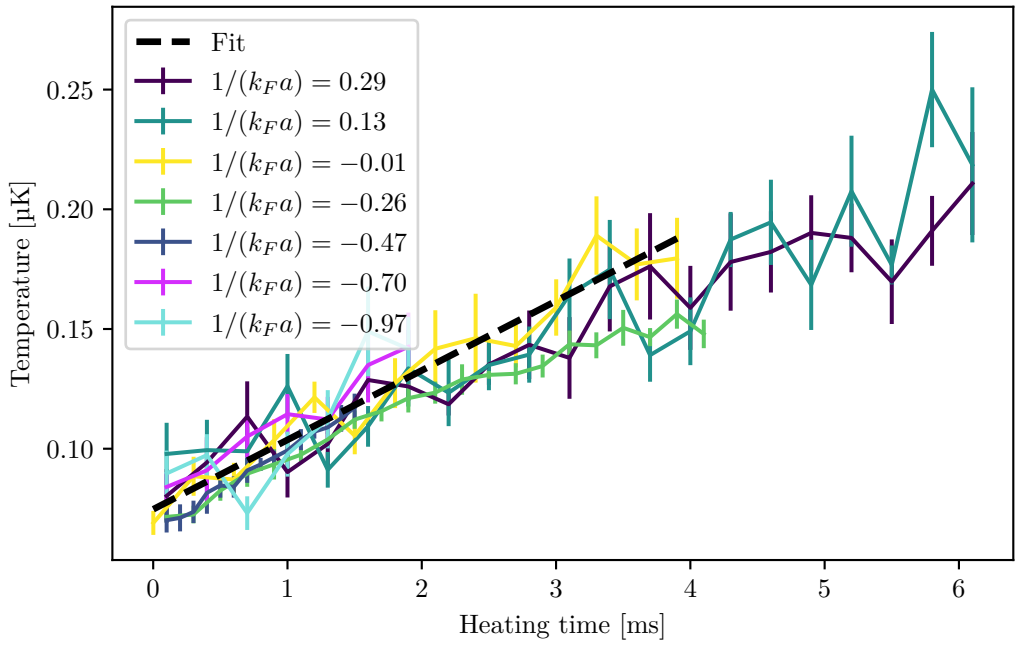

Figure 5.16: Thermometry results. Each colored curve represents one interaction value with each datapoint consisting of 30 separate pictures that are averaged for one temperature measurement. The black dashed line is a fit to all datapoints up to 4 ms with parameters given in equation (5.8), where the condensate has vanished for all interactions.

5.4 Phase diagram of the BEC-BCS crossover

With the conversion from heating time to temperature through the in-trap thermometry and the network predicting the condensate fraction, it is possible to map the BEC-BCS crossover phase diagram. For this, a large dataset of ≈ 3200 pictures was taken within a range of interactions of $-1 \leq 1/(k_Fa) \leq 2.8$. At each interaction, each heating time is taken twice for averaging. The resulting phase diagram is shown in figure 5.17. The absolute temperature T is calculated with the temperature calibration given by equation (5.8). T_F is separately calculated for every datapoint, using the atom number extracted from the time-of-flight pictures, taking into account the Gaussian shape of the potential. The grey dashed lines highlight the interaction values at which the network was trained. The white datapoints mark the temperature value at which the condensate fraction falls below 0.005. This criterion was chosen because for some datapoints far on the BEC side, the network predicts a very small but non-vanishing condensate fraction for some datapoints. For all other interactions, this value is equal to just taking the first datapoint that predicts a condensate fraction of 0. At unitarity we estimate the critical temperature to be $T_C/T_F = 0.166(12)$, which is consistent to the value of $T_C/T_F = 0.167(13)$ determined in ref [22]. We estimate the value of the critical temperature in the BEC limit by taking the mean of all points with $1/(k_Fa) > 1$ to achieve $(T_C/T_F)_{\mathrm{BEC}} \approx 0.229(3)$. In the unitarity regime, we predict

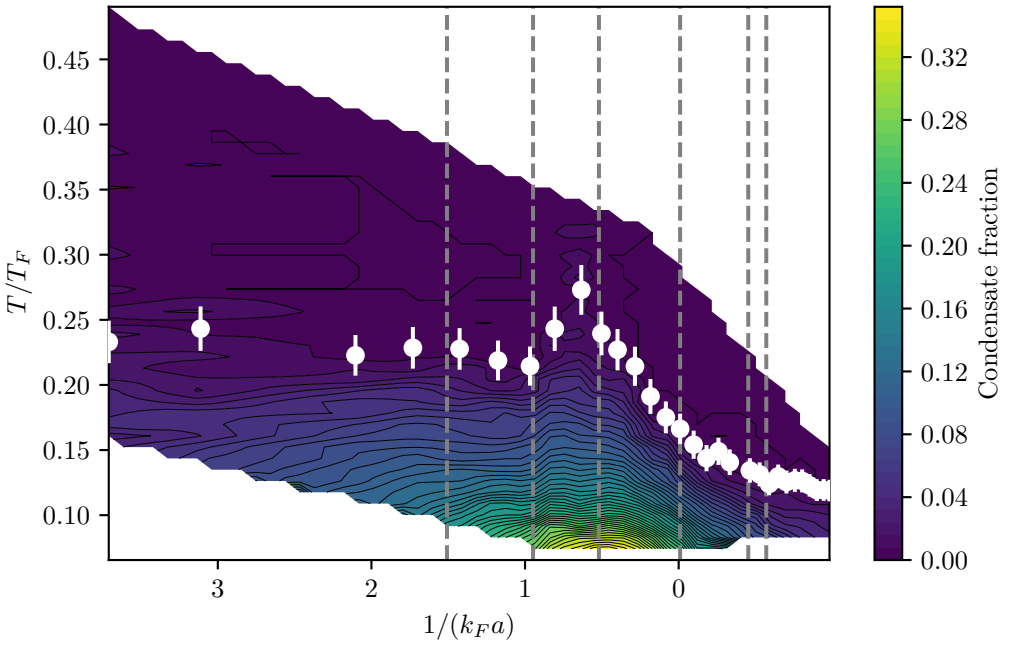

Figure 5.17: Phase diagram of the BEC-BCS crossover. The temperature scale in units of the Fermi temperature is determined by the temperature calibration given in equation (5.8). The increase in initial temperature is due to a decrease in the Fermi temperature caused by atom loss on the BEC side. Grey dashed lines indicate the interaction values where the network was trained. The white datapoints mark the extracted critical temperature.

a maximum in the critical temperature of $(T_C/T_F)_{\max} = 0.27(2)$ around $1/(k_F a) \approx 0.63$ in agreement with what the theoretical approaches in figure 5.2 suggest. Figure 5.18 shows our measured critical temperature in comparison to theory and simulation. Towards the BEC side, our datapoints agree well with the theoretical approaches presented in refs [21, 38, 165–167]. At unitarity and small negative values of $1/(k_F a)$, the measured T_C agrees well with the simulations presented in refs [19, 21, 168], however, for larger negative values our curve deviates from the theoretical datapoints. In the limit of weak interactions, the critical temperature is expected to approach the Gorkov limit [72] given by equation (2.51) which should act as an upper limit for the critical temperature and is approximately a factor of two lower than the mean-field BCS limit. Our datapoints deviate significantly from this behaviour, with higher values for T_C in the regime of $1/(k_F a) < -0.5$. A similar slowing of the decay of the critical temperature was observed in the early phase diagram of ref [18] shown in figure 5.1, however, for these measurements the thermometry can be expected to introduce errors towards stronger interactions. One cause for the slowed decay could be the suppression of pair fluctuations due to the presence of the harmonic trap which can cause an increase in T_C/T_F compared to the Gorkov limit [181]. A possible second reason are effects of the rapid ramp towards high fields, like pair formation during the ramp. Both cases are consistent with the observation of the slowed decay in our experiment and ref [18] as both

measurements use a harmonic trap and the rapid ramp. In our case, however, we use the rapid ramp only on the training data, which aims to decouple the results from the exact shape of the ramp. Alternatively, ref [18] suggests that the observed condensate could arise from condensed atomic pairs above the Feshbach resonance. These pairs would be smaller than the interparticle distance and arise from two-body effects with a lifetime that is potentially affected by the presence of the Fermi sea and mean-field effects. Note that the prediction of a finite condensate fraction from the neural network at $1/(k_F a) \approx -1.0$ is consistent with the rapid ramp observation and not an artifact of the neural network, see figure 5.7. To further investigate, we develop a novel way of detecting the critical temperature without the need for the rapid ramp, based on an autoencoder, see chapter 6. The network architecture used for all phase diagrams is given in appendix A.1.

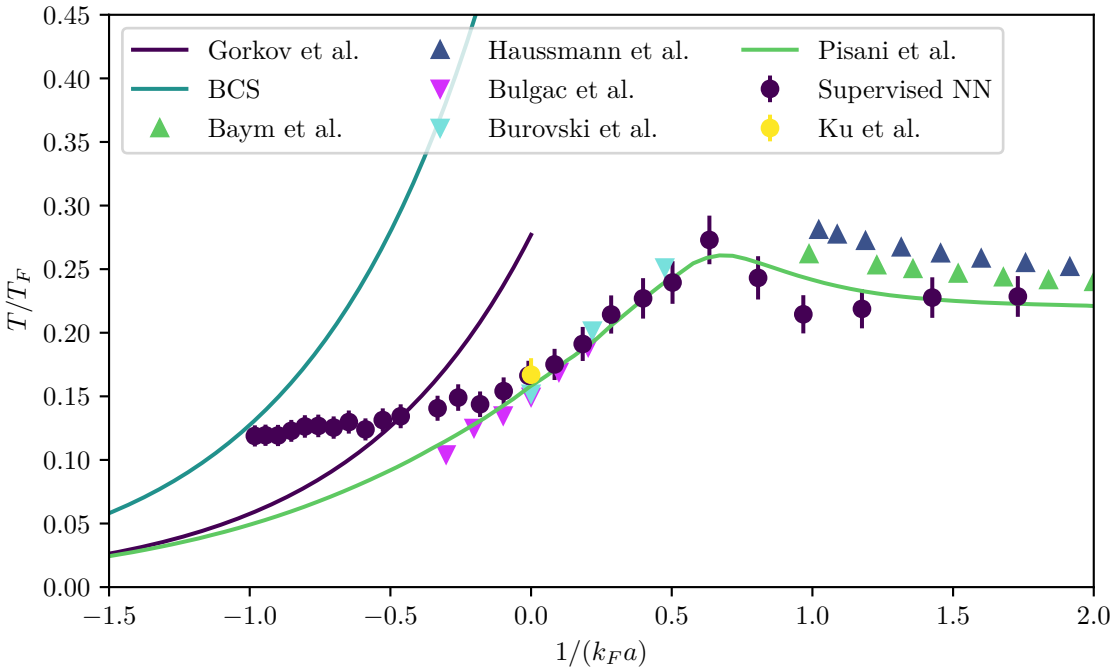

Figure 5.18: Critical temperature across crossover. The data from our supervised neural network approach (purple datapoints) is the same that was shown in figure 5.17 as white datapoints. We show the BCS limit (turquoise line) and the Gorkov approximation [72] (purple line). We compare our results to the theoretical approaches of Haussmann et al. [38], Baym et al. [165–167], Bulgac et al. [19] and Burovski et al. [168], as well as the experimentally determined value from Ku et al. [22].

5.4.1 Difference between BEC and BCS side

Figure 5.17 shows the phase diagram, when trained on the full training data. However, something interesting happens when instead the training data is pre-selected. Using only the two interactions furthest on the BCS side for training yields a phase diagram that performs comparable to the fully trained network in the BCS regime, but predicts a sharp drop of the condensate fraction and critical temperature in the regime $0.5 \leq 1/(k_F a) \leq 1$. The resulting phase diagram is shown in figure 5.19. Something similar happens, when training on the two highest $1/(k_F a)$-values. Again, the network performs as expected in the BEC limit, but shows a sharp drop in a transition regime $0 \leq 1/(k_F a) \leq 0.5$, see figure 5.20. This completes the picture of the neuron activations drawn in section 5.2.2, where the feature the network uses to determine the condensate fraction changes in a similar regime. This enables the network to effectively detect either tightly bound Feshbach molecules or Cooper pairs depending on the data provided during training. From a physics point of view both kinds of pairs are inherently different, but connected through a smooth transition, when approaching unitarity. Note that the quality and smoothness of the phase diagrams may seem worse as a consequence of the reduction in training data. Phase diagrams including training data at interactions closer to the crossover regime can be found in appendix A.2 and do not change the picture significantly. Being able to distinguish between Cooper pairs and Feshbach molecules might help to further explore the unitarity regime and gain an understanding of the dynamics of Cooper pair formation and pair evolution across the Feshbach resonance. The presented neural network approach therefore not only constitutes a clean way of measuring the critical temperature and condensate fraction with high accuracy, but also opens the way to new measurements that were not possible before in a regime that is difficult to access theoretically.

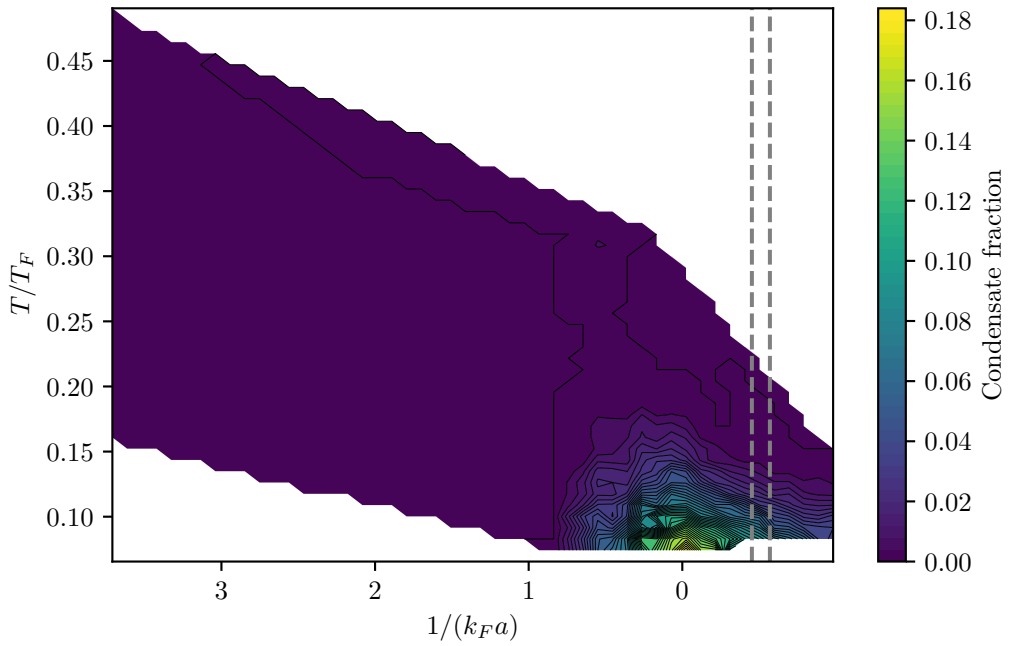

Figure 5.19: Phase diagram with training data on the BCS side. Only data at $1/(k_F a) \approx -0.63$ and $1/(k_F a) \approx -0.5$ (grey dashed lines) was used to train the network. Note, the overall quality might change due to the limited data taken for training. A phase diagram including $1/(k_F a) \approx -0.01$ can be found in appendix A.2.

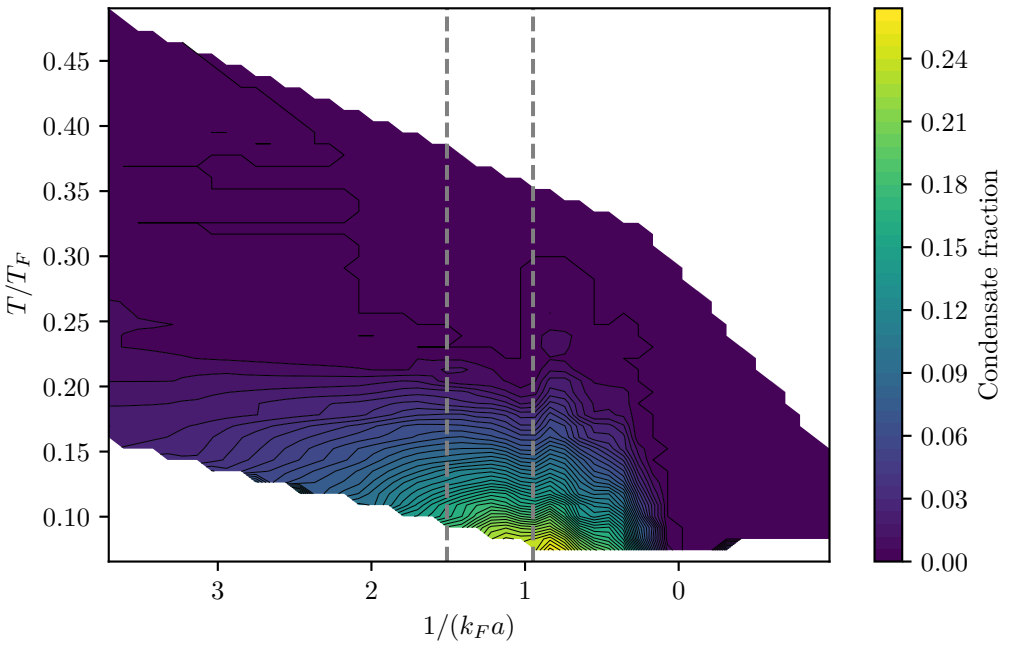

Figure 5.20: Phase diagram with training data on the BEC side. Only data at $1/(k_F a) \approx 1.04$ and $1/(k_F a) \approx 1.66$ (grey dashed lines) was used to train the network. Note, the overall quality might change due to the limited data taken for training. A phase diagram including $1/(k_F a) \approx 0.57$ can be found in appendix A.2.

CHAPTER 6

Unsupervised Detection of the critical Temperature

In order to use supervised machine learning techniques as introduced in the previous section it is required to have labelled data to train on. In many cases, however, it is not straightforward how to label data. In these cases, unsupervised learning techniques might provide a way of structuring data. Here, we use an autoencoder to achieve a low-dimensional representation of the phase diagram data. Analysing the distribution of datapoints, we extract meaningful information from the experimental parameters of the datapoints. Carefully analysing the distribution of datapoints in the representation allows us to extract the critical temperature. Importantly, using only the phase diagram data foregoes the use of the rapid ramp technique, which provides an important crosscheck on the T_C measurement presented in the previous section.

6.1 A low-dimensional representation for the phase diagram data

As introduced in section 3.5.2, autoencoders can be used to find efficient representations of complex data. The general idea is to use experimental data - in this case the time of flight pictures from the phase diagram data - as input and output of a network, that has an information bottleneck at the center. The network is then trained to reproduce the input picture as the output[1]. Due to the information bottleneck in the network it is not possible to transfer all information from input to output layer. In order to minimise the training error, the network has to find an efficient low-dimensional representation to yield an output that captures all essential information. By using only the *encoder* part of the network and the bottleneck as output layer, it is possible to access this representation.

Inspired by the results in section 5.2, we use an autoencoder network that roughly resembles the network architecture from before. This means a number of convolutional plus pooling layers followed by fully connected layers up to the bottleneck (details of the network architecture are given in appendix A.1). The bottleneck consists of two neurons, yielding a two dimensional representation of the data. Training the network and putting the neuron activations on two

[1] To be exact, the input is used as a label to apply supervised learning procedures, however, since this does not require extra labels, this technique is usually referred to as unsupervised.

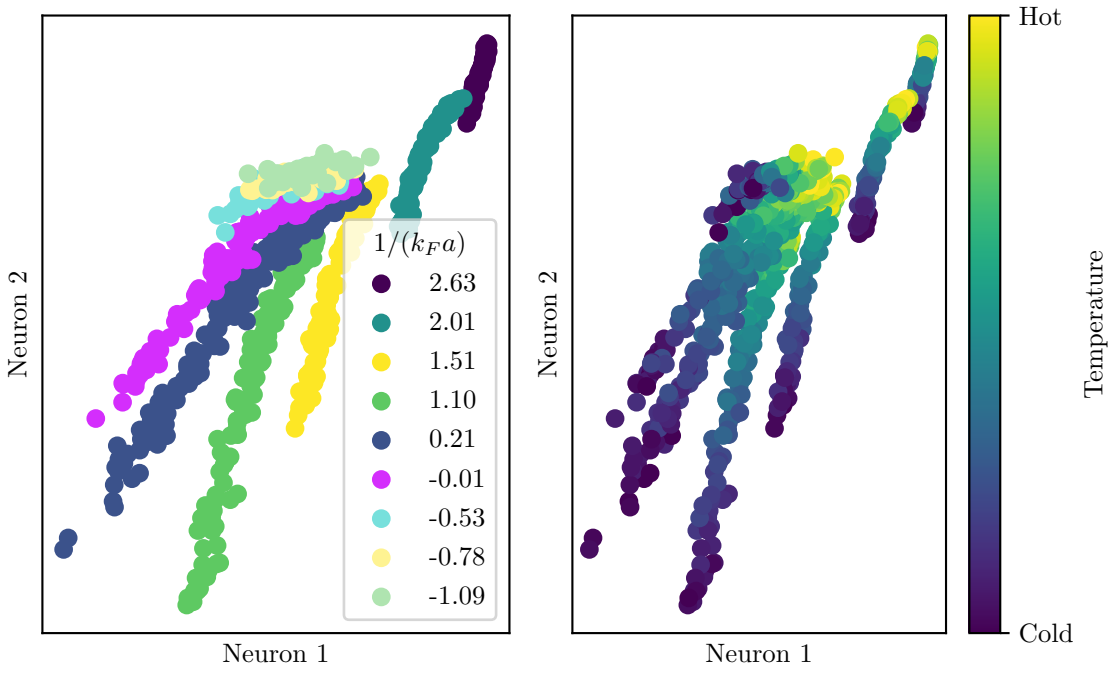

Figure 6.1: Dimensional reduction with the autoencoder technique for selected values of the magnetic field across the BEC-BCS crossover. Both plots show the same distribution. The coloring in the left plot indicates the interaction strength of the corresponding datapoints and the coloring in the right plot is proportional to the temperature of the gas.

axes yields figure 6.1. The left plot shows, that the network can distinguish between different interactions in the crossover. Interestingly, towards the BEC side, the representation seems to change significantly as compared to the BCS side and unitarity, as the total distance between the point clouds increases. The panel on the right side shows the same plot as on the left with the color code representing the temperature of the corresponding datapoint. The lines corresponding to the interactions are ordered according to the temperature, showing the representation also seems to take temperature into account. Interestingly, most of the lines seem to meet for high temperature, suggesting the network learns that a hot gas is more *similar* than a cold gas for different interaction values.

6.2 Similarity between interactions

One strength of the autoencoder representation is, that it provides a way of *hashing* the data [182]. This means that more similar datapoints will be closer in the projected space than less similar datapoints. To analyse the similarity between the datapoints at different interactions, we fit a linear function to each line representing one interaction value. Since many of tthe lines intersect at one point, we choose the angle of the line as a measure of similarity. The angle of each line versus the interaction is shown in figure 6.2. From unitarity towards the BCS side, the angle tends towards a fixed value with functional dependence that seems to be

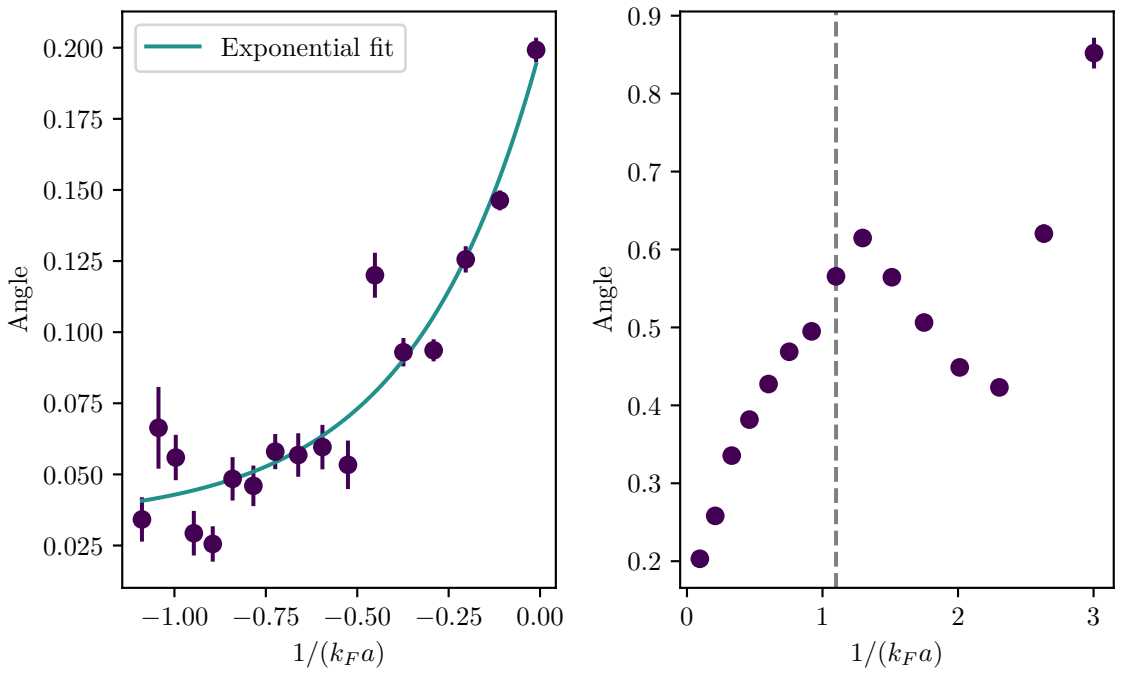

Figure 6.2: Similarity of datasets of different interaction strengths within the low dimensional representation. Dataset similarity is characterised by the angle of a linear fit to the different magnetic fields. **(Left)** Towards the BCS side of the resonance, the increase in similarity is characterised well by an exponential function. **(Right)** Far on the BEC side the similarity measure breaks down due to the dataset changing the position in addition to the angle.

well described by an exponential function. This is consistent with the physics of the BCS Hamiltonian that describes the system through the gap parameter $\Delta \propto e^{-\pi/(2k_F a)}$. Towards the BEC side, the trend continues and seems to slow down towards $1/(k_F a) \approx 1$ consistent with the BEC limit. However, since the position of the lines in the projected space changes, it is questionable, whether the angle still makes sense as a similarity measure towards the far BEC side. The dashed line in the figure marks the point, where a visible spatial shift of the line takes place and the similarity description breaks down.

6.3 Extracting the critical point

Another strength of the representation generated by the encoder network is that sudden changes within the data can be detected by observing the shape of the clusters in representation space. Upon closer inspection of the "interaction lines" in figure 6.1 for example, a small kink along the line becomes visible. The left plot of 6.3 shows three example interactions. This kink is not visible for all interactions, but well pronounced in the region $-0.5 \leq 1/(k_F a) \leq 0.4$. Keeping in mind that the kink might suggest some kind of critical point, where the representation of the network changes due to a change in the system, it is possible to analyse the position of the kink with respect to experimental quantities. The position of the kink

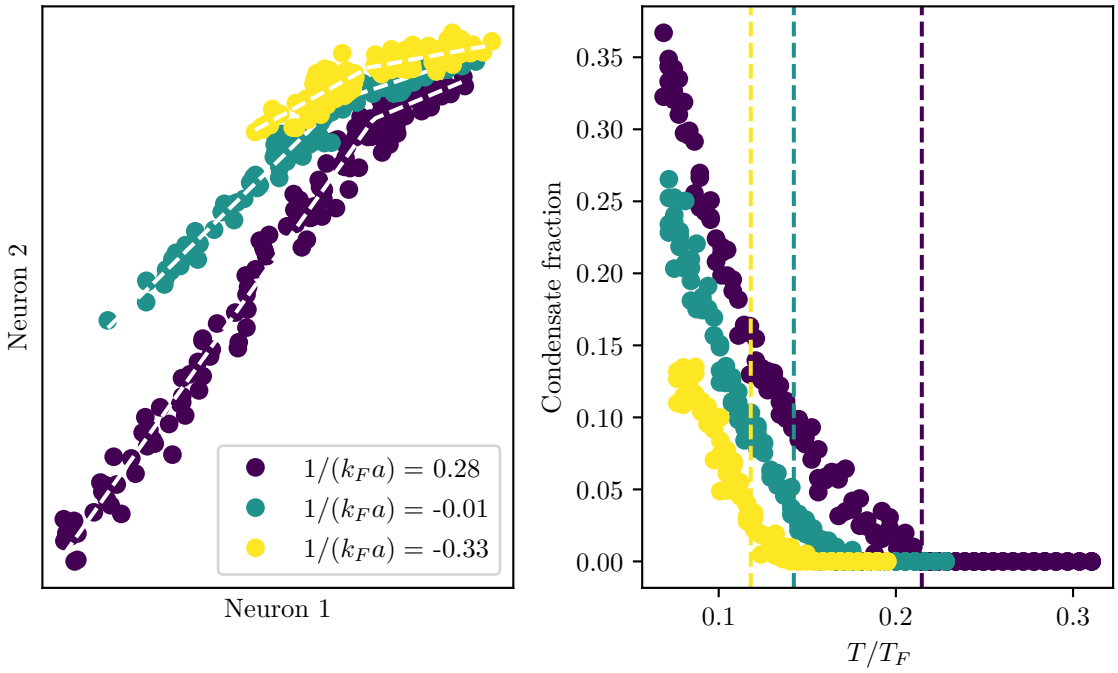

Figure 6.3: Extracting the critical point from the autoencoder representation. The plot on the left side shows example representations for different interactions obtained from the encoder network. The white dashed lines show piecewise linear fits to extract the position of the kink. The right plot shows the condensate fractions as predicted by the supervised neural network from section 5.2. The dashed lines show the corresponding temperature calculated from the position of the kink in the left plot using in-trap thermometry.

is extracted by piecewise linear fits within parameter space (white dashed lines). As the temperature is ordered along the interaction lines, the temperature corresponding to the kink position can be extracted by averaging the temperature of the datapoints in a small region around the kink position. The right panel in figure 6.3 shows the extracted temperature corresponding to the kink (dashed lines) as compared to the condensate fraction curves at the corresponding interactions predicted by the neural network in section 5.2. It is visible, that the kink position is close to the critical point for all shown interactions with the tendency to predict lower values for the critical point towards unitarity and the BCS side. It should be stressed that the critical point is not sharp, as any forming condensate around the critical point is a small signature within the data. Therefore, the kink is more likely to be a smooth curve, introducing an intrinsic fitting error. The extracted critical temperature in units of T_F for different interactions is shown in figure 6.4, together with the data from figure 5.18. We extract a critical temperature of $T_C/T_F = 0.142(14)$ at unitarity, compatible with our previous measurement of $T_C/T_F = 0.166(12)$ and the experimental value of $T_C/T_F = 0.167(13)$ from ref [22]. The values extracted from the low-dimensional representation agree well with the theoretical predictions of Bulgac et al. [19], Burovski et al. [168] and Pisani et al. [21]. In comparison to the supervised datapoints, the autoencoder

does not seem to predict a deviation from the theory. However, the range of the autoencoder analysis only extends to a region where the deviation is small, which does not allow for a definitive conclusion in this regard. Around unitarity and on the BEC side, the predictions of the neural networks and the theory agree within errorbars.

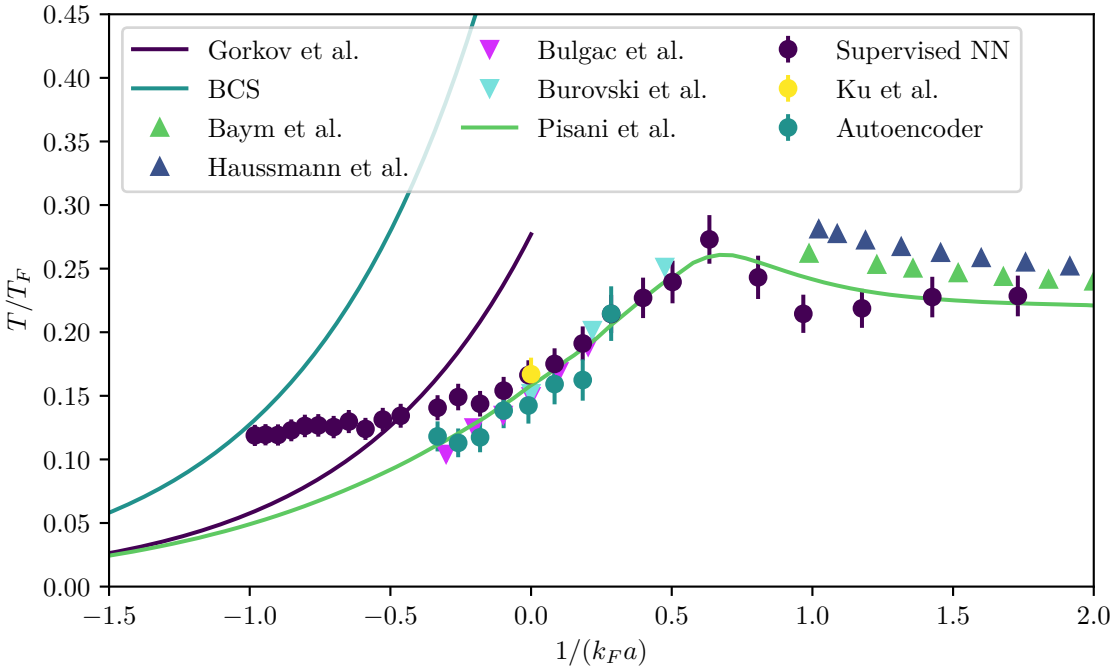

Figure 6.4: Critical temperature across the crossover with autoencoder data. This figure shows the same data as figure 5.18 with additional datapoints from the autoencoder approach (turquoise dots). The autoencoder data agrees well with the experimental and theoretical approaches in the range of investigated interactions, but does not allow for a conclusion regarding the slow decay towards the BCS side.

6.4 Improvements

The autoencoder used for the presented analysis is trained with gradient based backpropagation. For deep autoencoders, however, this can cause problems in the training process, as errors are damped when backpropagating through the network. Using deeper autoencoders might provide benefits, such as an exponential reduction of the computational cost representing certain kinds of functions, a (possible) exponential decrease in the amount of training data to learn certain functions and a better compression of the representation of the data as compared to shallow encoders [110, 123]. Training deeper autoencoders efficiently might require advanced training methods [123], but might also yield more expressive representations.

6.5 Summary

Using an autoencoder with two neurons in the bottleneck, we found a two-dimensional representation of the phase diagram data. Analysis of the distribution of datapoints within this representation showed that an encoder network based on convolutional and fully connected layers is able to discover the concept of temperature and interaction from just time-of-flight pictures. Using the angle of linear fits to points of equal interaction as a similarity measure, we showed that the exponential trend of similarity towards the BCS side of the resonance was recovered. Analysing the curves of equal interactions in more detail, we showed that a critical point represented through a kink in the datapoint distribution coincided with the critical point of the onset of condensation to a reasonable degree. It should be stressed again here, that the autoencoder only learns from the data, it does not have any concept of condensate, temperature or interaction beforehand. The results show that unsupervised learning methods based on autoencoders are able to discover physical concepts from first principles and identify critical points and phase transitions even on unclean experimental data (as compared to numerical data). We compared the extracted critical temperatures to the results from our supervised neural network approach, the experimental reference from ref [22] and different theoretical approaches and found good agreement.

CHAPTER 7

Higgs Mode in a strongly interacting Fermionic Superfluid

This chapter describes the observation of the Higgs mode in a strongly interacting fermionic superfluid. The Higgs mode, a collective mode in the superfluid, is excited by modulation of the gap parameter via radio-frequency dressing. This work has been published in ref [1]. Theoretical modeling has been done by Johannes Kombe and Jean-Sébastien Bernier in the group of Prof. Dr. Corinna Kollath at the University of Bonn [183]. The data shown here was also discussed in the theses of Tim Harrison and Alexandra Behrle [129, 130]. My contribution to this work was taking data, calibrating the system and interpretation of the results.

7.1 The Higgs mode

When the symmetry of a physical system is spontaneously broken, new collective modes can appear. In the case of a BCS superconductor as introduced in chapter 2, such a symmetry is broken when transitioning through the critical point of the thermal to superfluid phase transition. This phase transition is described by the superconducting order parameter Δ. In general, this is a complex number and can be written in terms of phase and amplitude $\Delta = |\Delta|e^{i\Phi}$. In the limit of s-wave scattering, however, the phase can be gauged out and does not play a role, which allows us to use Δ and $|\Delta|$ interchangeably. Consequently, the amplitude of the order parameter is equal to the superfluid gap energy. When the amplitude of the complex order parameter becomes finite when transitioning the phase border from a thermal gas to a superfluid, the convex energy landscape deforms and takes the form of a *mexican hat* potential, shown in figure 7.1. In this case the phase symmetry of the system breaks down and it chooses an arbitrary phase corresponding to a position in the valley along the hat. This situation exhibits two distinct modes: For one the massless Nambu-Goldstone-, or Phasemode, which corresponds to movement along the hat valley. The perpendicular mode in radial direction is called (Anderson-)Higgs- or amplitude mode and corresponds to an oscillation of the amplitude of Δ. The stability of the Higgs mode against decay into lower lying modes such as the Goldstone mode has to be secured by the presence of additional symmetries in the system [42]. It manifests in oscillations of the absolute value of

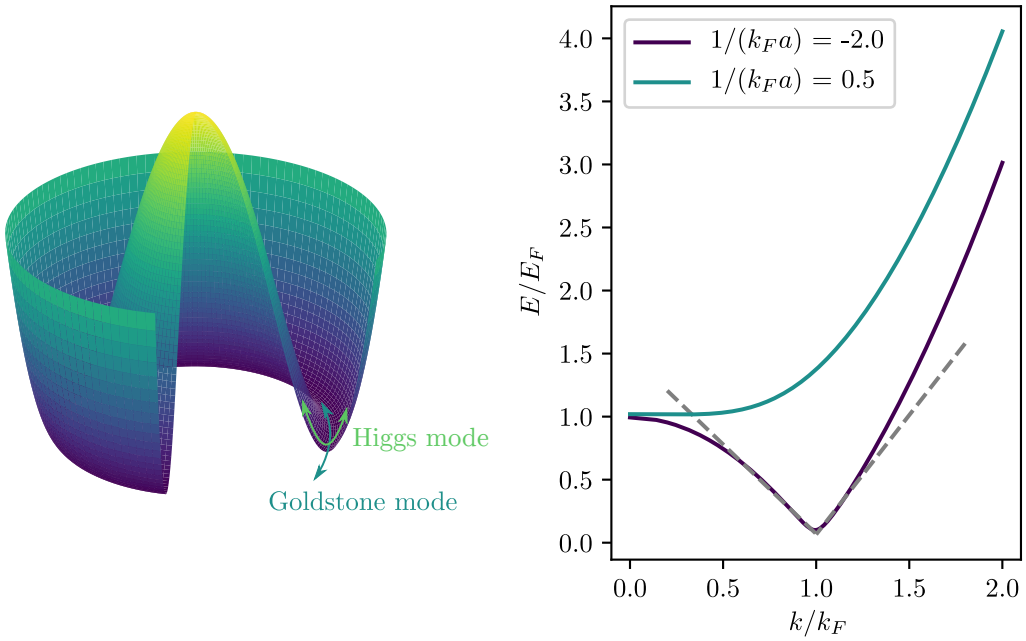

Figure 7.1: Mexican hat and particle hole symmetry. (Left) The mexican hat potential shows two distinctive modes: The Nambu-Goldstone mode which moves along the rim of the hat and the perpendicular, radial Anderson-Higgs mode. Stability of this mode is guaranteed with particle hole symmetry. **(Right)** This plot shows the dispersion of a Fermi gas in the BCS regime (purple) and the strongly interacting BEC regime (turquoise). On the BCS side, particle hole symmetry is granted around the Fermi surface because the energy values are the same for a hole (slightly below the Fermi surface at $k/k_F = 1$) and for a particle slightly above the Fermi surface. This can be imagined by two lines having the same but opposite slope around the Fermi surface (grey dashed lines). This requirement is lifted gradually when transitioning to the BEC regime, which allows decay into lower lying modes.

the order parameter with a frequency of $\nu_H = 2\Delta/h$. The most famous occurrence of the Higgs mode is in the standard model of particle physics [184–186], where it is responsible for the mass term of elementary particles. Stability of the Higgs mode against decay into lower lying modes in this case is guaranteed by Lorentz invariance. In the non-relativistic, low energy regime of condensed-matter physics, Lorentz invariance does not play a role and the existence of a stable Higgs mode cannot be taken for granted [42]. The role of the additional symmetry under certain conditions can be played by particle-hole symmetry. One low-energy system with approximate particle-hole symmetry is the BCS model introduced in section 2.3.2. Particle hole symmetry is realised close to the Fermi surface, where particles and holes exhibit the same energy after excitation which can be seen by equal energy values slightly above and below the Fermi surface, see figure 7.1. Indeed, evidence for the Higgs mode has been found in conventional BCS superconductors [32–34]. However, all these measurements rely on indirect measurements of the Higgs mode either through coupling to charge-density waves or non-linear terahertz spectroscopy, as the Higgs mode does not linearly couple to the

electromagnetic field [187]. In this chapter, we present a direct parametric excitation scheme of the Higgs mode. With this, we are not only able to measure the associated frequency which we can compare to the predicted and measured gap values of other works, but we can also study the behaviour of the Higgs mode towards unitarity, where the particle hole symmetry breaks down partially. It is unclear whether in this regime a stable Higgs mode exists, or whether it becomes overdamped [42, 188–192]. Other systems in which the Higgs mode was observed are quantum ferromagnets [193], superfluid ^{3}He [194], ultracold atoms in a lattice close to the superfluid/Mott-insulator transition [195, 196], spinor Bose gases [197] and Bose gases strongly coupled to electrical fields [198].

7.2 Modulating the order parameter

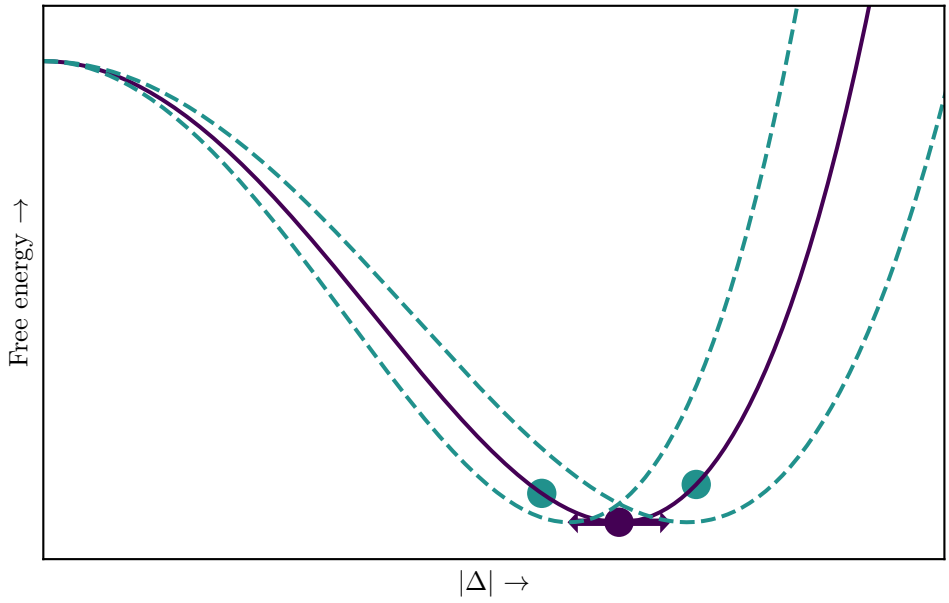

Figure 7.2: Excitation by modulation. One way to excite the Higgs mode is to modulate the absolute value of the order parameter with the correct frequency $\nu_H = 2\Delta/h$. This can be illustrated by imagining a *wiggling* of the hat.

One way to excite the Higgs mode is by modulation of the absolute value of the order parameter $|\Delta|$ [189], see figure 7.2 for illustration. The most straightforward way to do so is to modulate the interactions between the atoms by modulation of the magnetic field, however, only single particle excitations have been detected with this scheme so far [199]. Here, we use a different scheme for excitation based on radio-frequency (rf) dressing of the states: We prepare a $|1\rangle + |2\rangle$ condensate at a given magnetic field B. We then rf dress the $|2\rangle$-state by applying off-resonant, red detuned rf radiation addressing the $|2\rangle \rightarrow |3\rangle$ transition. This

drives Rabi oscillations between states $|2\rangle$ and $|3\rangle$, effectively modulating the populations N_2 and N_3. For an isolated two-level system, the time evolution of the population in the $|2\rangle$ state is given by

$$p_{|2\rangle} = 1 - \frac{\Omega_R^2}{\Omega_R'^2} \sin^2\left(\frac{1}{2}\Omega_R' t\right), \tag{7.1}$$

with Ω_R being the on-resonance Rabi frequency and $\Omega_R' = \sqrt{\Omega_R^2 + \delta^2}$ being the effective Rabi frequency and δ the detuning from resonance. This situation is more complicated in the real system, as the presence of the dispersion in the system effectively introduces a momentum dependency to the detuning $\hbar\delta \to \hbar\delta_k = \hbar\delta - E_k - \epsilon_k$ with $\epsilon_k = \hbar^2 k^2/(2m) - \mu$ and $E_k = \sqrt{\epsilon_k^2 + \Delta^2}$, see figure 7.3. In order to prevent single particle excitations in the sample, we choose a red detuned rf drive to modulate the population off-resonantly, which will lead to a modulation of $|\Delta|$ as we will see in the next section.

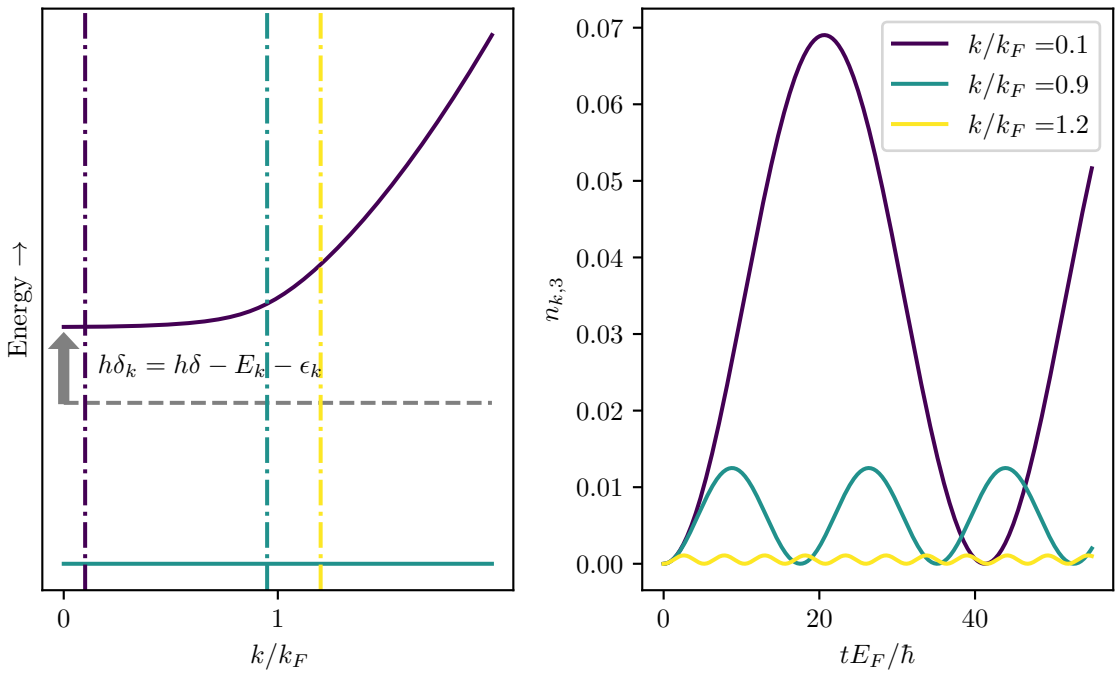

Figure 7.3: Momentum dependency of the detuning at $1/(k_F a) = -0.3$. (Left) We apply rf radiation with detuning δ with respect to the bare transition frequency between ground (turquoise) and excited state (purple). Due to the dispersion of the excited state, the effective detuning is momentum dependent. The dashed-dotted lines are color coded to correspond to the different momenta shown on the right side. **(Right)** Rabi oscillations for different momenta for $\Omega_R = 0.04 E_F$ with a bare detuning of $\delta = -0.3 E_f$. For red detuning, high momenta will experience a higher driving frequency due to this effect.

7.2.1 Theoretical modeling

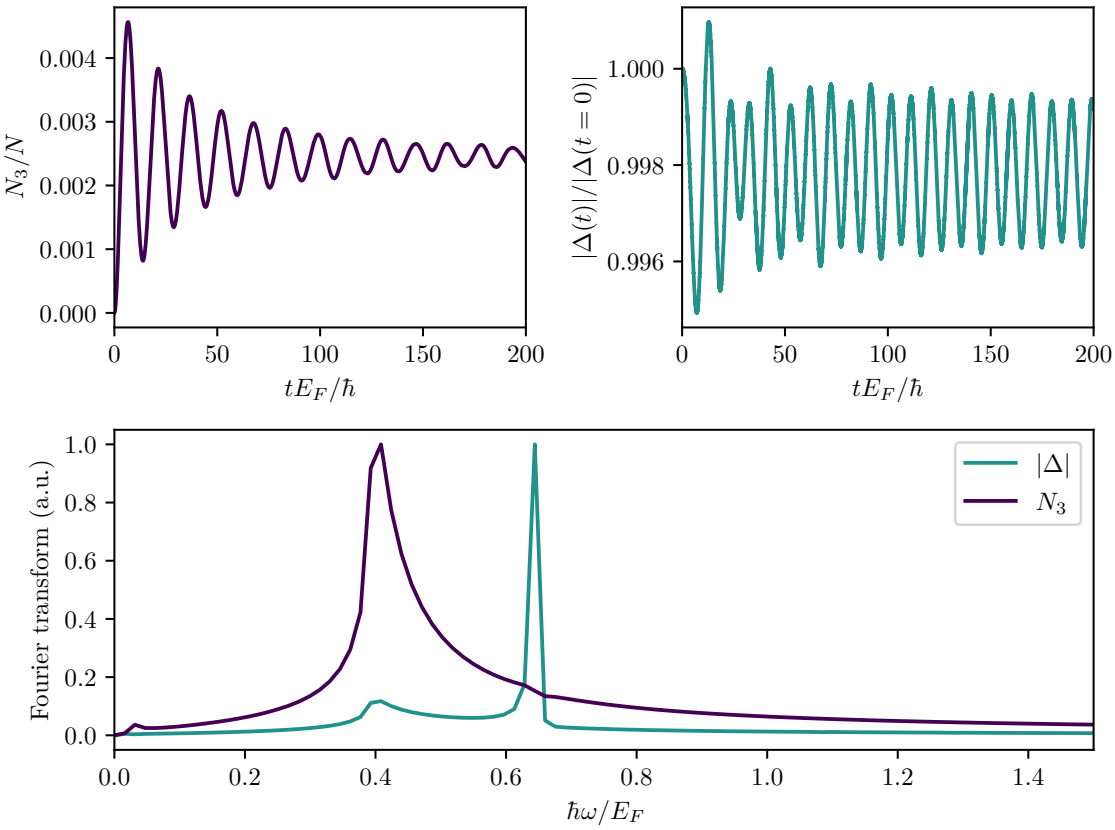

Figure 7.4: Simulated response to the rf drive. Numerical simulations at $1/(k_F a) = -0.6704$, $\hbar\Omega_R = 0.0353 E_F$ and $\hbar\delta = -0.3247 E_F$, resulting in $\alpha \approx 0.4\,\%$. **(Top left)** The simulated rf drive causes a modulation in the population of state $|3\rangle$. The decaying amplitude of the modulation is caused by dephasing effects within the trap. **(Top right)** The rf drive causes a persistent modulation in the absolute value of the superfluid gap $|\Delta|$, allowing for spectroscopic measurements of the Higgs mode. **(Bottom)** Fourier transform of the population (purple) and absolute value of the gap (turquoise). Although the drive modulates the population at a lower frequency, the strongest response of the gap happens at a frequency corresponding to the predicted Higgs mode frequency.

The important question that remains to be answered is whether the modulation of the population in the states is able to cause a modulation effect in the order parameter. To investigate the effect of the modulation scheme, we simulate the rf drive. For this, we combine the mean field Hamiltonian for states $|1\rangle$ and $|2\rangle$

$$H_{\mathrm{BCS}} = \sum_{\mathbf{k}} \epsilon_{\mathbf{k}} \left(n_{\mathbf{k},1} + n_{\mathbf{k},2} \right) + \sum_{\mathbf{k}} \left(\Delta^{\star} c_{-\mathbf{k},2} c_{\mathbf{k},1} + \Delta c^{\dagger}_{\mathbf{k},1} c^{\dagger}_{-\mathbf{k},2} \right), \tag{7.2}$$

with $\Delta = \frac{g}{V} \sum_{\mathbf{k}} \langle c_{-\mathbf{k},2} c_{\mathbf{k},1} \rangle$ being the mean field gap with volume V and interaction

strength g, with a term H_{rf} describing the rf drive and dispersion of state $|3\rangle$. We treat interactions between states $|1\rangle$ and $|2\rangle$ on the mean-field level and decouple them within the s-wave Cooper pair channel. State $|3\rangle$ is treated as a free band and coupled to state $|2\rangle$ through H_{rf}. Within the rotating wave approximation H_{rf} becomes

$$H_{\mathrm{rf}} = \sum_{\mathbf{k}} \left(\epsilon_{\mathbf{k}} - \hbar\delta\right) n_{\mathbf{k},3} + \frac{\hbar\Omega_R}{2} \sum_{\mathbf{k}} \left(c^{\dagger}_{\mathbf{k},3} c_{\mathbf{k},2} + c^{\dagger}_{\mathbf{k},2} c_{\mathbf{k},3} \right), \tag{7.3}$$

where δ is the detuning with respect to the resonance frequency. Combining this to the total Hamiltonian

$$H = H_{\mathrm{BCS}} + H_{\mathrm{rf}}, \tag{7.4}$$

gives a closed set of equations for the expectation values

$$\hbar\frac{\partial}{\partial t} \langle c_{-\mathbf{k},2} c_{\mathbf{k},1} \rangle = i \left\{ -2\epsilon_{\mathbf{k}} \langle c_{-\mathbf{k},2} c_{\mathbf{k},1} \rangle - \frac{\hbar\Omega_R}{2} \langle c_{-\mathbf{k},3} c_{\mathbf{k},1} \rangle + \Delta \left(n_{\mathbf{k},1} + n_{-\mathbf{k},2} - 1 \right) \right\} \tag{7.5}$$

$$\hbar\frac{\partial}{\partial t} \langle c_{-\mathbf{k},3} c_{\mathbf{k},1} \rangle = i \left\{ -\frac{\hbar\Omega_R}{2} \langle c_{-\mathbf{k},2} c_{\mathbf{k},1} \rangle - \left(2\epsilon_{\mathbf{k}} - \hbar\delta\right) \langle c_{-\mathbf{k},3} c_{\mathbf{k},1} \rangle + \Delta \langle c^{\dagger}_{-\mathbf{k},2} c_{-\mathbf{k},3} \rangle \right\} \tag{7.6}$$

$$\hbar\frac{\partial}{\partial t} \langle c^{\dagger}_{-\mathbf{k},2} c_{-\mathbf{k},3} \rangle = i \left\{ \Delta^{\star} \langle c_{-\mathbf{k},3} c_{\mathbf{k},1} \rangle + \hbar\delta \langle c^{\dagger}_{-\mathbf{k},2} c_{-\mathbf{k},3} \rangle - \frac{\hbar\Omega_R}{2} \left(n_{-\mathbf{k},2} - n_{-\mathbf{k},3} \right) \right\} \tag{7.7}$$

$$\hbar\frac{\partial}{\partial t} n_{\mathbf{k},1} = -2 \,\mathrm{Im} \left(\Delta^{\star} \langle c_{-\mathbf{k},2} c_{\mathbf{k},1} \rangle \right) \tag{7.8}$$

$$\hbar\frac{\partial}{\partial t} n_{-\mathbf{k},2} = -2 \,\mathrm{Im} \left(\Delta^{\star} \langle c_{-\mathbf{k},2} c_{\mathbf{k},1} \rangle \right) + \hbar\Omega_R \,\mathrm{Im} \left(\langle c^{\dagger}_{-\mathbf{k},2} c_{-\mathbf{k},3} \rangle \right) \tag{7.9}$$

$$\hbar\frac{\partial}{\partial t} n_{-\mathbf{k},3} = -\hbar\Omega_R \,\mathrm{Im} \left(\langle c^{\dagger}_{-\mathbf{k},2} c_{-\mathbf{k},3} \rangle \right). \tag{7.10}$$

Here, the number densities are $n_{\mathbf{k},m} = \langle c^{\dagger}_{\mathbf{k},m} c_{\mathbf{k},m} \rangle$. From equation (7.5), we can already see that the time evolution of the gap is proportional to the number density of state $|2\rangle$, $\partial/\partial t \Delta \propto \partial/\partial t \langle c_{-\mathbf{k},2} c_{\mathbf{k},1} \rangle \propto \Delta n_{-\mathbf{k},2}$. We solve these equations numerically with discretized time and momentum and using the self-consistency condition $\Delta = \frac{g}{V} \sum_{\mathbf{k}} \langle c_{-\mathbf{k},2} c_{\mathbf{k},1} \rangle$ at each time step, ensuring both the convergence of the time step dt and momentum spacing. Typical values taken are $dk/k_F = 5 \times 10^{-4}$, $dt = 5 \times 10^{-4}\, \hbar/E_F$ and the cutoff for the momentum sum is $E_c = 100\, E_F$. The momentum resolved spectral weight of the gap can be computed as

$$A_{\mathbf{k}}(\omega) = \frac{V}{g} \left| \mathfrak{F} \left\{ |\Delta_{\mathbf{k}}(t)| - \frac{1}{T} \int_0^T dt \, |\Delta_{\mathbf{k}}(t)| \right\} \right|. \tag{7.11}$$

The oscillations in the population of state $|3\rangle$ and the resulting oscillations in the order parameter are shown in figure 7.4 together with the respective Fourier transformed spectra.

We can also see that, although the drive is detuned with respect to the Higgs mode frequency, the oscillations in $|\Delta|$ remain at the Higgs frequency $\omega_H = 2\Delta/\hbar$. Figure 7.5 shows the momentum resolved spectra for a given drive. Here, we can identify three main contributions: For one, the rf drive (yellow crosses), that bends upwards for high momenta as a consequence of the momentum dependent detuning explained in figure 7.3. Secondly, the excitation of quasiparticles (green triangles) that follows the excitation spectrum of BCS quasiparticles with a minimum approaching k_F for weak attractive interactions. Lastly, a strong contribution at $\hbar\omega = 2|\Delta|$ resulting from the Higgs mode (purple stars). As a consequence of the collective nature of the Higgs mode, the energy is the same throughout all momenta.

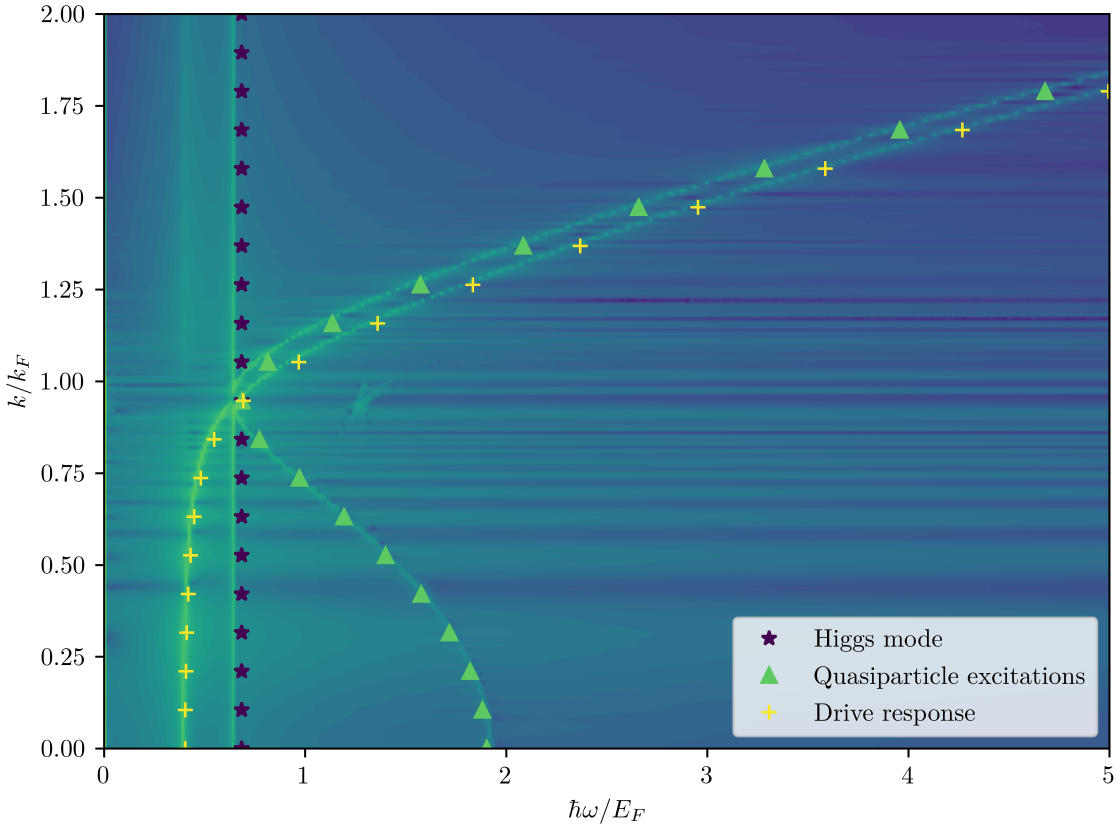

Figure 7.5: Momentum resolved spectra. Spectral weight of the momentum resolved gap $A_k(\omega)$ on a logarithmic color scale at $1/(k_F a) = -0.63$. It is possible to distinguish between three different excitation mechanisms. The Higgs mode (purple stars), as a collective mode does not change its resonance frequency for any value of k. The response to the drive (yellow crosses) bends upwards for high k, as expected from the momentum dependent detuning shown in figure 7.3. The quasiparticle excitation at $2E_k$ (green triangles) follows the expected dispersion relation. Note, that the small discrepancy between the symbols and the curves can be explained by the use of the mean field solution for chemical potential and gap to calculate the corresponding datapoints.

7.3 Experimental implementation and calibration

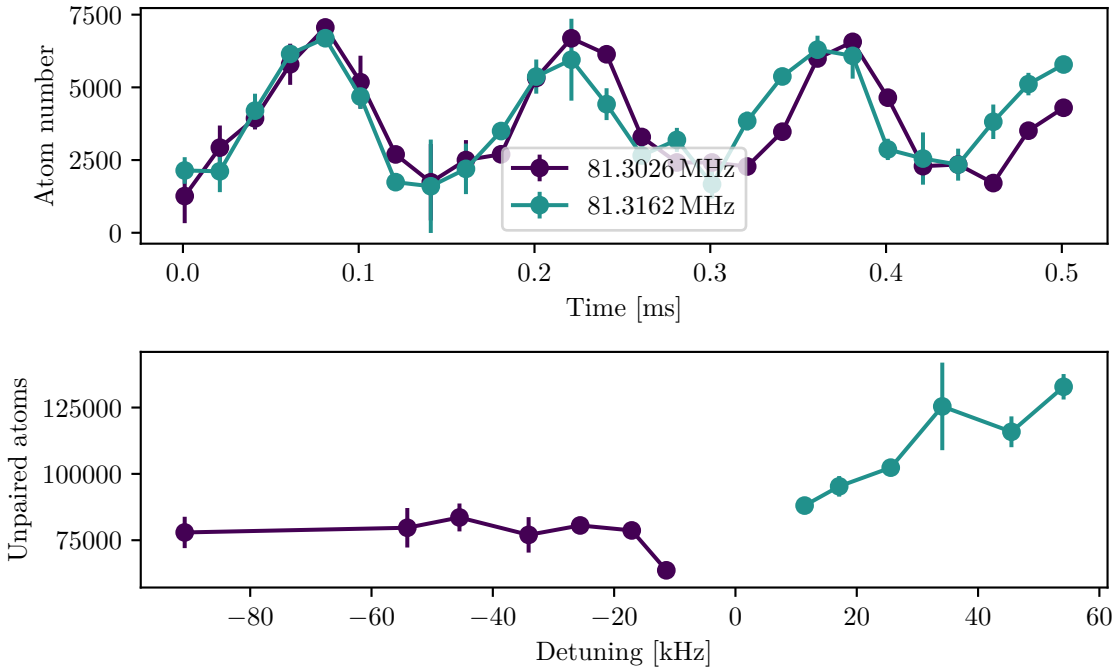

Figure 7.6: Rf drive calibration. (Top) Atom number in state $|3\rangle$ versus time during the rf modulation. Shown are two curves on either side of the resonance. **(Bottom)** We observe an increase in unpaired atoms with a blue detuned rf drive as compared to the red detuned rf drive. This is due to single-particle excitations into the continuum caused by the blue detuned rf. Unpaired atoms have been measured on the BEC side, where the binding energy of the tightly bound molecules causes a strong energetic offset for the imaging frequency.

The system under investigation is an equal mixture of $\approx 4 \times 10^6$ ^{6}Li atoms in states $|1\rangle$ and $|2\rangle$ at $T/T_F \approx 0.07(2)$. The final trap frequencies after evaporation and recompression of the dipole trap are measured to be $(\omega_x, \omega_y, \omega_z) = 2\pi \times (91, 151, 235)$ Hz. Interaction strength is varied by tuning the magnetic field between 740 G to 1 000 G, corresponding to interaction parameters of $-0.8 \leqslant 1/(k_F a) \leqslant 1$ throughout the whole crossover regime. The Fermi energy in the center of the gas for each considered interaction strength is $E_F \approx h \times 34(3)$ kHz. The rf radiation required for the rf dressing of the sample is applied by a 15 mm × 15 mm square coil with two windings. It is placed above the UKEA viewport with roughly 8 mm distance to the atom cloud. The rf signal at ≈ 80 MHz is generated by a synthesizer[1] and amplified with a 100 W amplifier[2]. In order to spectroscopically search for the Higgs mode by modulation of the population, it is necessary to calibrate the excitation, i.e. the amplitude and frequency of the modulation. The population of state $|3\rangle$ during the rf drive can be written as

[1] Agilent MXG N5183A
[2] Kuhne Electronics KUp PA BB 003055-100A

$$p_{|3\rangle} = \alpha \sin^2 (\Omega_{\mathrm{mod}} t) \,. \tag{7.12}$$

Here, $\Omega_{\mathrm{mod}} = \sqrt{\Omega_R^2 + (\delta - \delta_0)^2}$ is the modulation frequency and α the modulation amplitude. We assume a lorentzian lineshape with respect to the detuning, i.e.

$$\alpha = \frac{\Omega_R^2}{\Omega_R^2 + (\delta - \delta_0)^2}. \tag{7.13}$$

Here, δ_0 is an empirical term, taking into account interactions in the final state of the spectroscopy, condensation energy of the initial state and averaging over different momentum states and densities in the trap. Rabi frequency and amplitude can be varied independently by varying the rf power $P \propto \Omega_R^2$ and the frequency of the rf drive. With this, it is possible to keep α constant, while changing the modulation frequency for spectroscopic measurements. In order to calibrate α and Ω_{mod}, we record several Rabi oscillation at different power for fixed δ. This provides information about the amplitude and Rabi frequency. The center frequency of the lorentzian is well calibrated if the Rabi oscillation for a set amplitude is equal for a red and blue detuned rf-drive. We check this at $\alpha = 4\,\%$, where the agreement with the lorentzian model is accurate to a few percent and Rabi oscillations in state $|3\rangle$ are still measurable. Rabi oscillations for equal detuning with opposite sign are shown in figure 7.6. For the spectra themselves, we choose a constant modulation amplitude of $\alpha = 0.5\,\%$ extrapolated from the lorentzian model in the calibration measurements, as this is too small an amplitude to be directly visible during Rabi oscillations. The modulation frequency is varied (while the amplitude is held constant) by simultaneously adjusting the detuning and power of the rf drive. The detuning is always chosen negative, i.e. red detuned. The reason for this is that blue detuned rf can cause single-particle excitations into the continuum. This can be observed by an increase in unpaired atoms, see figure 7.6. After modulation, we perform the rapid ramp and 15 ms time of flight, which is approximately a quarter period of the residual harmonic potential generated by the Feshbach coils during ballistic expansion. This procedure maps the initial momentum distribution to positions in the absorption image [146]. We fit a bimodal distribution to the momentum distribution and calculate the condensate fraction, which provides us with a sensitive measure for the excitation of modes in the quantum gas.

7.4 Spectroscopic measurements

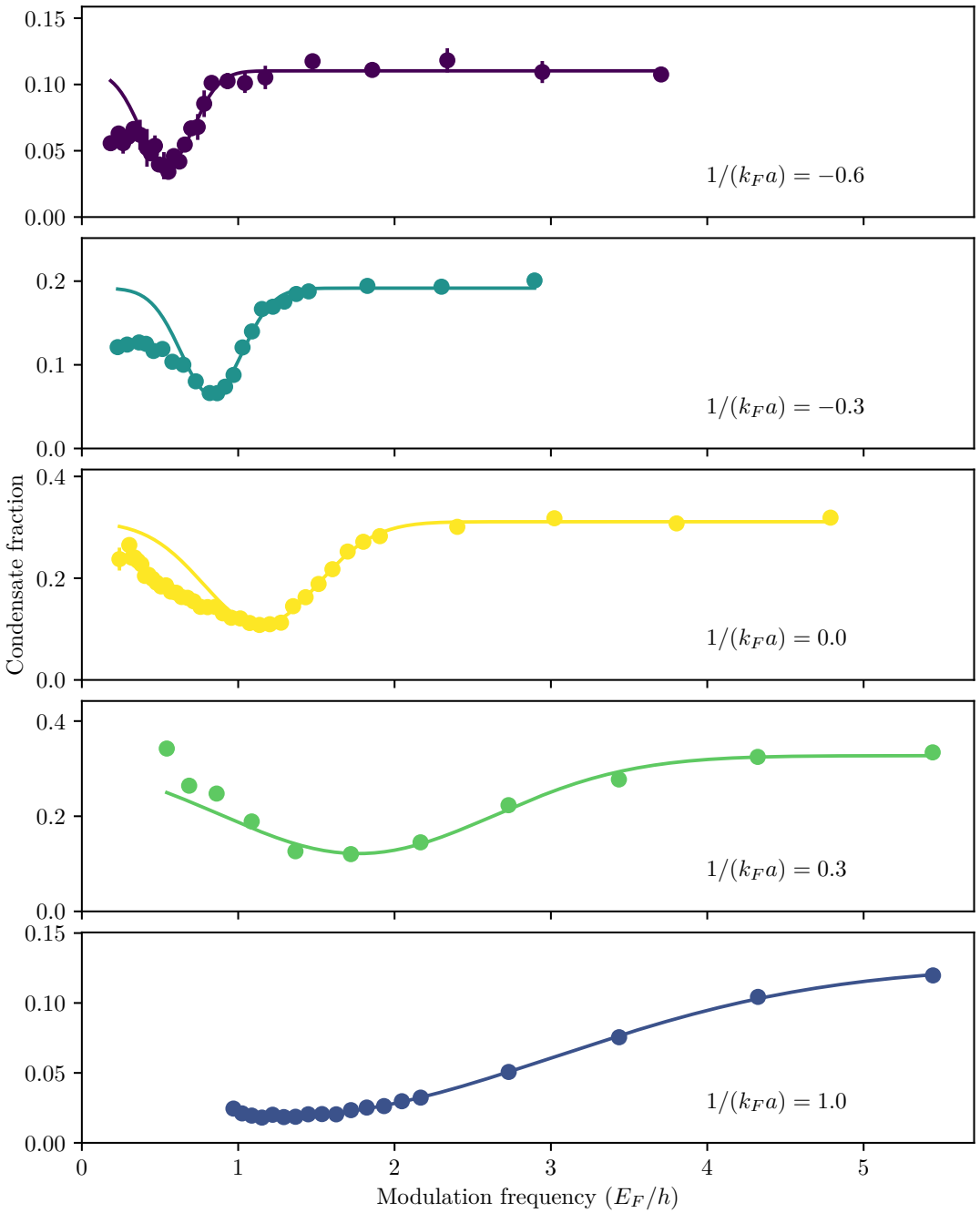

Figure 7.7: Spectra throughout the BEC-BCS crossover regime. We record the condensate fraction after rapid ramp for different modulation frequencies for an rf drive of 30 ms at an amplitude of $\alpha = 0.5\,\%$. Lines show gaussian fits to the high frequency side of the resonance. A shift in resonance frequency and a broadening towards the BEC side can be observed.

In order to observe the Higgs mode, we take spectroscopic measurements and search for a response of the system. Experimentally, we vary the modulation frequency of the population and expect a decrease in the condensate fraction once we reach a frequency that is able to excite the Higgs mode and hence adds energy to the system. We take several spectra at different interaction strengths throughout the BEC-BCS crossover, some of which can be seen in figure 7.7. Each spectrum is measured by first preparing the Fermi gas and subsequent adiabatic ramping to the final magnetic field setting the interaction. For a constant amplitude $\alpha = 0.5\,\%$, detuning and power of the rf drive are varied to scan the modulation frequency. We modulate the sample for 30 ms before the rapid ramp and observe the condensate fraction extracted from a bimodal fit. Starting on the BCS side, the spectra show a constant condensate fraction towards high modulation frequencies, but a pronounced dip towards low frequency. On the low-frequency side of the dip, the condensate fraction does not recover to its high-frequency value. One contribution to this asymmetric lineshape stems from the momentum dependence of the effective driving frequency, see figure 7.3. This effective detuning can cause a resonant excitation for high momenta, even though for low momenta the modulation frequency is below resonance. On the BCS side of the resonance up to unitarity, we observe a clear resonance. For lower fields with $1/(k_F a) < 0$, the energy peak is gradually washed out and broadens significantly. Towards the BEC limit $1/(k_F a) \approx 1$, we cannot observe a resonance and conclude that the Higgs mode is absent. For different interactions, a pronounced shift in resonance frequency as well as width can be observed. We extract both quantities by fitting a gaussian to the high frequency side of the resonance.

To verify the collective-mode nature of this resonance, we perform a number of checks. Firstly, we verify that the frequency and shape of the resonance is not dependent on the amplitude or duration of the modulation, see figure 7.8. We separately check the effect of different modulation times by observing the decrease in condensate fraction for different amplitude and modulation time and verify that we are in the regime of linear response, see figure 7.8. An interesting consequence of the collective nature of the Higgs mode is the fact that the frequency of the mode is the same throughout the whole sample and determined by the superconducting gap evaluated at the maximum density of the gas [189, 200–203]. This is in contrast to quasiparticle excitations, which have a momentum dependency due to their dispersion relation. To verify the momentum independence, we split the measured momentum distribution in radial averages with distances of $0.02\,k_F$ in radial direction. Each radial average represents a different momentum range in the cloud. The position of the resonance is then represented by a dip in the radially averaged optical density, as shown in figure 7.9. It can be seen, that the position of the resonance is stable throughout the momentum distribution of the condensate.

The width of the spectral feature can be explained by the numerical simulations. For this, we simulate the effect of different driving frequencies. For each driving frequency, we generate a plot like figure 7.5. As the rf drive applies to all momenta within the trap, we integrate each plot along the momentum axis. From the resulting Fourier spectra, we numerically extract the spectral width underneath the peak corresponding to the Higgs mode. This of course has some error, as for some momenta the Higgs peak overlaps with the other contributions, however, this is just a small correction, as most of the time the Higgs peak is the dominant feature (see figure 7.4). For high driving frequencies, we find a non-vanishing contribution to the Higgs weight, however, we attribute this to the excitation of quasi-particles in a

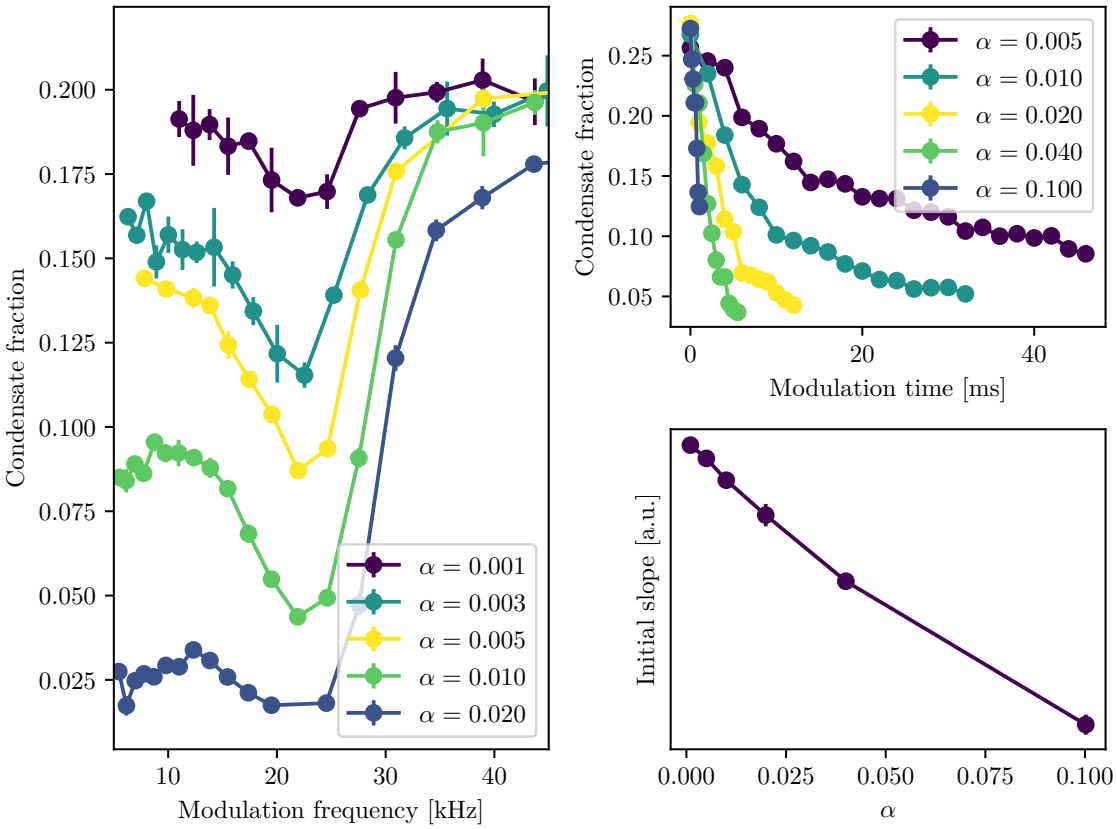

Figure 7.8: Modulation dependency of the spectra. **(Left)** Shown here are spectra at $1/(k_F a) = -0.4$ for several different values of α. Although the shape on the low frequency-side due to off-resonant excitation of higher momenta changes, the center position and shape of the resonance does not. **(Top right)** Changing the amplitude of the modulation changes the time scale of the depletion of the condensate fraction. **(Bottom right)** Fitting the first five datapoints of the curves from the top right plot, we can verify that the response to the modulation is proportional to the amplitude of the modulation, which confirms that we are in the linear response regime.

homogeneous system that vanish in a harmonically trapped system. A comparison between the experimental and simulated data at $1/(k_F a) = -0.6$ is shown in figure 7.10. The fitted widths to both curves are in the order of Δ and agree within the error bars.

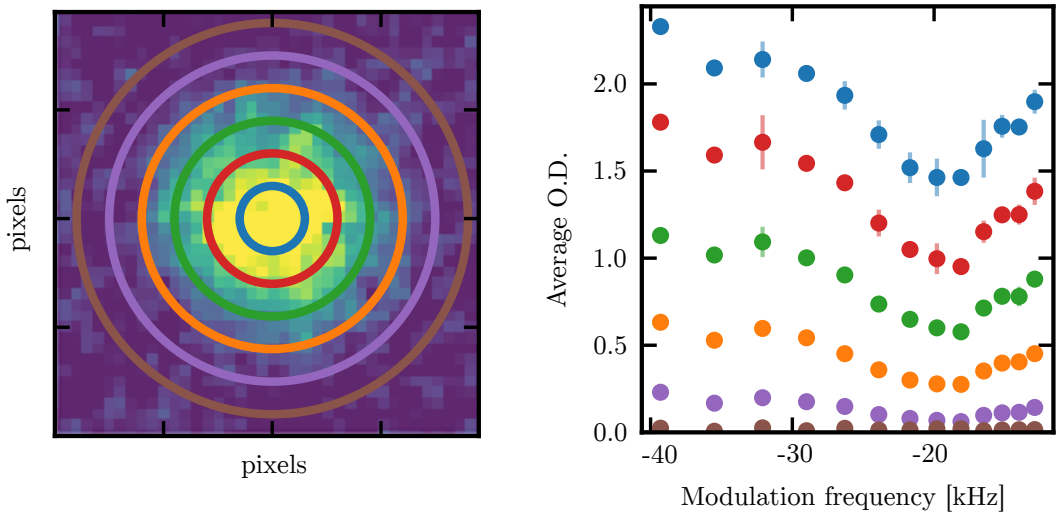

Figure 7.9: Momentum dependency of the spectral feature. The position of the mode does not change for different momentum components of the cloud extracted in time of flight. The colored rings on the left correspond to curves of the same color on the right.

7.5 Results

A stable Higgs mode is predicted to occur at a frequency corresponding to two times the gap energy $\hbar\omega_H = 2|\Delta|$. Figure 7.11 shows the measured frequency of the resonance feature as determined from fits to the high frequency side of the resonance for different values of $1/(k_F a)$ evaluated at the center of the sample. As for a stable Higgs mode, the frequency is directly connected to the gap value $|\Delta|$, it can serve as a measure for the superfluid gap parameter. In the crossover regime, the exact value of $|\Delta|$ is unknown and theoretical calculations and experimental determination is difficult. We compare our approach to measurements [40, 41] and several numerical calculations [19, 35–39] and find overall good agreement. Compared with the experimental values, our approach yields larger values. Both references rely on a fit to an onset of a spectral feature, whereas our approach fits a skewed spectral feature. Both approaches are prone small systematic uncertainties, which might explain the difference. An upper bound is given by mean field theory (dashed line), which is known to overestimate the gap value.

The Full-width-at-half-maximum (FWHM) of the resonance can be extracted from the Gaussian fits as well. The width of the excitation resonance itself can be estimated from our calculations to be on the order of the superfluid gap Δ. The linewidth can hence not be interpreted as the decay rate of the Higgs mode and only serves as a lower limit of the lifetime. Towards the BCS side of the resonance, the FWHM of the feature agrees well

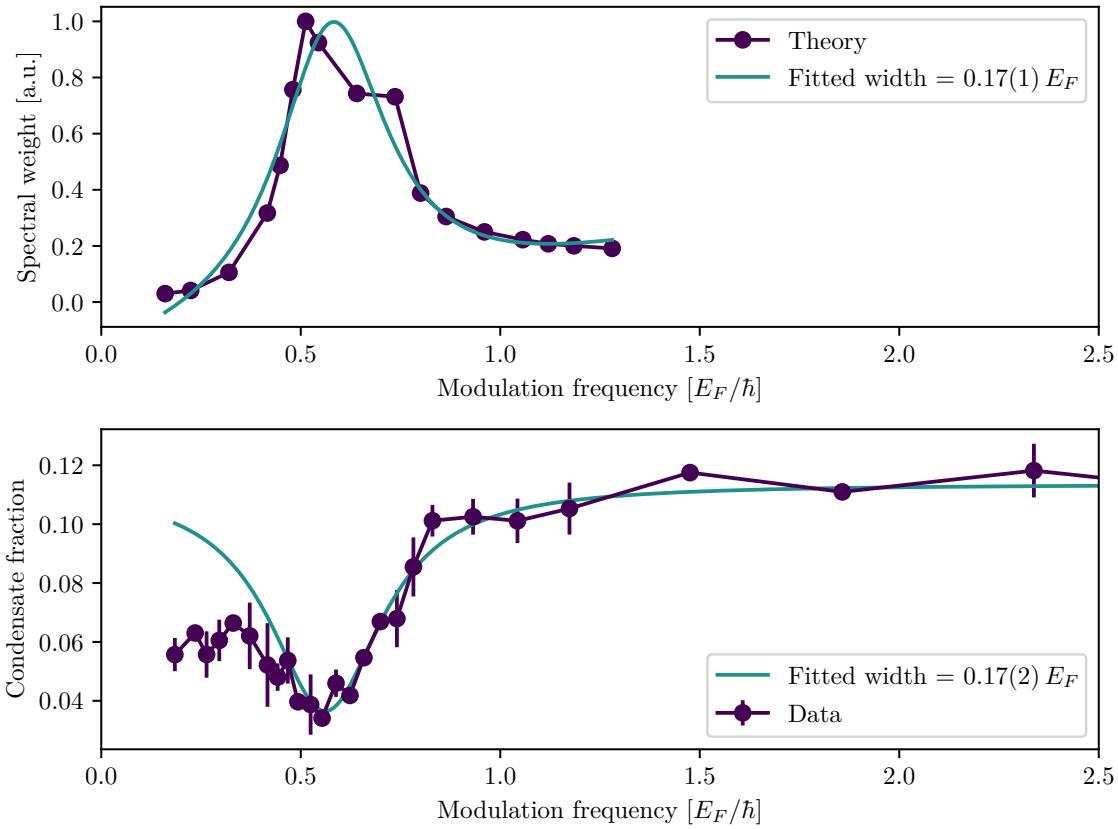

Figure 7.10: Comparison between theoretically predicted weight and experimental results. Both features are fitted with a lorentzian lineshape, in case of the theory curve with an additional linear background.

with this estimation, however, towards the BEC side the feature broadens significantly (see figure 7.7 for a direct comparison). This indicates that the Higgs mode becomes unstable, most likely due to the gradual violation of particle-hole symmetry on the BEC side of the resonance.

In summary, the presented excitation schemes presents the first direct excitation scheme for the Higgs mode in a strongly interacting fermionic superfluid. The superfluid gap is a quantity of great interest in the BEC-BCS crossover, as it is directly connected to the physics of Cooper pairs. So far, measuring it directly was difficult, however, measuring the frequency of the Higgs mode provides a way to measure the absolute value of the order parameter in the regime of a stable Higgs mode. Furthermore, with shorter modulation time and stronger drive, the measurement of gap dynamics might become feasible. With improved momentum resolution our scheme might also pave the way to explore the decay mechanisms of the Higgs mode in more detail.

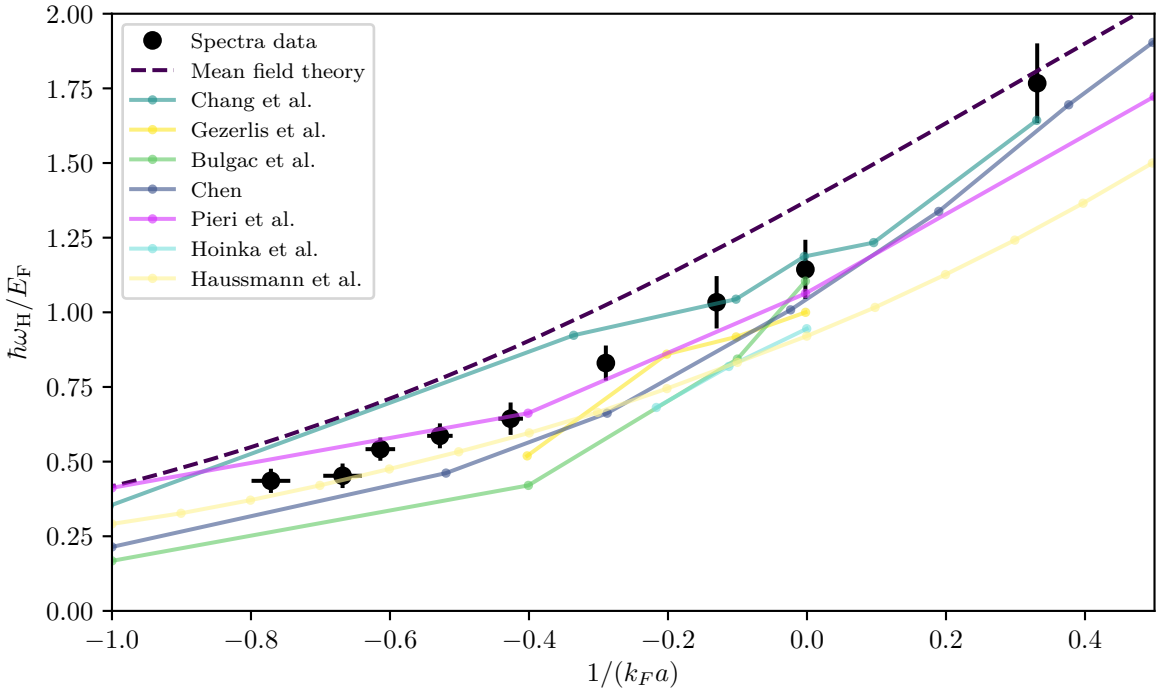

Figure 7.11: Measured Higgs mode frequencies. Measured Higgs mode frequencies (black dots), compared with different theoretical and experimental predictions of the gap energy for different interactions. The corresponding Higgs frequency is calculated via $\hbar\omega_H = 2|\Delta|$. The dashed line corresponds to the mean field solution, which is known to overestimate Δ. Additionally to our result, we show the work by Chang et al. [35], Gezerlis et al. [36], Bulgac et al. [19], Chen [37], Pieri et al. [39], Hoinka et al. [41] and Haussmann et al. [38].

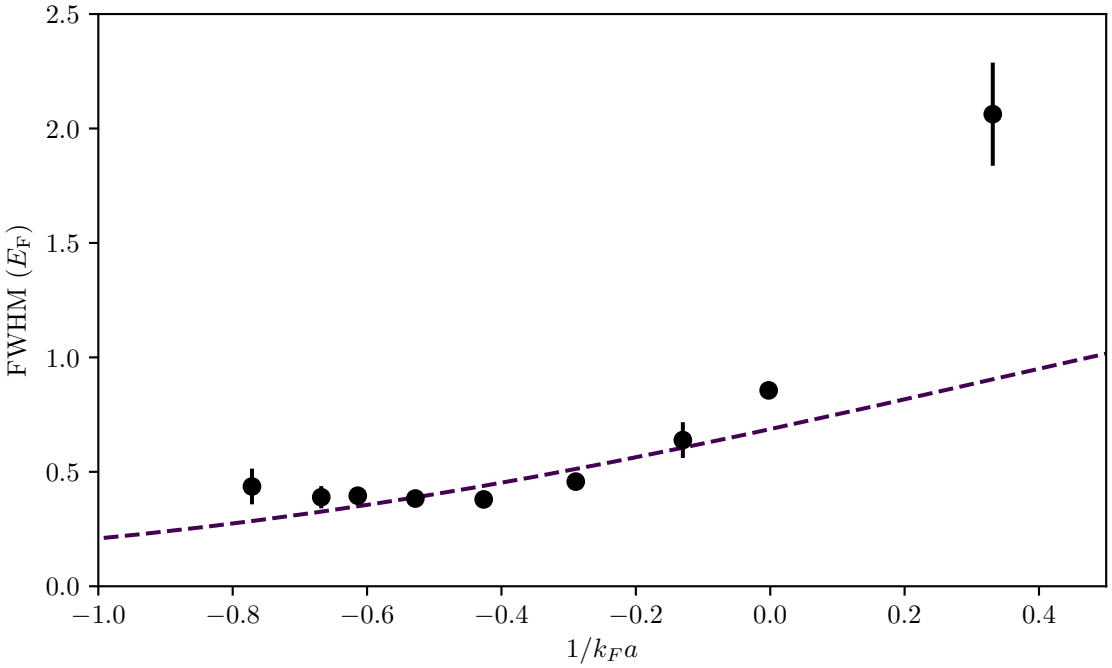

Figure 7.12: Width of the Higgs mode feature. Measured width of the resonance feature as determined by a gaussian fit to the high-frequency side (black dots) and mean field Δ (dashed line) for different interaction strengths.

CHAPTER 8

Rapid Quenches in the BEC-BCS Crossover

This chapter investigates the topic of fast quenches and non-equilibrium physics in the BEC-BCS crossover. Two kinds of interaction quenches will be presented: Firstly, by a rapid radio-frequency transfer of half the population to a different state. This will change the interactions of the constituents of the pairs, as the position and shape of the corresponding Feshbach resonance changes. Secondly, a method for very fast magnetic field quenches will be presented. For this, a dedicated quench coil has been designed, manufactured and tested. This coil setup can change the magnetic field by $\approx 36\,\mathrm{G}$ within $\approx 3\,\mu\mathrm{s}$ and is therefore faster than any inherent time scale of the system. The contents regarding radio-frequency quenches are currently available as a preprint on arXiv [2]. A publication detailing the design of the quench coil used for the magnetic field quenches is in preparation for publication [204].

8.1 Motivation and previous work

The motivation for the investigation of non-equilibrium properties of systems is varied across many fields [44]. In his 1944 book *What is life?* [205] Erwin Schrödinger concludes, that life has to continually increase entropy, to avoid decaying to an *inactive* equilibrium state. Non-equilibrium physics phenomena are not only found in field of active matter [206] or biology [207], but also in condensed matter systems [45, 46] and cold atom or trapped ion systems [208]. Non-equilibrium dynamics of superconductors and superfluids might hold the key to understanding how to *switch* superconductivity on and off at will. Among the first to investigate infinitesimally small instantaneous changes of the order parameter Δ in the context of the BCS Hamiltonian were Volkov and Kogan in 1974 [209]. They showed that the absolute value $|\Delta|$ acquires coherent oscillations at a frequency of $2|\Delta|/\hbar$ attributed to the Higgs or amplitude mode (whose existence in a strongly interacting superfluid we proved in chapter 7), decaying over time with a power law dependency $\propto t^{-1/2}$. It was later found, that the short-time behaviour of the system can be described by the integrable BCS model and that for longer times the dynamics reduces to more simple effective Hamiltonians with only a few degrees of freedom, even for strong excitations [43, 210–212]. Although the general situation is a quantum many-body problem, the time evolution can exhibit large anharmonic oscillations, which are reminiscent of the physics of collapse and revival known from few-particle dynamics [213]. Using different methods, the results of Volkov

and Kogan were later expanded upon by Yuzbashyan et al., who not only confirmed time damped oscillation for a certain range of quenches, but also showed that for certain quench parameters, persistent oscillations where expected [212, 214]. The connection between both regimes was investigated by Barankov and Levitov [211], who showed a continuous transition between both regimes. Including the full range of possible quenches $\Delta_i \to \Delta_f$ from an initial gap value Δ_i to a final value Δ_f, three distinct regimes have been shown [43], see figure 8.1:

1. In Region I, for large Δ_i and small Δ_f, the order parameter vanishes exponentially,

$$\Delta(t \gg \Delta_i^{-1}) \propto (\Delta_i t)^{-1/2} \exp(-2\Delta_i t). \tag{8.1}$$

2. In Region II, where both are comparable the order parameter exhibits a power law decay with oscillations at $2|\Delta|/\hbar$.

$$\Delta(t) \propto \Delta_\infty + A(t)\sin(2\Delta_\infty t + \alpha), \qquad A(t) \propto t^{-1/2} \tag{8.2}$$

3. In Region III, for large Δ_f, the order parameters exhibits persistent oscillations.

$$\Delta(t) \propto \Delta_\infty + A \cdot \sin(2\Delta_\infty t + \alpha) \tag{8.3}$$

Here, Δ_∞ describes the long time value of the gap parameter $\Delta_\infty = \Delta(t \to \infty)$. As the above results are calculated from integrable theory, it remains under discussion, how applicable they are to a Fermi gas system close to unitarity with $1/(k_F a) \leqslant 1$ and large quasiparticle scattering rates, where the observability and lifetime of coherent dynamics is critically dependent on the relaxation mechanisms and lifetimes of quasiparticle excitations. For holographic superconductors using AdS/CFT correspondence, depending on the quench parameters, oscillating and steady-state final states depending on the strength of the quench have been found. However, these are only valid on unitarity, where the scattering length diverges [215].

8.1.1 What does rapid mean?

In an experimental setting, naturally, instantaneous changes of the interaction are not possible. However, for the system to *perceive* the change as instantaneous, the time required for the quench has to beat the timescales of the relevant dynamics. The most natural time scale is the time associated with the Fermi energy $\tau_F = \hbar/E_F$, which in the case of the presented measurements is on the order of $\approx 4.7\,\mu s$. The second energy scale relevant to the system is the timescale associated with the gap energy, namely $\tau_\Delta = \hbar/\Delta$. The timescale for incoherent processes is given by the quasi particle relaxation time $\tau_\epsilon = \hbar E_F/\Delta_f^2$. This time scale varies with interaction between unitarity and $1/(k_F a) = -1$ between $60\,\mu s$ and $150\,\mu s$. Another important timescale is given by the natural timescale of the harmonic trap. Here, the dynamics in a given direction is determined by the inverse of the trap frequency $\tau_{\text{trap}} = 1/\nu_i = 2\pi/\omega_i$. Depending on the exact trap frequencies, for the following experiments,

this will be in the order of milliseconds. All relevant numbers will be given more accurately when needed.

8.2 Fast quenches by rapid state transfer

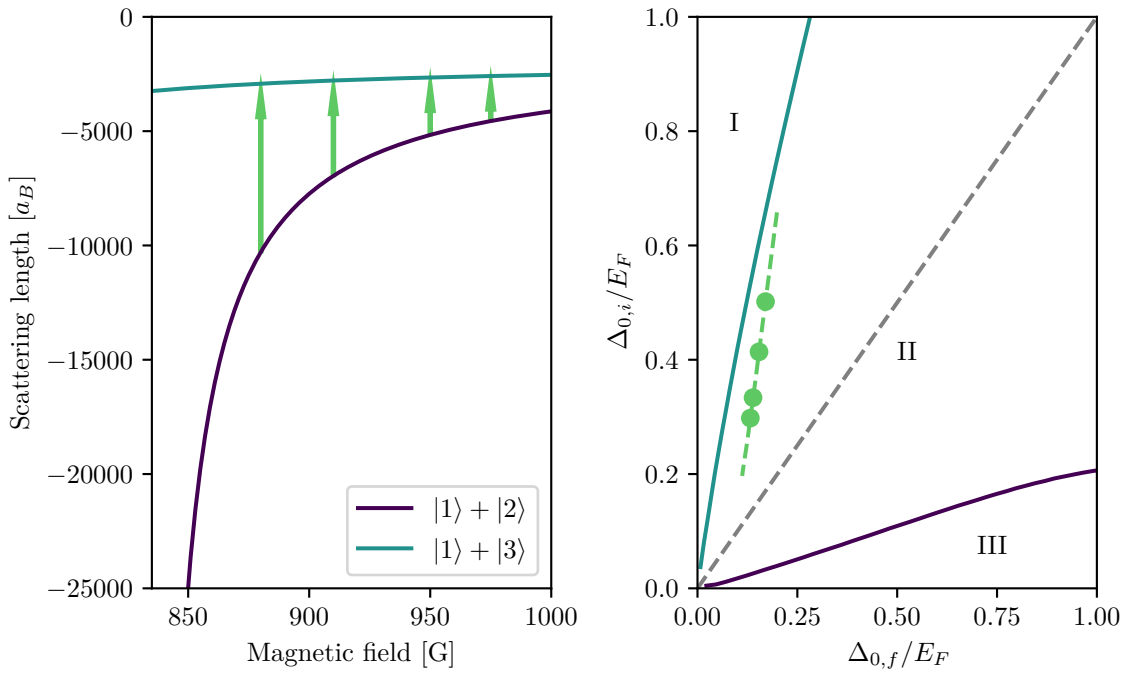

Figure 8.1: Concept of rf quenches. (Left) The initial cloud is prepared in a equal mixture of $|1\rangle$ and $|2\rangle$. By applying a π-pulse on the $|2\rangle \rightarrow |3\rangle$ transition, the mixture is changed to an equal mixture of $|1\rangle$ and $|3\rangle$. Consequently, due to the shape and position of the Feshbach resonances (indicated by the lines), different interaction quenches can be realised (green arrows show the quenches used in the experiment). **(Right)** The quench phase diagram adapted from ref [43] with regions I, II and III indicated by the labels. As the range of quenches is limited by the difference between the interactions on the Feshbach resonances, only a subset of quenches can be probed. The green datapoints indicate the used quench parameters, the dashed green line shows potential quench parameters dictated by the relative position of the Feshbach resonances. The values for Δ are calculated via mean-field theory.

The first realised method of quenches applied to our Fermi gas is achieved through a fast radio-frequency transfer of one state of the two state mixture to a different, third state. For this, we prepare a Fermi gas in a mixture of states $|1\rangle$ and $|2\rangle$ and then do a fast state transfer $|2\rangle \rightarrow |3\rangle$. This constitutes a rapid change of interaction between the constituents of the pairs. Part of this work has been described in the theses of Dr. Alexandra Behrle [129] and Dr. Timothy Harrison [130], however, since then the analysis of the present data was improved upon, additional data was added and the interpretation of the results changed significantly. The work presented in this chapter is currently in preparation for publication and available as preprint [2].

Final magnetic Field [G]	Transfer efficiency [%]
834	67.0
850	81.3
865	86.5
880	88.0
895	90.7
910	97.2
1000	98.4

Table 8.1: Efficiency of the rf-transfer on the $|2\rangle \rightarrow |3\rangle$ transition for different magnetic fields settings.

8.2.1 Antenna and calibration

To drive the $|2\rangle \rightarrow |3\rangle$ rf-transition, we use the same $15 \times 15\,\mathrm{mm}^2$ square coil that was used in chapter 7 to modulate the populations. This coil has two windings and is placed above the UKEA viewport with roughly 8 mm distance to the atom cloud. The rf-signal at $\approx 80\,\mathrm{MHz}$ is generated by a Synthesizer[1] and amplified with a 100 W Amplifier[2]. The magnetic field at which the quench takes place is tuned between 834 G and 1 000 G, corresponding to $0 \leq 1/(k_F a_i) \leq -0.9$ and $-1.2 \leq 1/(k_F a_f) \leq -1.4$, where a_i and a_f correspond to the initial scattering length on the $|1\rangle + |2\rangle$ Feshbach resonance and the final scattering length on the $|1\rangle + |3\rangle$ Feshbach resonance, respectively. In order to determine the efficiency and the time needed for a full transfer from $|2\rangle$ to $|3\rangle$, we recorded several Rabi oscillations. By maximising the amplitude of the oscillations, we made sure to be on resonance. The measured transfer efficiency at different magnetic fields is given in table 8.1. We observe a decrease in the transfer efficiency towards unitarity at 834 G, most likely due to heating effects and dephasing caused by strong interactions. To circumvent any dynamics arising from a three-body mixture at low transfer efficiency, we only choose fields with a transfer efficiency higher than 88 % for quenches. The final duration of a full transfer is $\tau = 28(2)\,\mu\mathrm{s} = 4.8\,\tau_F$, which is slow as compared to τ_F, but comparable to τ_Δ and τ_ϵ and much faster than τ_{trap}.

8.2.2 Data acquisition

For the quenches, we prepare an equal mixture of states $|1\rangle$ and $|2\rangle$ with $N = 1 \times 10^6$ atoms per state. After evaporation in the optical dipole trap, the temperature of the sample is $T = 0.07(2)\,T_F$. In harmonic approximation, the trap frequencies of the final trapping configuration are $(\omega_x, \omega_y, \omega_z) = 2\pi \cdot (110, 151, 235)\,\mathrm{Hz}$ and the Fermi energy in the center of the trap is $E_F \approx h \times 29(3)\,\mathrm{kHz}$ for all quenches. After evaporation we ramp the magnetic field to the field value corresponding to the desired interaction strength within 200 ms with a subsequent waiting time of 200 ms for equilibration. After the quench plus variable waiting time and before imaging, the weakly bound pairs on the BCS side are converted to tightly bound molecules via the rapid ramp technique (see section 4.3). The rapid ramp procedure starts at the same time the dipole trap is switched off for time-of-flight imaging. In order to

[1] Agilent MXG N5183A
[2] Kuhne Electronics KU PA BB 003055-100A

image the momentum distribution, the cloud expands for 15 ms, which is a quarter period of the remaining potential generated by the curvature of the Feshbach fields. This maps the momentum distribution to the density of the cloud and therefore enables the imaging of the bimodal momentum distribution [146]. Note that after the rf-transfer the cloud is in a mixture of states $|1\rangle$ and $|3\rangle$. Consequently, all rapid ramp parameters are adjusted according to the $|1\rangle + |3\rangle$ Feshbach resonance with its zero crossing around 569 G.

8.2.3 Analysis

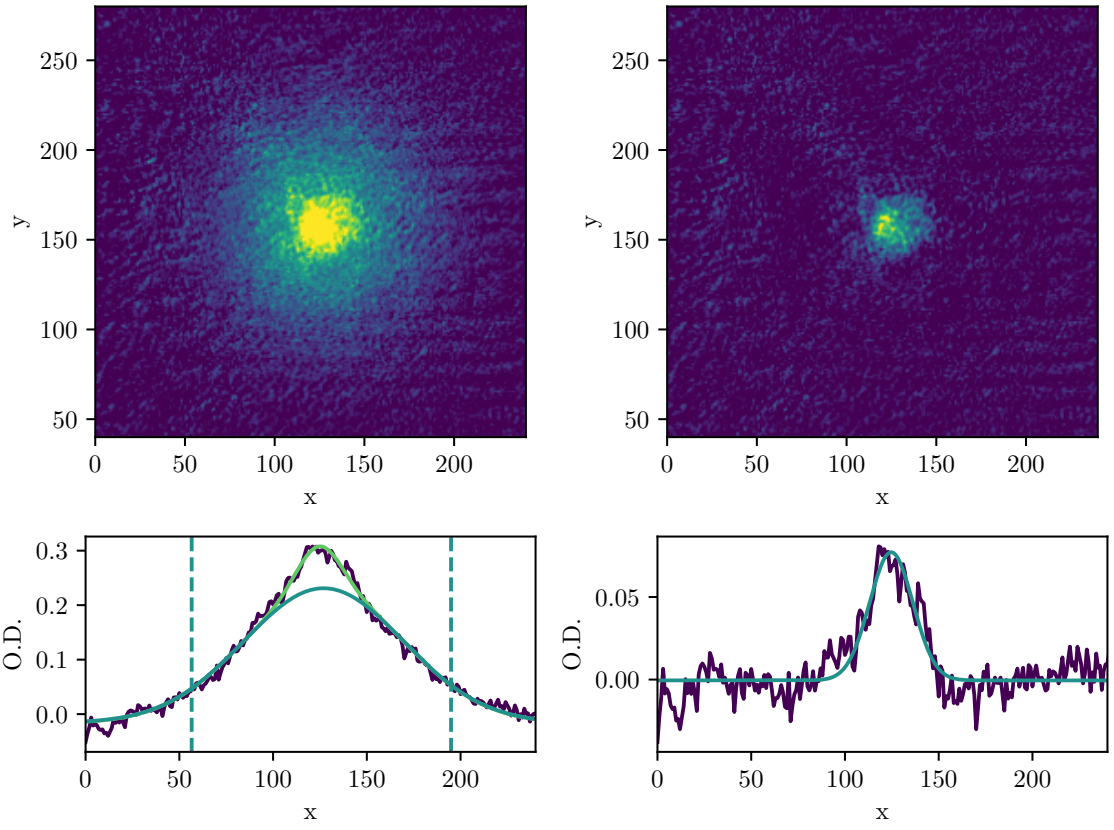

Figure 8.2: Illustration of the analysis for the rf quench data. (Top left) Averaged OD picture taken at 910 G corresponding to $1/(k_F a_f) = -1.22$. **(Bottom left)** Mean along the y-axis of the above picture. Indicated are the boundaries of the thermal fit (dashed lines), the thermal fit (solid turquoise line) and the thermal plus condensate fit (solid green line). **(Top right)** Averaged OD picture after subtraction of the thermal fit. **(Bottom right)** Average along the y-direction of the above picture. Indicated is the fit to the remaining density distribution (turquoise line).

The usual procedure for the analysis of bimodal distributions is to fit an inverted parabola centered within a Gaussian distribution to the optical density picture. In order to gain more flexibility, we modify this fitting procedure in the following way: To get initial guesses for the size of the condensate and thermal distribution, we fit the usual bimodal distribution.

We then apply a cut at three times the size of the parabola and mask out the center region. From this masked picture, we fit a Gaussian distribution to the wings of the cloud. As the center region containing the condensate is masked out, the edges used for fitting only contain thermal atoms and the resulting Gaussian distribution should describe the thermal background. In the second step, we subtract the fitted thermal distribution from the optical density picture. This leaves a picture that only contains the condensate part of the cloud. This condensate is again fitted with a Gaussian distribution. The different steps of the procedure are highlighted in figure 8.2. From the parameters of both Gaussian distributions several quantities can be calculated. One example is the condensate fraction, which is defined via

$$\text{condensate fraction} = \frac{N_{\text{condensate}}}{N_{\text{thermal}} + N_{\text{condensate}}}. \tag{8.4}$$

As optical density is proportional to atom density (see section 4.2 for details), the number of atoms in the condensate or thermal part is proportional to the integral over the corresponding Gaussian distribution $N_i \propto \int dxdyg_i(x,y) = \pi\sigma_x\sigma_y A_x A_y$. Here, $i \in [\text{condensate}, \text{thermal}]$ and $g(x,y)$ is a two-dimensional Gaussian distribution. This way of fitting has two major advantages as compared to directly fitting a bimodal distribution. Firstly, for low condensate fraction, fitting the parabola often just corrects the misshaping of the underlying thermal distribution due to imaging effects or noise by making the parabola as broad as the Gaussian distribution and with very low amplitude. This might reduce the fitting error, as most of the weight is distributed in the thermal cloud, however, this does not reflect the condensate. Subtracting the thermal distribution in the first step and using good initial guesses makes sure that the dominant signal for fitting is the condensate. Secondly, by fitting two independent gaussians, all parameters for the condensate and thermal distribution are independent of each other. Notably, this for example lifts the restriction of the parabola being centered on the thermal cloud, which enables the detection of the motion of thermal and condensate parts separately.

8.2.4 Time evolution

After evaporation close to the Feshbach resonance at 795 G, the magnetic field is adiabatically ramped to its final value within 200 ms, setting the initial interaction parameter $1/(k_F a_i)$. After an initial wait time of 200 ms for equilibration, the rf-transfer is carried out. Followed by this is a variable wait time before rapid ramp and time of flight starts, allowing for the investigation of the time evolution of the fitted and calculated parameters. Qualitatively, the situation at the quench time can be considered as follows: The initial state at interaction strength $1/(k_F a_i)$ has a superconducting gap of Δ_i and a chemical potential μ, which govern the density profile in the trap. After quenching to a weaker interaction strength $1/(k_F a_f)$ by the rf transfer, the gap is reduced to a nominal value of Δ_f and the chemical potential increases. Both effects trigger internal dynamics and an adjustment of the density distribution which we investigate after the quench. Example curves for the width of the condensate part and the condensate fraction can be seen in figure 8.3 for a quench from $1/(k_F a_i) = -0.65$ to $1/(k_F a_f) = -1.28$, with all other curves looking qualitatively similar. For the width of the

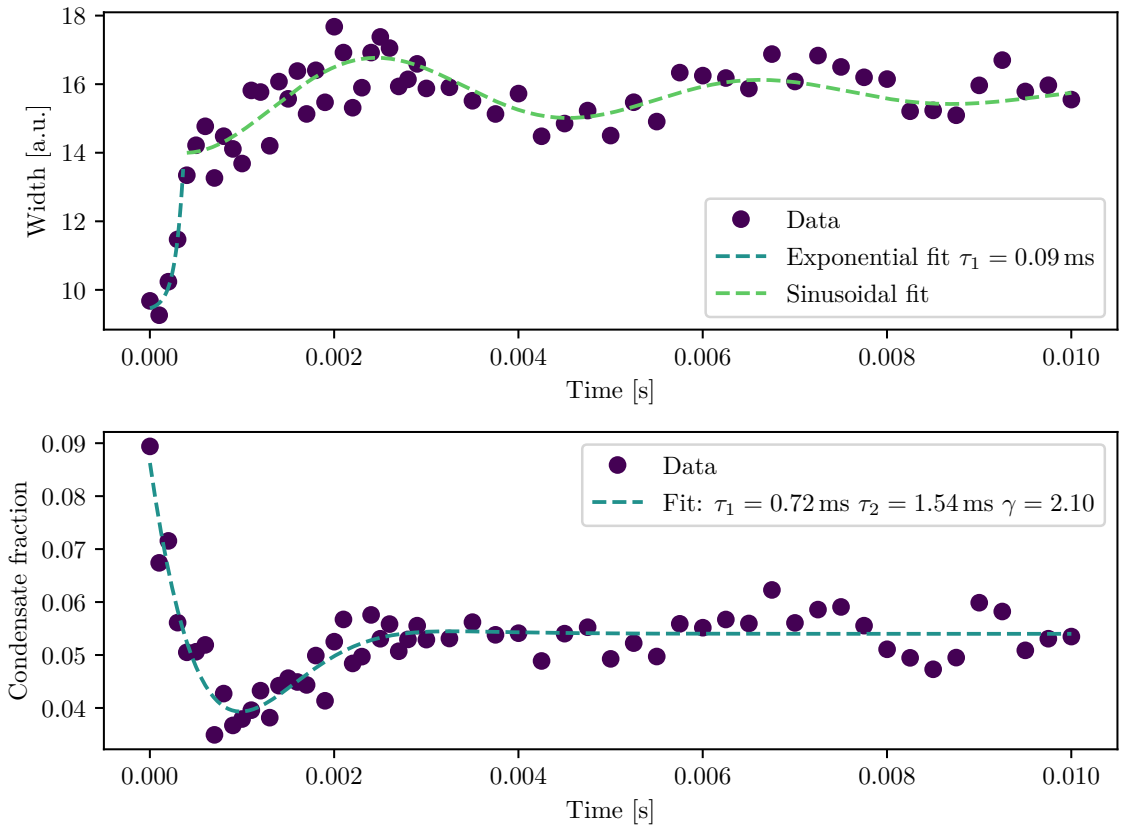

Figure 8.3: Time evolution of density and condensate fraction. The purple datapoints are taken at $1/(k_F a_f) = -1.28$ after a quench from $1/(k_F a_i) = -0.65$. **(Upper)** Evolution of the width of the condensate part of the cloud. Indicated are the initial exponential increase (turquoise dashed line) and an exponentially decaying sinusoidal fit to the subsequent oscillation (green dashed line). **(Lower)** Evolution of the condensate fraction with a fit of the function given by equation (8.6) (turquoise dashed line).

condensate, two features are of interest. For one, a fast initial broadening and secondly, a decaying oscillation that does not significantly change throughout the range of interactions we observe. The latter can be explained by trap dynamics. As the time scale of the quench is orders of magnitude faster than the time scale of the trap $\tau_{\mathrm{trap}} \gg \tau_{\mathrm{quench}}$ and the perturbation is proportional to the initial atom density, the interaction quench serves as an excitation mechanism for the collective monopole mode. We verify this by solving a hydrodynamic model of the monopole mode for our trap geometry [216, 217]. This model does not include interaction-dependence, but only uses the equation of state for the unitary Fermi gas and does therefore not possess any dependency on the interaction. The resulting frequency from the model agrees well with the fitted frequency from an exponentially decaying sinusoidal fit to the oscillation of the condensate width and thermal width. This verifies that the Fermi gas is still hydrodynamic as opposed to non-interacting. We verify the correct scaling of the monopole mode frequency by varying the trap confinement. The initial increase in width of

the condensate up to $\approx 400\,\mu s$ is fitted with an exponential function

$$w(t) = w_0 + w_1 \exp\left(\frac{t}{\tau_0}\right). \tag{8.5}$$

The resulting timescale τ_0 in the order of $\tau_0 \approx 100\,\mu s$ is comparable to the quasiparticle excitation time $\tau_\epsilon = \hbar E_F/\Delta_f^2$ with Δ_f calculated from equilibrium mean-field theory. This hints at the redistribution of excess quasiparticles as underlying mechanism for the initial increase in size. We do not observe a corresponding broadening in the width of the thermal cloud.

Looking at the condensate fraction, we observe two timescales. For one, an initial decrease and secondly, a revival. To extract both time scales, the time evolution of the condensate fraction is fitted with an initial exponential decrease and a stretched exponential increase for the revival

$$f(t) = A_i \exp\left(-\frac{t}{\tau_1}\right) + A_f \left(1 - \exp\left[-\left(\frac{t}{\tau_2}\right)^\gamma\right]\right). \tag{8.6}$$

Here, A_i and A_f are the initial and final condensate fraction, respectively. γ is the stretch factor, which is fitted to $\gamma = 2.1(1)$ throughout all datasets. The condensate fraction reaches a steady state in $\approx 500\,\hbar/E_F$. τ_1 has a weak dependence on the interaction parameter $1/(k_F a_f)$ and τ_2 is about twice τ_1. For slow quenches, no revival of the condensate fraction or oscillations corresponding to the Higgs mode are expected [218, 219]. The timescale of the superfluid is surprisingly slow and the timescale of the revival is comparable to the timescale of the trap rather than the quasiparticle relaxation time. During the initial decay of the condensate fraction, the width of the condensate nearly doubles. This is in contrast to the width of the thermal background which also increases, but only by less than ten percent.

8.2.5 Theoretical comparison

In order to verify the dynamics and counter-check the interpretation, theoretical simulations on the attractive Hubbard model have been performed by Johannes Kombe, Jean-Sébastien Bernier and Corinna Kollath [183]. The simulations are done in the limit of low filling with $N = 80$ lattice sites on a 1D lattice, where the lattice effects are minimised, with populations $N_{|1\rangle} = N_{|2\rangle} = 8$ in state $|1\rangle$ and $|2\rangle$, respectively. The interaction parameters are given by $U_{12} = -6\,J$, $U_{13} = -2\,J$ and $U_{23} = -2\,J$ with tunneling rate J. The energy offset to the third level is $V_3 = 50\,J$. The simulated rf drive has frequency $\hbar\omega_{\mathrm{rf}} = 52.9\,J$, amplitude $\hbar\Omega_{23} = 1\,J$ and pulse duration $J\tau_{\mathrm{rf}} = 3.45\hbar$. A trapping potential of $V_{\mathrm{trap}} = 0.001\,J$ is present for all levels. We solve the system numerically using the time-dependent density matrix renormalization group method and verify good convergence of the numerical results. Specifically, we take into account the time dependence of the transfer and interactions in the final and initial state. The results are shown in figure 8.4. Here, the blue shaded area marks the calibrated duration of the pulse, ending at $tJ = 3.45\hbar$. The initially empty level $|3\rangle$ is populated during the rf drive and stays constant afterwards. However, atoms continue to

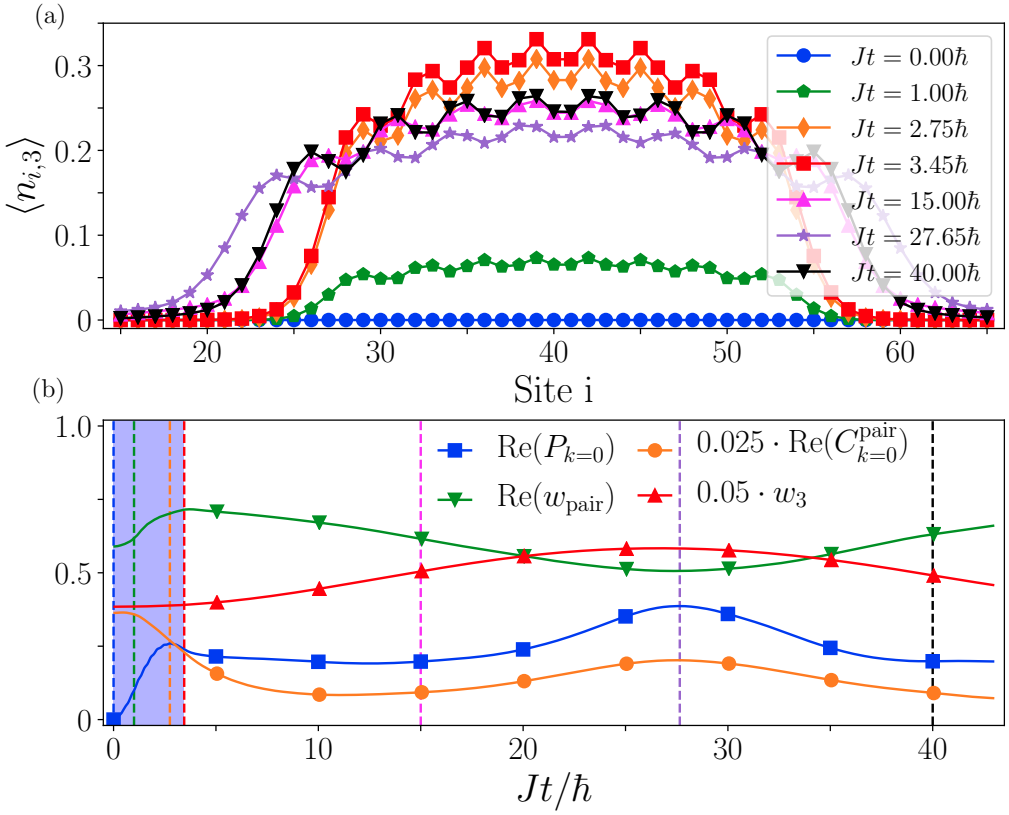

Figure 8.4: Theory of the rf quenches. (Top) Snapshots of the density distribution of state $|3\rangle$ in the trap, starting from an empty population at $t = 0$. **(Bottom)** Time evolution of different observables: The second moment of the density distribution in state $|3\rangle$ w_3^2, Coherence between pairs of atoms in states $|1\rangle + |3\rangle$ $\mathrm{Re}(P_{k=0})$, the second moment of the pair correlation distribution in momentum space w^2_{pair} and the pair coherence rescaled by density $C^{\mathrm{pair}}_{k=0}$. The shaded area marks the duration of the rf-pulse. Dashed vertical lines correspond to the time steps in the upper panel.

redistribute, broadening the density profile shown in the upper plot up to $tJ = 27.65\hbar$ with a subsequent contraction. This signals the excitation of the collective monopole mode by the rf-quench. With the application of the rf pulse, initially the number of pairs formed between $|1\rangle$ and $|3\rangle$ increases. With this, the pair coherence $\mathrm{Re}(P_{k=0})$ and the second moment of the pair correlation momentum distribution w^2_{pair} display a strong initial increase, as shown in the lower plot of figure 8.4. These quantities are related to the condensate fraction and the width of the cloud, respectively. After the initial increase, the pair coherence starts to decrease before the end of the rf-pulse due to the loss of coherence over long distances which is no longer compensated by the increase of atoms in state $|3\rangle$. Looking at the long distance pair distribution rescaled with density $C^{\mathrm{pair}}_{k=0}$, which is not complicated with density dynamics, an initial decrease corresponding to a decay in pair coherence can be seen. At later times,

the pair coherence oscillates with a maximum at the time, where the density distribution is largest, allowing for longer range coherence. Consequently, the long pair coherences measured by the zero-momentum peak are influenced by the monopole oscillations in the density, and therefore also show oscillations with the monopole frequency. The oscillations of the width of the pair momentum peak and its amplitude are not in phase. The dynamical effects uncovered by this model are in good agreement with the experimental findings and support strongly the interpretation provided earlier.

8.2.6 Conclusion

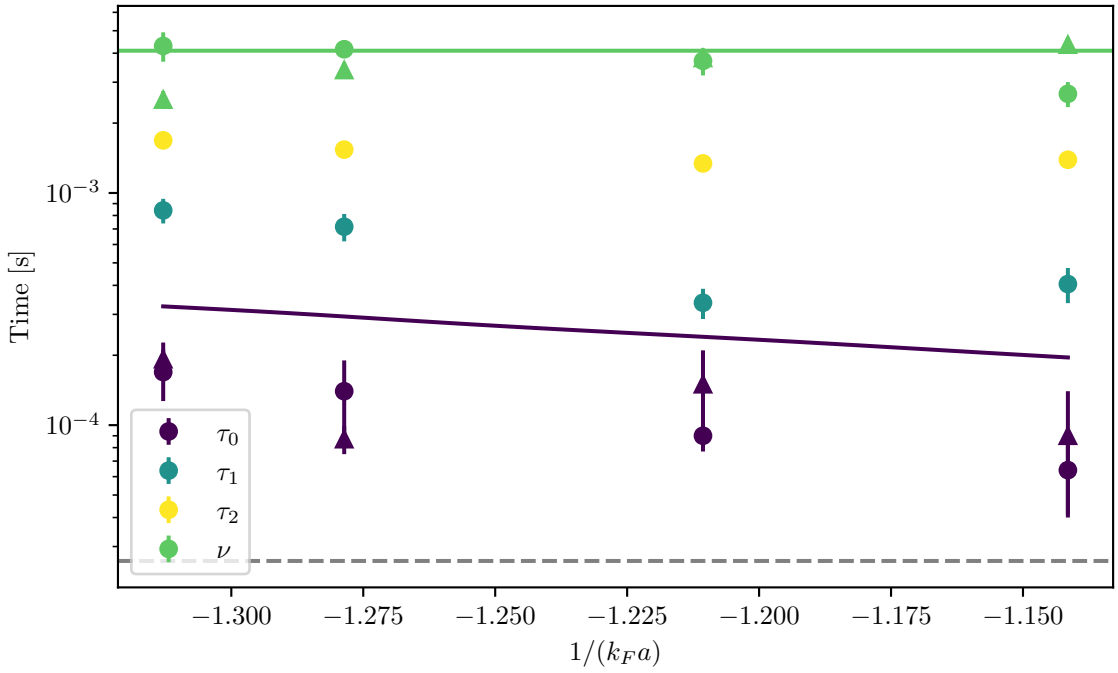

Figure 8.5: Time scales of time evolution after rf quench. The green line shows the calculated hydrodynamic frequency, which is constant for all interactions. The purple line shows the quasiparticle relaxation time. The grey dashed line indicates the duration of the rf-quench.

Figure 8.5 shows a summary of all detected timescales. For some interaction values two datapoints corresponding to the two observed spatial directions are shown. We find that the rf-quench excites the monopole mode in the harmonic trap, visible in the oscillations of the thermal and condensate cloud width. The condensate width shows a fast initial increase by approximately a factor of two. This is attributed to a redistribution of particles with a timescale that is in agreement with the quasi particle relaxation time. The condensate fraction shows a fast initial decay on a timescale comparable or larger than the quasi-particle relaxation time. After this initial decay, the condensate fraction recovers to a steady state. These observations are in agreement with a theoretical model for a one-dimensional system that fully models the dynamics of three internal levels, taking final state interactions into

account.

8.3 Fast magnetic field quenches

In the previous section we showed that an interaction quench through a fast state transfer with a duration in the order of the quasiparticle relaxation time can excite dynamics on several timescales. For one, it excites the monopole mode on the order of the trap time scale and secondly, a redistribution of quasi particles on the time-scale of the quasi particle relaxation time takes place. However, one thing that was not observed is the Higgs mode at a frequency of $2|\Delta|/\hbar$, whose existence was detected by rf modulation (see chapter 7) and is expected to manifest itself for quenches faster than the Fermi time [189]. The Fermi timescale is on the order of $\tau_F = 4.7\,\mu s$, which is fast compared to the duration of the rf-transfer $\tau_{rf} = 28(2)\,\mu s$. As the duration of the pulse scales with the square root of the power, another factor of ≈ 50 in power would be needed achieve $\tau_{rf} \approx \tau_F$. It is apparent, that without any modifications to the quench scheme in general, the Fermi time scale cannot be beaten by the rf-transfer method.

Another way to quench the interaction is by rapidly changing the magnetic field. Due to the presence of the Feshbach resonances, this will change the interaction parameter $1/(k_F a)$. In our case, the difficulty in this procedure is to do so fast enough. The fastest way to change the magnetic field within our setup is realised in the rapid ramp procedure described in chapter 4.3. Here, the time scale of the decay of the magnetic field is $\tau_{RR} = 86\,\mu s > 30\,\mu s$ and therefore not faster than the rf-transfer. The reason for this are eddy currents induced in the metal chamber surrounding the atoms and the mutual inductance of the large magnetic field coils. In order to do faster quenches on the order of the Fermi time, we had to develop a new ansatz for magnetic field quenches: We designed a small (as compared to the Feshbach coils) coil setup consisting of two coils with counterpropagating currents that was placed close to the atoms on the UKEA viewport on top of the chamber. By choosing the diameter and number of windings per coil accordingly, a magnetic field was designed that has a vanishing gradient, but a finite offset field at the atom position. Together with a fast switching circuit, we used this new coil setup to do fast magnetic field quenches. Additionally, the small size of the coil setup and therefore small magnetic flux, prevents both eddy currents and a large mutual and self inductance. A publication detailing the design of the coil setup and circuit is in preparation for publication [204].

In the following section, the design and performance of the new coil setup and switching circuit will be detailed. The alignment of the coil with respect to the atom position will be described in detail. Finally, there will be an outlook on the data taken so far with this quench coil setup.

8.3.1 Design and positioning

The idea behind the quench coil setup was to create a reliable way to change the magnetic field by tens of Gauss within a short time frame of less than $5\,\mu s$. Using the narrowest of the three Feshbach resonances (between states $|1\rangle + |3\rangle$), this would amount a reasonable change of the interaction parameter $1/(k_F a)$. To match the required timescale, we had to make sure that self- and mutual inductance as well as induced eddy currents in the metallic

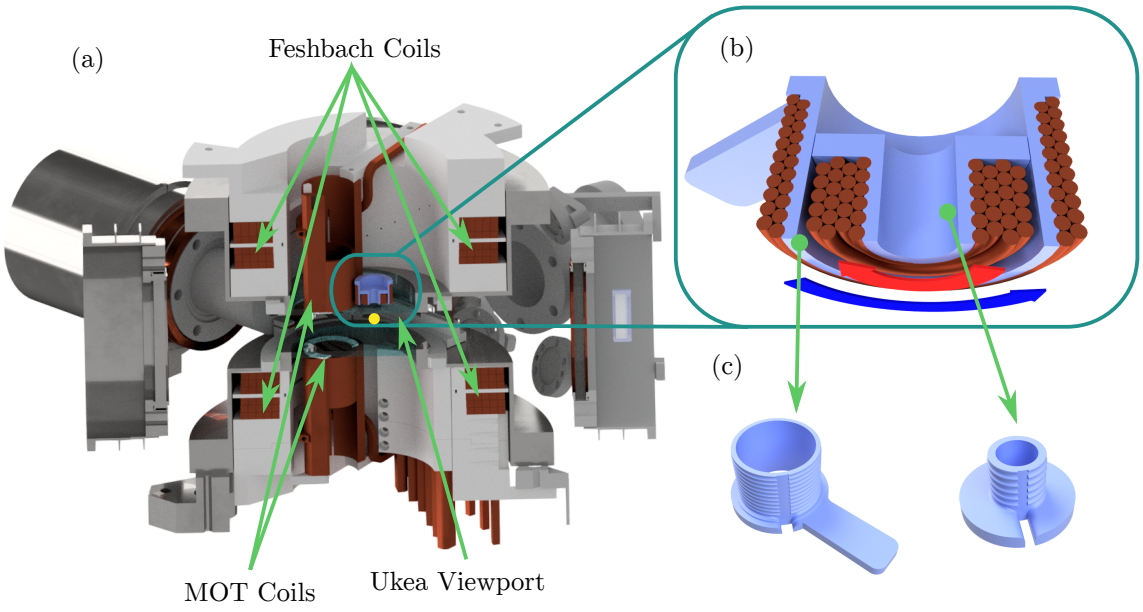

Figure 8.6: Details of quench coil design and placement within the experiment. (a) Positioning of the quench coil within the experimental setup. The coil is lowered to the UKEA viewport of the main chamber with a translation stage from the top. Indicated in yellow is the atom cloud position. (b) Detailed view of the quench coil. The inner layers and outer layers carry current in opposite directions. (c) Two assembly parts are used to wind the coils. The inner part (right) is slid into the outer part (left) and fixed with glue. Pictures of the final setup can be found in appendix A.3.

chamber were prevented. The easiest way to do so was to use a coil setup that is small as compared to the chamber and the Feshbach coils, as this would lower the mutual inductance and lower the magnetic field flux through the chamber walls. It would also enable the coil to be placed directly above the UKEA viewport, allowing for a close proximity to the setup and a higher field at the atom position. Another requirement for the setup was imposed by the trap depth of the dipole trap. If the coil setup would cause a large gradient, this would lower the depth of the trap in the corresponding direction. Small gradients can be compensated by counteracting compensation gradients or an increase in the trap depth in the corresponding direction, however this would potentially increase the trap frequencies and change the Fermi energy.

The final coil setup was chosen to consist of two coils of different size and winding number in series with counterpropagating currents (see figure 8.6). This design creates a magnetic field which has a zero crossing in the magnetic field gradient at a finite offset field at a position z_0 with respect to the edge of the coil setup. Choosing the number of windings and winding layers for each coil and the relative position of the coils accordingly, the position of this point and the offset field can be selected. Placing the coil on the UKEA viewport

dictates a distance of ≈ 8 mm from the lower end of the coil setup to the atom cloud. The position of the coil setup with respect to the experiment chamber is shown in figure 8.6 and a picture of the finished coil setup is shown in appendix A.3.

8.3.2 Magnetic field

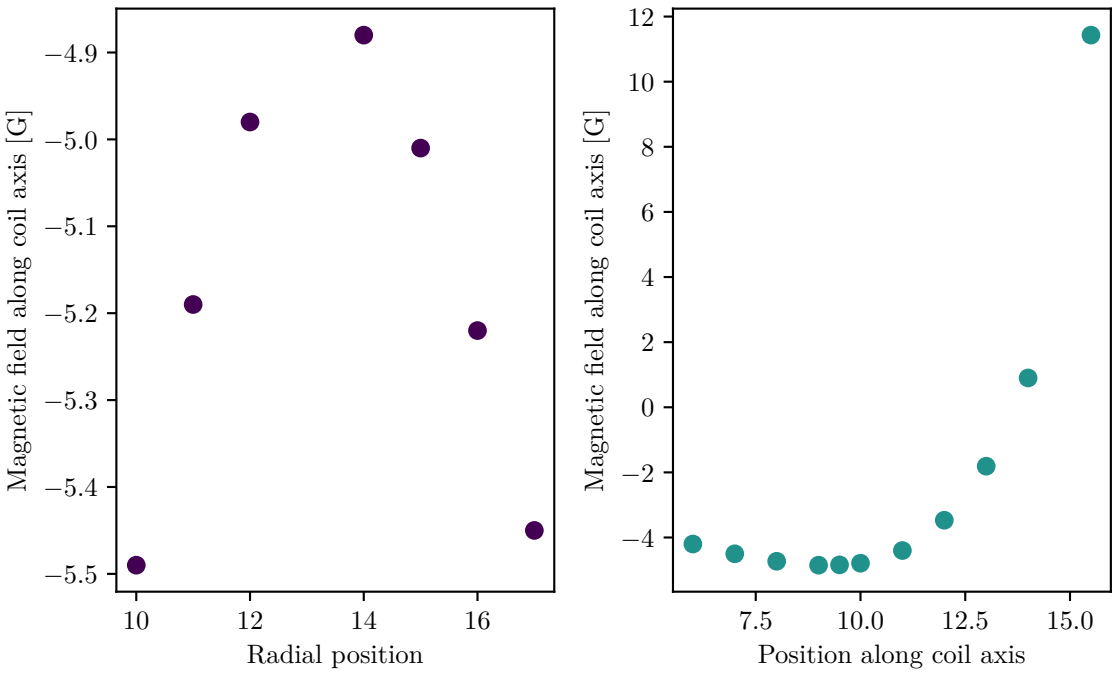

Figure 8.7: Magnetic field distribution of the quench coil. Magnetic field of the quench coil along the coil axis at 5 A for varying position in radial (left) and longitudinal direction (right) measured with a Hall probe. As expected, the magnetic field has a non-zero minimum along the coil axis. Note, that the datapoint at 16 mm is the closest the Hall probe sensor can go towards the coil. This results in an approximate distance $z_0 = 9(2)$ mm.

The magnetic field of different combinations of winding numbers and winding layers was calculated by simulating the Biot-Savart law for the corresponding coil setup. In the end, we chose the coil setup indicated in figure 8.6, consisting of 22 windings, arranged in 2 layers for the outer coil with an inner diameter of 22 mm and 22 windings, arranged in 4 layers for the inner coil with an inner diameter of 6 mm. This design yields $z_0 \approx 9$ mm, which gives 1 mm distance to the UKEA viewport, allowing for some leeway when aligning the coil and preventing the coil touching the UKEA viewport and scratching or heating the glass.

We designed the rigid body parts used to wind the coil in Autodesk Inventor and subsequently 3D printed the parts with the fused filament fabrication method. The used material for the body is polylactide, which is cheap and easy to print, allowing to manufacture several different design iterations. It is furthermore reasonably sturdy and stable to deformation to up to 65 °C [220], which is higher than the limit temperature that we felt comfortable putting

on top of the viewport. The body consists of two independent parts, each corresponding to one of the coils (see figure 8.6 (b) and (c)). Each part has groves on the outer side, allowing for easy and equidistant winding. The coils are wound with 1 mm diameter polyimide coated copper wire. After winding the coils separately, the outer layers are coated with epoxy to prevent movement of the windings and increase stability. Afterwards, the small coil is slid into the center of the larger mount and again fixed with epoxy. Additionally, two temperature sensors[3] are glued to the body to monitor the temperature and prevent overheating with an interlock system. The outer part is connected to a plastic post, which lowers the coil onto the experiment chamber from a 3D translation stage[4] placed on the breadboard above the chamber.

The magnetic field of the coil was tested in constant current mode for a low current of 5 A with a Hall probe. The results can be seen in figure 8.7. The left panel shows the magnetic field along the coil axis for a displacement in the radial direction. As expected, the magnitude of the field shows a minimum on the coil axis. Variation along the coil axis shows that for a distance of $z_0 = 9(2)\,\mathrm{mm}$ the field has a zero-gradient, non-zero offset field extreme value. In the end, the goal is to place the atoms in this distance on the coil axis. As the maximum field is proportional to the current in the coils, the maximum field difference for the quenches can be calculated via

$$\Delta B_{\mathrm{final}} \approx \frac{I_{\mathrm{set}}}{5\,\mathrm{A}} \cdot 4.5\,\mathrm{G}. \tag{8.7}$$

In the experiment we use a 40 A power supply[5], resulting in a maximum field change of $\Delta B_{\mathrm{final}} \approx 36\,\mathrm{G}$. The total resistance of the coil is $0.1\,\Omega$ and the total inductance was calculated to be 6 µH. The range of possible quenches with this coil with respect to the phase diagram calculated in ref [43] is shown in figure 8.8.

8.3.3 Control circuit

The control circuit for a fast turnoff of the coil was designed and assembled by Dr. Akos Hoffmann in our group and is shown in figure 8.9. It consists of two separate sections. For one, the current switch and secondly, the switch signal processor.

Switching off the current through the coil is done by a high-speed power MOSFET driver (Microchip Technology MCP1407) in combination with a power MOSFET (IXYS IXFN140N20P). The high peak output current of 6 A of the driver allows switching times down to 30 ns. The drain-source on-state resistance of the power MOSFET is less than $18\,\mathrm{m}\Omega$, resulting in a total DC resistance of less than $50\,\mathrm{m}\Omega$ for coil, wiring and MOSFET. The high drain-source breakdown voltage of 200 V allows high inductive voltage spikes, even though a varistor (EPCOS Inc S20K17) is used. When the current through the power supply is cut off by the MOSFET, the current in the coil is driven to zero by dissipating the magnetic field energy with an AC coupled RC snubber circuit. This circuit consists of a carbon composition

[3] PT100
[4] Thorlabs PT3/M
[5] Delta Electronics S280

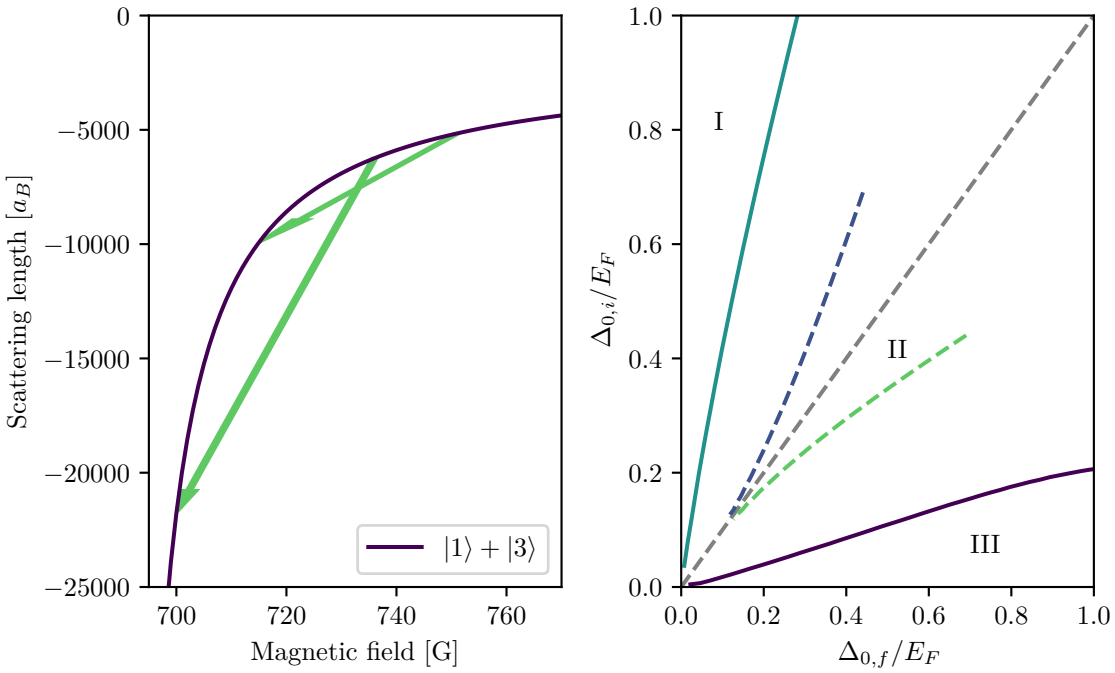

Figure 8.8: Range of possible magnetic field quenches. (Left) A change of $\Delta B_{\text{final}} = 36\,\text{G}$ towards lower magnetic field constitutes to a variable change of interaction strength due to the shape of the Feshbach resonance. The green arrows show two quenches originating from a different initial magnetic field. **(Right)** The range of possible magnetic field quenches within the phase diagram of ref [43]. Aligning the magnetic field of the quench coil parallel to the Feshbach field yields weak to strong quenches with $\Delta_i < \Delta_f$ (green dashed line), aligning the quench coil magnetic field anti-parallel yields quenches with $\Delta_i > \Delta_f$ (blue dashed line). Changing the current through the quench coil enables us to realise all quenches between both curves.

resistor with no inductance (RC07GF100J) and three parallel 200 V capacitors (SMD 1812). The components for the snubber circuit where optimised for fast switching by simulating the switching behaviour with LTSPICE. In order to not compromise the switching behaviour, the leads from the circuit to the coil have to be kept as short as possible. We achieved this by placing the circuit in an insulated box directly on the breadboard above the upper viewport, next to the translation stage that controls the quench coil position. This keeps the distance between circuit and coil at less than 15 cm. The power cables are directly connected to the MOSFET.

In order to process the TTL signal coming from the digital cards of the experiment control, we use a switch signal processor. It consists of an fast optical coupler (ON Semiconductor HCPL-0601), a single Schmitt-Trigger inverter (Texas Instruments SN74LVC1G14) and a single retriggerable monostable multivibrator (Texas Instruments SN74LVC1G23). This design combines the external switch signal and the internal timing signal to switch and protect the coil. To protect the coil against overheating, additionally to the temperature interlock, the maximum time of high output of the multivibrator is determined by capacitor

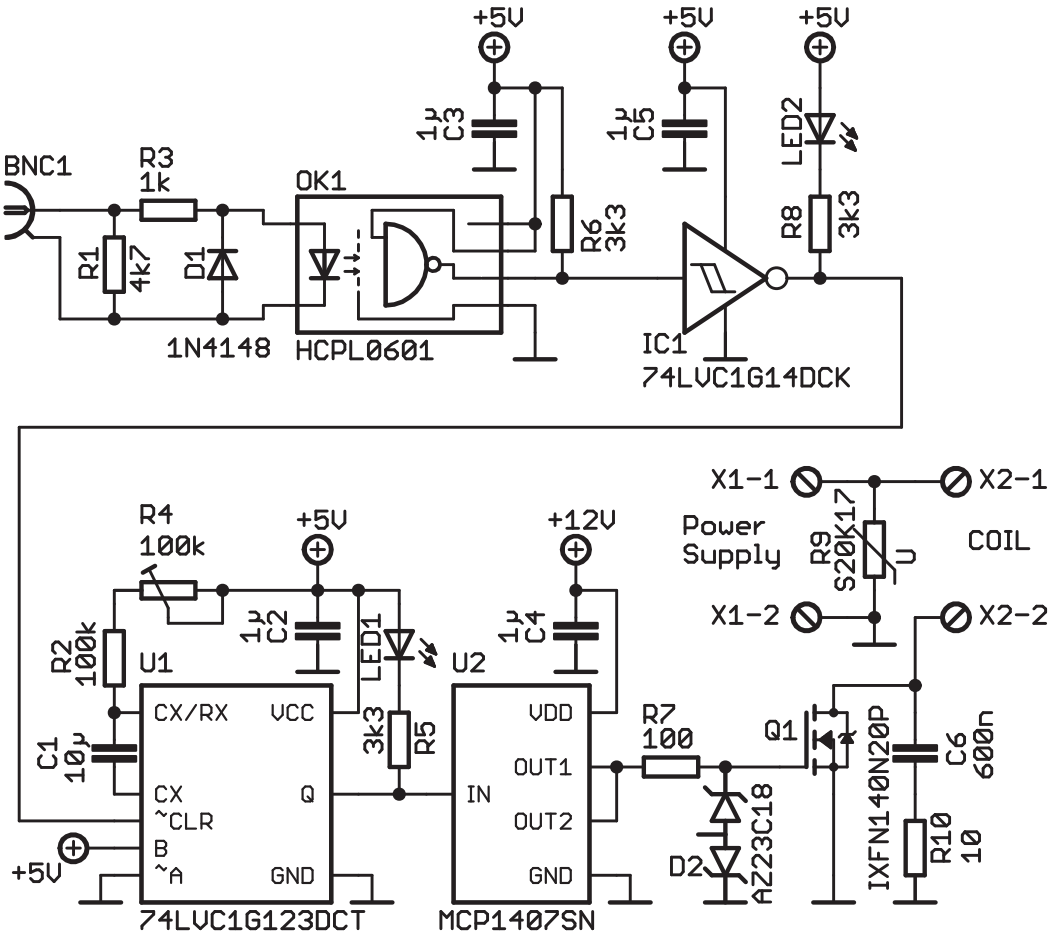

Figure 8.9: Quench Coil control circuit. The circuit consists of two basic parts, a switching circuit and a switch signal processor.

C1 between the CEXT and REXT/CEXT pins and by the set value of the potentiometer in the voltage divider connecting REXT/CEXT with the V_{CC} pin. In our case, we set the potentiometer such that the maximum on-time is 1 s, which is long as compared to the time the quench coil is needed in the sequence but still short as compared to the full duration of the sequence of about 30 s. This will prevent any damage to the coil and chamber in case the temperature interlock fails.

8.3.4 Switching performance

The performance of quench coil and circuit were tested with a pickup coil. The pickup coil is a simple 15 mm × 15 mm square coil with two windings and was connected to an oscilloscope. Having the pickup coil close to the quench coil when switching, induces a change in magnetic

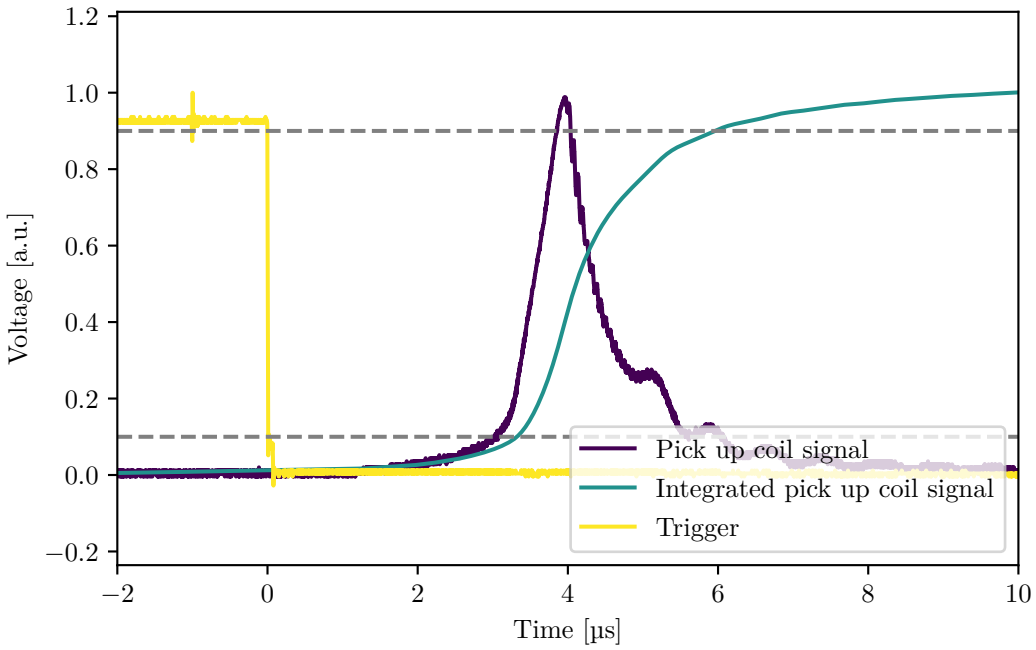

Figure 8.10: Bench test for quench coil switching time. The purple curve shows the voltage signal induced in a pickup coil close to the quench coil, when switching the quench coil off. The integrated signal (turquoise) is proportional to the magnetic field, the dashed grey lines show the 10 % and 90 % values of the magnetic field. The yellow curve is the TTL Trigger used to switch the magnetic field.

flux through the pickup coil loops and a counteracting induced voltage, that can be measured with an oscilloscope. The result is shown in figure 8.10. The yellow signal is the TTL Trigger that causes the turnoff and the purple signal is proportional to the voltage induced in the pickup coil. The magnetic field generated by the quench coil is then proportional to the integral over the pickup coil signal, shown in turquoise. From this, it can be seen, that the 10 % to 90 % rising time is $\approx 3\,\mu s$. The measurement was repeated in the vicinity of different metallic objects and at the final quench coil position above the chamber, however, the performance did not degrade visibly. With a quench time of 3 µs as compared to a Fermi time of $\tau_F = 4.7\,\mu s$ fast quenches are possible.

8.3.5 Alignment to the atom position

Having the coil tested and verified its performance, the next step was to put it into the experiment. For this, the coil was fixed to a plastic post and connected to a 3D translation stage on top of the breadboard above the chamber. By sending the imaging beam (which naturally hits the atoms) from below the chamber upwards, we achieved a coarse initial alignment of the coil to the atom position. We chose the initial setting of the translation stage such that for a maximum setting of $z = 20\,mm$ the coil touches the upper window.

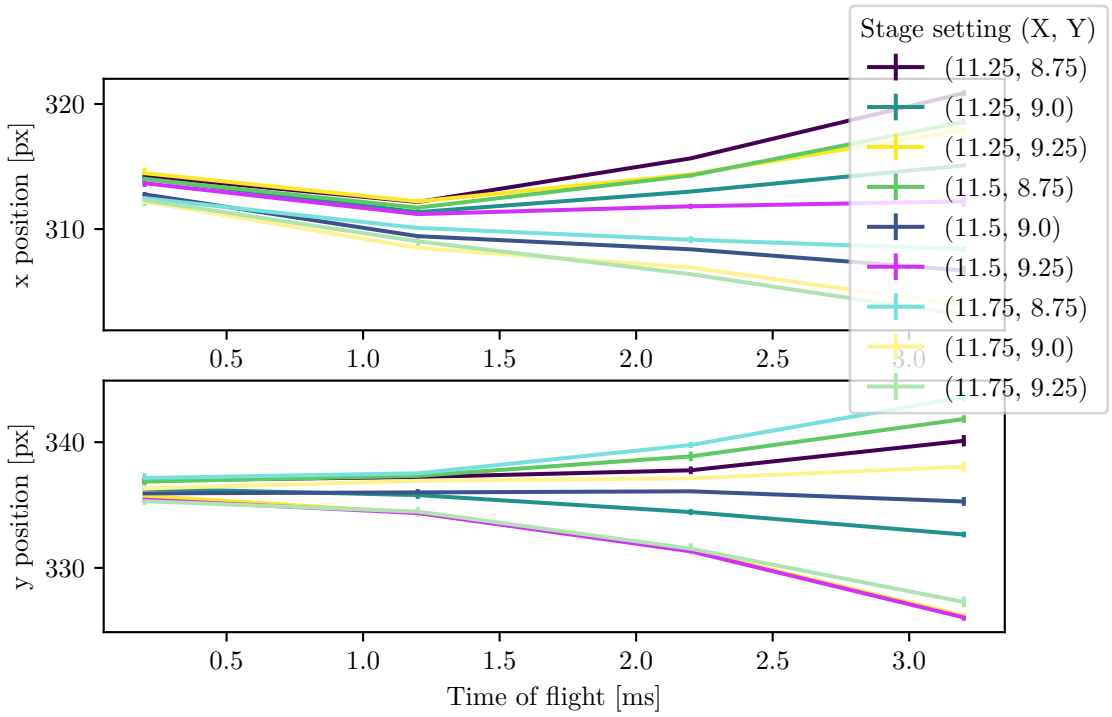

Figure 8.11: Quench coil alignment along the x (top) and y (bottom) direction. The translation stage is set to $z = 19.25\,\mathrm{mm}$. Misalignment along the radial direction will accelerate the atoms outwards. The initial tilt of the curves in the upper panel can be explained by a misalignment of the imaging axis with respect to the gravitational axis.

Consequently, in order to not scratch the window, the coil has to be moved upwards towards decreasing z values for every movement in radial direction. The alignment in x and y direction was done by monitoring the movement of the cloud in time of flight. As misalignment in radial direction creates a positive gradient (see figure 8.7) and the atoms are high-field seekers, the atoms will move further, the larger the misalignment is. By minimising the movement along x and y, when imaging from below, the radial alignment was done. The results are shown in figure 8.11 for different settings of the translation stage.

The alignment in z direction was more complicated, as the method of using the displacement is complicated by the movement of the atoms by the gravitational force and the anti-confinement of the Feshbach field in z-direction. It was furthermore important to get the most accurate reading along z direction, as after evaporation this direction has the lowest trap depth. To get an accurate measure of the magnetic field and gradient, we performed radio-frequency spectroscopy [221–223] on the $|2\rangle \rightarrow |3\rangle$ transition, which is very sensitive to small changes in the magnetic field. In the left panel of figure 8.12, several radio-frequency spectra for different z alignments are shown. The decrease in atom number for some spectra is caused by a strong gradient due to misalignment, effectively ejecting atoms from the trap. Extracting the center positions by applying lorentzian fits and recalculating frequency into magnetic fields yields the turquoise datapoints in the right panel of the same figure. By fitting a quadratic function to the datapoints, we estimate the position of the maximum to be

at a translation stage setting of $\tilde{z}_0 = 19.97(11)\,\mathrm{mm}$. The resulting linear gradient is shown in green. In principle, the alignment procedure in radial and magnetic field axis directions have to be iterated, however, we found that this was not necessary. The position of the minimum with respect to the atoms is surprising, as from our calculations, it should have been around 19 mm. This could have several reasons: For one, the atoms could be further away from the upper window as expected. This was never measured in detail before and could date back to small initial misalignments of the dipole trap. A second reason could be a slight tilt of the coil, which presents an angle to the coil axis relative to the viewport. In this case, the distance between the lower end of the coil to the atoms projected onto the normal axis of the window is reduced. As scratching and damage to the UKEA viewport had to be prevented by all means, we set the quench coil to a final value of $\tilde{z} = 19.75\,\mathrm{mm}$. This introduces a small gradient to the atoms, however, atom loss can be prevented by adjusting the power of the dipole beams accordingly or by compensating the gradient with the magic layer of the Feshbach coils.

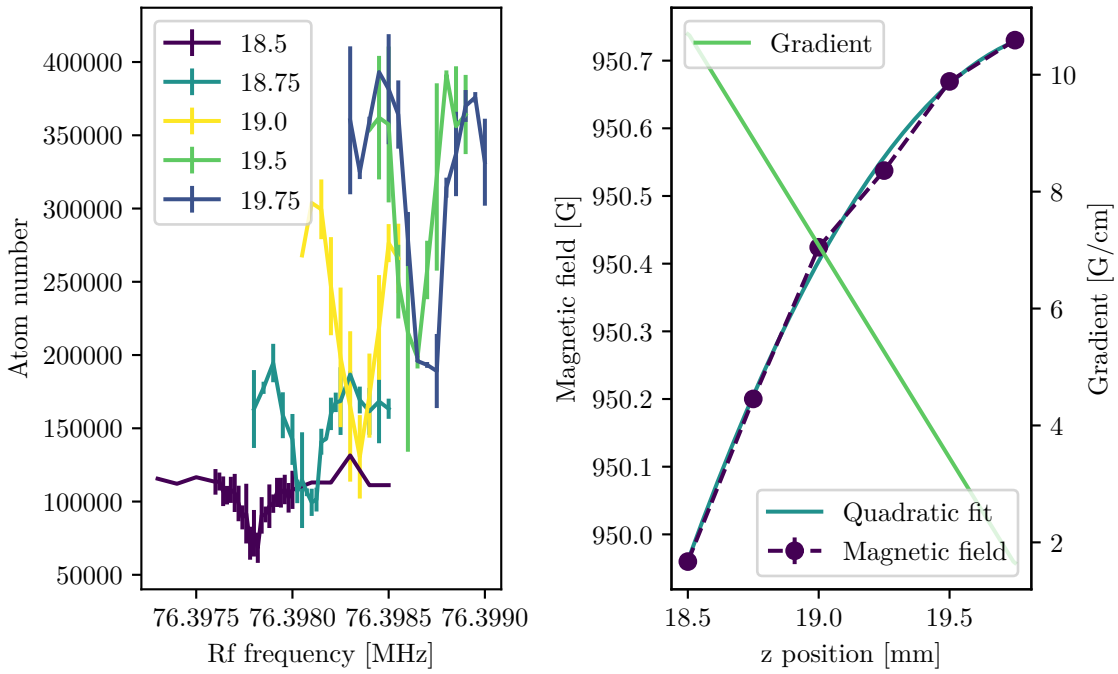

Figure 8.12: Alignment of the quench coil with rf-spectroscopy. (Left) Radio-frequency spectra for different quench coil alignments along the z-direction. The position of the dip indicates the resonance condition. The total number of atoms per curve changes, because with stronger misalignment the resulting gradient ejects atoms from the dipole trap. **(Right)** Resonance position (turquoise dots) in Terms of the magnetic field against quench coil alignment. A quadratic fit (purple curve) is used to determine the maximum position, which is predicted to be at $z = 19.97\,\mathrm{mm}$. Also shown is the gradient (green line) determined from the quadratic fit. Resulting error bars from the resonance fit are shown, but not visible. This measurement has been performed with a quench coil current of 25 A, as compared to the final 40 A to decrease the atom loss. Accordingly, the relative field (minus offset) and gradient have to be scaled.

8.3.6 Time domain dynamics

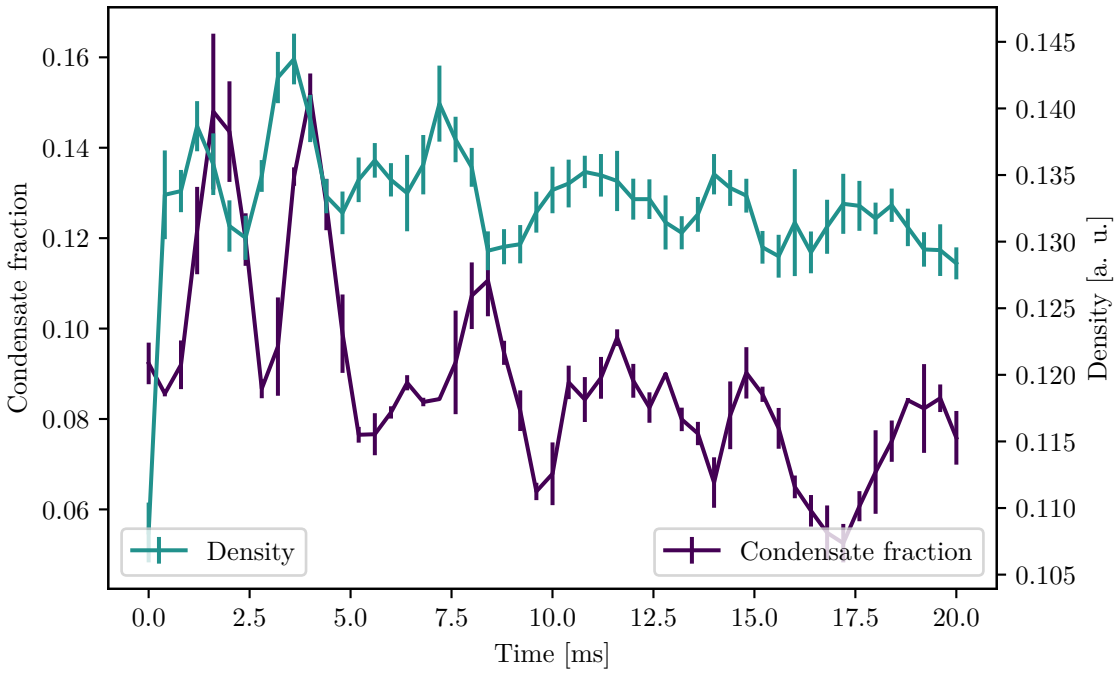

Figure 8.13: Time evolution of the condensate fraction and the density after a rapid quench with the quench coil. This curve is taken close to unitarity at $1/(k_F a) \approx -0.06$. A delay between the characteristic features present in both curves can be seen.

With the quench coil in place and ready to be used with the atoms, it is possible to observe the time dynamics after a rapid magnetic field quench. Here, we observe the dynamics of a quench from weakly interacting to strongly interacting $1/(k_F a_i) > 1/(k_F a_f)$, meaning the magnetic field generated by the quench coil is parallel to the magnetic field created by the Feshbach coils. We prepare a sample of $N \approx 2 \times 10^5$ ^{6}Li atoms in states $|1\rangle$ and $|3\rangle$ at a temperature of $T/T_F \approx 0.06(2)$ before ramping up the coil. This is different to the other experiments in this thesis, where mostly states $|1\rangle$ and $|2\rangle$ are used. The reason for this difference are the favorable properties of the $|1\rangle + |3\rangle$ Feshbach resonance. Namely, the width of this resonance is narrower allowing a potentially larger quench with the difference in magnetic field the quench coil can produce. The coil is used at the maximum possible current of 40 A given the maximum rating of the power supply and the limitations of the switching circuit. This generates an additional magnetic field of $\Delta B_{\text{quench}} \approx 36\,\text{G}$. After evaporation and adiabatic sweep to the magnetic field B_f corresponding to the final interaction strength $1/(k_F a_f)$, the coil current is ramped up within 1 s. Afterwards, the cloud is held for an additional 10 ms at the initial field $B_i = B_f + \Delta B_{\text{quench}}$. After the coil is suddenly turned off for the quench, the cloud is subjected to a variable hold time in the dipole trap. To acquire the condensate fraction, we perform a rapid ramp with a subsequent bimodal fit to the momentum distribution. Additionally, we estimate the density by taking the in-trap cloud

size with the ANDOR0 and ALTA1 imaging systems to cover all three spatial directions. The time evolution of condensate fraction and density is shown in figure 8.13. The evolution of both quantities takes the form of an oscillation-like behaviour with changing frequency and a possible decay. Interestingly, by checking the corresponding features in both curves, it seems there is a time delay between the quantities, with the condensate fraction being delayed with respect to the density. By fitting several peaks corresponding to features in both curves and comparing the center position, it is possible to quantify the time delay. The results of this analysis show a dependence on the interaction strength $1/(k_F a_f)$ and are shown in figure 8.14. To quantify if the delay is dependent on the strength of the quench and / or the final (or initial) interaction value, we perform two sets of two quenches that have the same quench strength, but different final interactions $1/(k_F a_f)$, shown in figure 8.15. It can be seen that the delay depends on the exact interaction strength. For subsequent investigations, the exact behaviour of the delay time can be measured and might yield new information about the formation dynamics and timescales of the superfluid.

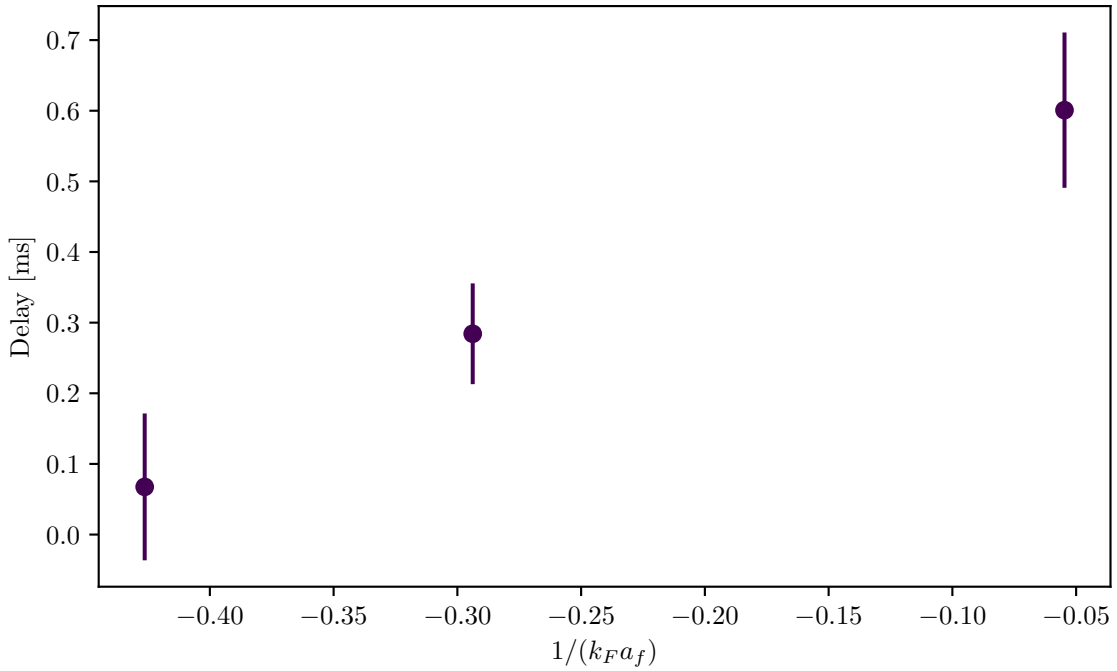

Figure 8.14: Comparison of the averaged delay time for quenches at different interactions. The delay is given by the average of the delay of the first four peaks for each curve and shows a dependency on the final interaction strength. All quenches were performed with maximum ΔB_{quench}.

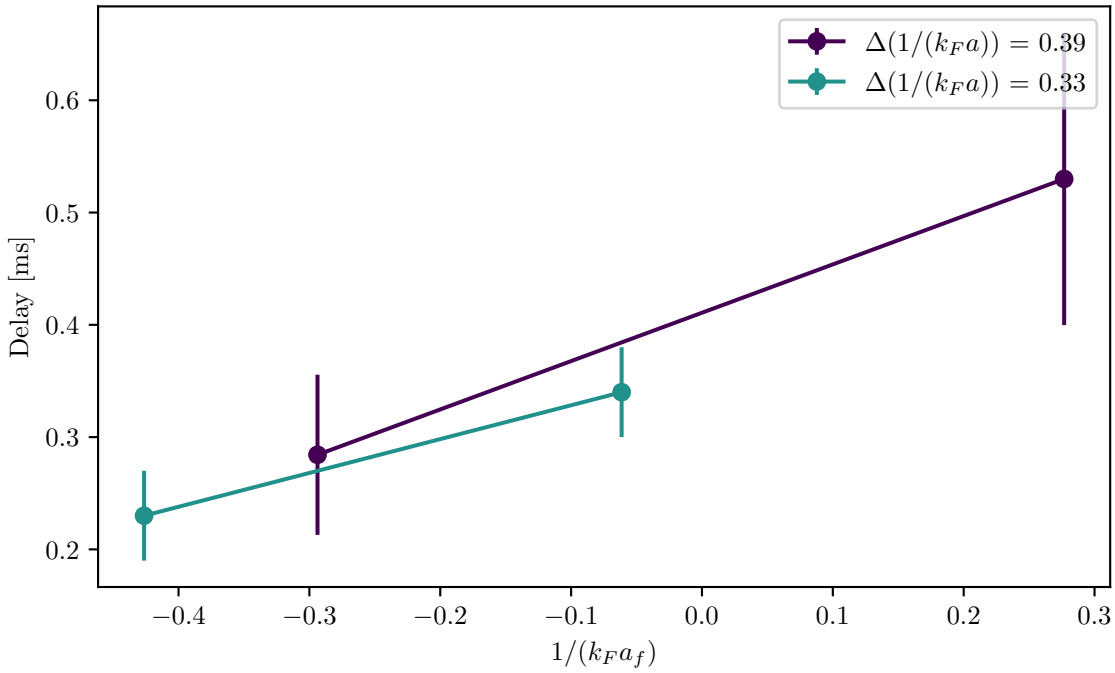

Figure 8.15: Evolution of the delay time for equal quench magnitudes. The magnitude of the quench $\Delta(k_F a)^{-1}$ is held constant, while $1/(k_F a_f)$ is varied. The different final interactions are achieved by having lower currents and therefore lower ΔB_{quench} closer to the resonance. The turquoise datapoints are analysed by finding a peak within the cross-correlation, as no reliable fitting of peaks in both curves was possible.

8.3.7 Results and outlook

In this section, we developed a coil system for fast changes of the magnetic field. We use this coil system to perform interaction quenches on our strongly interacting Fermi gas system. We showed that the coil setup is able to beat all relevant time scales in the Fermi gas. In the last section, we gave an overview over the long time dynamics after such a quench, showing an interaction dependent time delay that could potentially be connected to the energy scales relevant in the formation of a superfluid.

The analysis in the previous section is mostly concerned with the long-time limit of the evolution. However, observing the Higgs mode will require much shorter timescales. Looking at the results of chapter 7, the Higgs mode is expected to appear at a frequency of $\nu_{\text{Higgs}} = 2\Delta/h \approx 20\,\text{kHz}$ corresponding to an oscillation period of $T_{\text{Higgs}} \approx 50\,\mu\text{s}$. So far, within these short times, we did not observe oscillations of the condensate fraction. This can, however, have several reasons which were not possible to rule out within the scope of this thesis and require further investigation. One reason why no oscillation of the condensate fraction can be seen could be a strong decrease of the oscillation amplitude due to the second ramp that is introduced by the rapid ramp to the BEC side, where the expected oscillation amplitude is strongly damped and therefore possibly hidden in the noise [189]. One way to

circumvent this problem would be to use a neural network to detect the condensate fraction without rapid ramp, as introduced in section 5. Investigating the short time dynamics of the condensate fraction more accurately therefore presents a route forward for the possible direct detection of the Higgs mode without help of spectroscopic methods that are insensitive to the coherent dynamics of the mode.

CHAPTER 9

Machine Learning for compensation tasks

This chapter of the thesis is special because it does not show an experiment conducted on the Na-Li apparatus described in the previous chapters. It rather introduces a very general way to use machine learning algorithms to speed up the process of compensation or optimisation tasks very common to physics experiments. The procedure used to achieve this will firstly be introduced through a general formulation of the compensation problem. Secondly, it will be demonstrated on an experimental system, namely stray field compensation on a single, trapped ytterbium ion. Naturally, the details and operation of this system are not within the scope of this thesis, but are part of the work of Pascal Kobel. His PhD thesis [224] will contain a detailed description of the experimental side and the implementation of this work. While trying to achieve a self-contained short version of the experimental details, this chapter will focus more on the conceptual part of the presented work. The contents of this chapter are currently in review for publication with a preprint being available on arXiv [3].

9.1 Compensating disturbance

This section aims to introduce the general problem of compensation or optimisation tasks (which from here on will be summarised by just compensation tasks). It will then present the general procedure for solving the given task, which usually is time-consuming. In order to improve the situation and reduce the number of required measurements, we will introduce the idea of an empirical model of the experimental system.

9.1.1 General problem description

In the context of physics experiments, the careful characterisation and compensation of external disturbances is increasingly important. For atomic clocks, quantum computers or cold atom experiments, external disturbances can for example lead to a decrease of quantum coherence and can therefore impact the performance of the system drastically. In a similar way, even for completely different kinds of physics, optimisation of parameters to compensate for disturbances are important, one example being the optimisation of beam parameters in particle physics [225]. The general problem is illustrated in figure 9.1 (a): The experimental system (purple circle) is exposed to a set of noise sources N_i (green) originating from the

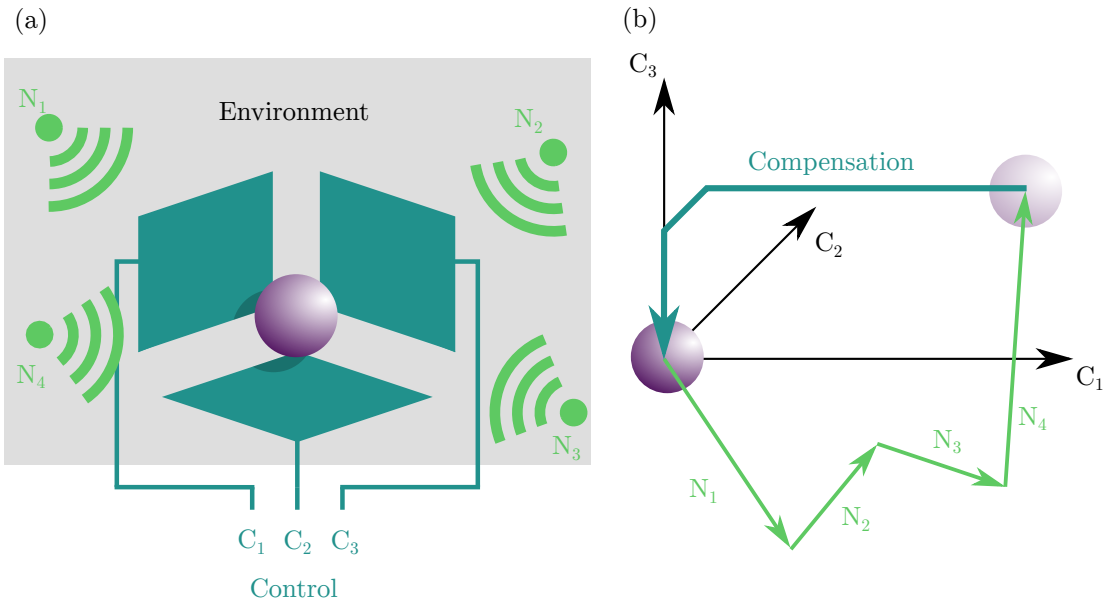

Figure 9.1: General problem description of a compensation task. (a) Real space representation of an experimental system (purple), being exposed to noise sources N_i (green) from the environment. The noise sources can be compensated by means of control parameters C_i (turquoise). **(b)** The same situation as shown in panel (a), but in parameter space, spanned by the control parameters. The goal of compensation is to counteract the displacement the system experiences through the noise sources. Color coding is equivalent to (a).

environment. The number and strength of the noise sources is arbitrary, as is their exact position, i.e. there is no known cross-correlation between them. The goal is to compensate the collective effect of all noise sources by means of tuning control parameters C_i (turquoise). Generally, the number of control parameters does not match the number of noise sources, but should at least be equal to the dimensionality of the system under investigation (e.g. for a spatial compensation one needs three linearly independent controls, or for steering the angle of a laser beam one needs two linearly independent controls). The most general way to look at this problem is by means of the parameter space spanned by the controls C_i, see figure 9.1 (b). Here, the combined effect of the noise sources (green) represents a displacement originating from adding the displacement vector of every single noise source. The goal of the compensation is to apply control parameters that compensate the noise displacement and push it back towards its optimal position. As the space is spanned by the unit vectors of the control parameters, this task is equal to finding a representation of the displacement vector in terms of the control vectors. This is a very general description, as noise sources come in a plethora of different kinds, such as stray magnetic or electric fields, acceleration or movement of the apparatus or thermal drifts. All of these sources need different means of compensation, but can be formulated in the same way.

9.1.2 Finding the compensation point

Finding a representation of the noise displacement vector by means of the control parameters is not an easy task, because the noise vector is not known a priori. To make things worse, generally, the displacement vector changes over time, which becomes especially cumbersome if the time scale of the temporal deviation of the noise sources is comparable to the time scale needed to find the correct compensation. It is apparent that an efficient and reliable way to compensate the noise sources is of high importance. One way to find the parameters needed to compensate for the noise sources, is by trial and error: Scanning the parameter space and monitoring the performance of the system will yield the compensation parameters, as long as the compensation point is within the scanned region of parameter space. Scanning the parameters, say, on a hypercubic grid in parameter space, the number of data points required for compensation scales as n^k for n datapoints per control parameter and k control parameters, i.e. k is the dimension of the underlying parameter space. Sometimes, the number of datapoints required for compensation can be reduced by an underlying (linear) model of the situation. Such a model might decouple the control parameters for finding the compensation point, such that a parameter scan might have to be done only once per direction in order to find the optimal point. The number of points then scales with $n \cdot k$. In the following sections, we measure the performance of different models with respect to this number. The number of measurements then might be written as $a_{\text{model}} \cdot n \cdot k$, where the additional factor a_{model} characterises how well the model performs as compared to the naive assumption of n datapoints per dimension.
A second way to find the compensation point of a given system that does not rely on scanning parameter space for every compensation is to model the system with all noise sources. This would enable an accurate prediction of deviations from the optimal state due to external deviations and the parameters required for compensation. Simulating such a model, however, is very complicated. To be precise, one needs knowledge of all fringe effects and the exact characterisation of all present noise sources. This is made more difficult by any potential time evolution of noise sources that has to be accounted for. Weighing up gain and effort of an universally applicable model, usually unbiased systematic search for an optimal setting within the parameter space is preferred.
Another way to look at the two options for compensation is the following: For systematic search, the initial effort of preparing the measurement is vanishing. However, at runtime (=during the compensation process), the inference takes as long as taking the required datapoints for compensation. For the universally applicable model, the initial effort is very high, as a realistic model of the experimental system and noise sources has to be created. However, once this is done, at runtime, the inference cost is vanishingly low, as the model *predicts* the compensation from its knowledge base. Our ansatz is to combine the best of both worlds: From the data taken during systematic search, we create an empirical model of the experiment system. This model can then be used to infer the compensation point with less datapoints than needed for a full search by accessing its knowledge base. Indeed, we will show that the amount of data required for building such an empirical model is low, and that the accuracy in our case is limited only by measurement error and not by the availability of data.

9.2 Creating an empirical model of the target system

This section investigates the functionality of an empirical model for the experiment system, as describe above. Afterwards, it presents two different ways of creating the empirical model that accelerates inference for generic compensation tasks. Firstly, we use principal component analysis (PCA) as introduced in section 3.5.1 to improve the performance of an existing linear model. Secondly, we present a solution based on a small fully connected neural network (see section 3.3.2). This solution does not require any underlying model and can therefore be used in cases, where the first method is not applicable. Furthermore, by changing the architecture of the neural network, the accuracy of the network predictions can be adjusted to decrease the needed data even further.

9.2.1 What does an empirical model learn?

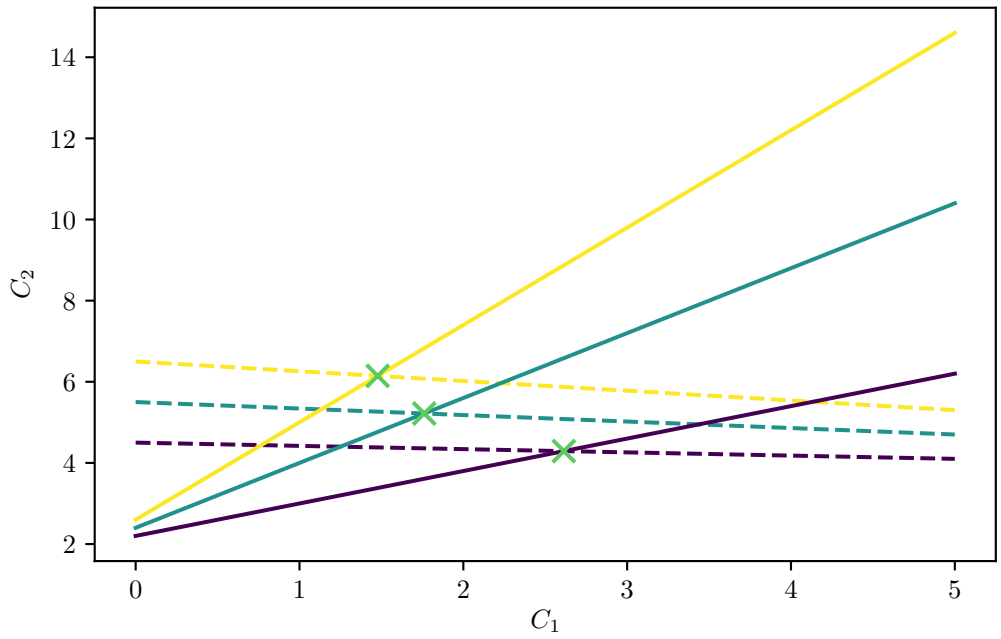

Figure 9.2: The compensation measurement in parameter space. Different colors indicate different noise characteristics. The dashed lines represent the optimal values of C_1, the solid lines represent the optimal values for C_2. The compensation points are indicated by the green crosses. Note, that in general the curves do not have to be straight lines, but could be more complicated.

To understand how the empirical model works, it is necessary to investigate the process of a compensation measurement a bit further. For simplicity, let $k = 2$ with C_1 and C_2 as control parameters. For unbiased systematic search, both control parameters are varied independently of each other, meaning, for every given value of C_1 it is possible to find an optimal value of C_2. Note, that C_1 and C_2 are in general not independent. This means the

optimal value of C_2 depends on the given value of C_1. The goal is to find the combination of C_1 and C_2 which represents a global optimal point.
What happens if the noise characteristic changes? Basically, the curves that represents all optimal values for C_1 and C_2 shift (and maybe deform) and so does the compensation point, which is represented by the crossing point of the optimal curves of both parameters in parameter space. The goal of the empirical model is to learn how this shift behaves - and predict the compensation point from the shape of one or more curves. The situation is illustrated in figure 9.2. Indicated are three compensation measurements at different times with different noise characteristics (indicated by different colors). The optimal curve for C_1 (dashed) and C_2 (solid) shifts for different times. The optimal compensation point is the point where curves of the same color intersect. We can now investigate the different approaches:

- For systematic unbiased search every combination of C_1 and C_2 is tested until the minimum point is found. This results in n^k datapoints.

- For a linear model, n points per curve are measured and fitted. From the extrapolation of the fit, it is possible to determine the compensation point. This results in $n \cdot k$ points total.

- An empirical model based on the linear model is able to use information from earlier measurements, to extrapolate the shape of the curves and predict the compensation point. This can for example mean that by learning from earlier data, the network is able to predict the shape of the optimal curve by just taking one datapoint per curve as an input. Such an improvement can be achieved by using principal component analysis on a representation of solutions from the linear model.

- Even stronger empirical models might be able to use less than n points on just a subset of curves by learning correlations between the shifting of the curves. This might result in less than n datapoints in total. This can be achieved through a neural network.

9.2.2 Principal component analysis for model improvement

The first ansatz for creating an empirical model of the target system is utilising the dimensionality reduction capabilities of principal component analysis, as introduced in section 3.5.1. For this, it is necessary to have a model of the system characterising the optimal curves. In case of the above example, this would be the assumption, that all optimal points for C_i form a line. PCA is then used to find common generalities between all possible solutions. Using only a subset of principal components, the number of possible solutions is reduced - requiring less probing to find the best solution from this subset and effectively reducing the dimension of parameter space. By taking into account only the strongest principal component, the subset of solutions is exactly those solutions, that can explain most of the underlying dataset. Using PCA makes use of the hidden correlations within the experimental system, by modeling the dependency of the compensation point on the control parameters. If all variance is along one direction in parameter space, PCA is able to parameterise the dependency through a single parameter, and therefore reduces the dimension of the problem down to 1, i.e. $a_{\mathrm{PCA}} = 1/k$ and therefore $a_{\mathrm{PCA}} \cdot n \cdot k = n$. This means in the optimal limit of fully correlated dependency,

the number of measurements needed to find the full compensation point is equal to the number of measurements needed to compensate in direction of one control parameter.

9.2.3 An artificial neural network as model

PCA for model improvement can already produce an impressive theoretical reduction in measurements by a factor of $1/k$. However, we want to go a step further by eliminating the need for a linear model and possibly reduce the number of required measurements even further. Neural networks, as presented in section 3.3.2, are a perfect candidate for this. They can act as a universal function approximator [93, 94] and in order to build them via supervised learning only require sufficient input and output data. By tuning the number of input neurons before training and evaluating the performance of the network, it is furthermore possible to achieve scaling between accuracy and number of required measurements. Without further careful analysis, the network is a black box, that represents the correlations introduced by the experimental system and, given the corresponding reduced input, will produce the compensation point as an output.
The architecture we propose is the following: As input layer take the amount of desired datapoints, e.g. one point of compensation for any direction C_i, resulting in k k-dimensional points in total. The output is a k-dimensional vector representing the compensation point for every C_i. The exact architecture in between the input and output layer is in principle arbitrary and should be optimised for any system independently. In order to work with as little training data as possible it should be chosen as small as possible without sacrificing accuracy. The training data can be generated from compensation measurements without neural network, by extracting a subset of points and labeling it with the final compensation point. This additionally inflates the training data, as for any $n \cdot k$ datapoints only a subset is picked and this subset can be mutated in $\binom{n}{l}^k$ ways, where l is the number of datapoints used as input. The number of datapoints required for compensation is then given by $a_{\mathrm{ANN}} \cdot n \cdot k$ with $a_{\mathrm{ANN}} = l/(n \cdot k)$.

9.2.4 Accuracy of the general approach

In order to determine the accuracy of a compensation method, we define a measure σ, which represents the standard deviation of predicted compensation points for several input data sets of a common noise manifestation in terms of their Euclidean distance in parameter space. For the following, we also verify that any predicted compensation point is within one standard deviation of the actual compensation point as determined by the linear model.

9.3 Compensating a trapped ions position

In order to verify the functionality of the proposed methods, we demonstrate them on an experimental system, namely a single trapped ytterbium ion within a radio-frequency Paul trap. After a brief introduction to the experimental setup, we compare a simple linear model, a PCA enhanced version of said model and an artificial neural network by measuring the accuracy of the different methods. We show that a reduction of input data is possible through PCA or the ANN up to $a = 1/k$ without reduced accuracy. In case of the neural network, by

sacrificing some accuracy, it is even possible to compensate with just one datapoint, meaning $a_{\mathrm{ANN}} = 1/(n \cdot k)$.

9.3.1 Brief introduction: Experimental system

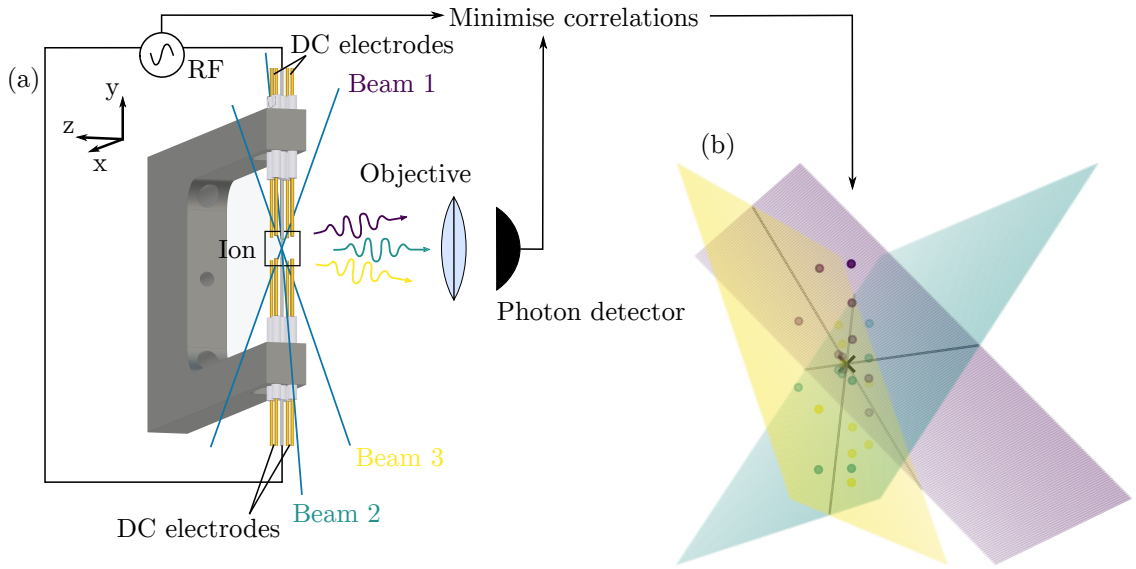

Figure 9.3: Ion trap setup and plane model. (a) Ion trap setup and sketch of the measurement procedure for Doppler correlation based compensation of micromotion. The ion fluorescence generated by illumination with one of three linearly independent beams is measured by a photodetector. The goal is to minimise the correlation between the rf signal on the rf electrodes and the ion fluorescence signal by applying a DC voltage on linear combinations of the gold DC electrodes. **(b)** Points of minimised micromotion form a hyperplane for each beam separately. The intersection point of all hyperplanes presents the point of minimum micromotion, the compensation point.

The system used for demonstration is a radio-frequency Paul trap used for capturing $^{174}Yb^{+}$ ions. It was originally built by Dr. Matthias Steiner and Dr. Hendrik-Marten Meyer during their PhD work [226, 227]. After moving from Cambridge to Bonn, it was brought back to service with help of Dr. Hendrik-Marten Meyer and Dr. Tim Ballance during my masters thesis [228]. Currently, the setup is further developed by Pascal Kobel, who will give a more detailed introduction of the setup in context of these measurements in his PhD thesis [224]. The relevant parts of the setup are shown in figure 9.3. The ytterbium ion is trapped between two etched tungsten needles, 100 µm apart. The needles act as rf-electrodes and provide a time-varying confining potential. Around the needles, recessed by 0.5 mm, gold dc-electrodes are placed and used to compensate for external fields and keep the ion in the center of the rf-potential. The ion is continually laser cooled and trapping times of hours can be achieved.

9.3.2 Stray field compensation

Stray electric fields surrounding the trapped ion can push the ion out of the symmetry point of the quadrupole trap. This will increase the amplitude of the motional component oscillating with the driving frequency of the trap. This fast micromotion decreases coherence and severely limits the systems use in quantum information processing [229] and atomic clocks [230, 231]. Numerous ways of compensating stray electric fields have been presented [232–237]. Here, we use the Doppler correlation technique [233, 238, 239], which relies on minimising the correlation between the ion fluorescence excited by a laser beam and the driving frequency of the trap. As this does only compensate the motion along one direction, the procedure has to be repeated with three linearly independent laser beams. The compensation voltages are applied to linear combinations of the gold electrodes, resulting in three different control parameters. Hence, the dimension of parameter space is $k = 3$. The measurement procedure as well as example data can be seen in figure 9.3.

9.3.3 The plane model

As mentioned above, minimising the correlation between the driving frequency and the ions fluorescence does only compensate the micromotion along the direction of the impinging laser beam. All points with compensation along this one direction then form a hyperplane in parameter space as long as the misalignment stays in the linear regime. Note that this plane does not have to be orthogonal to the beam axis, but its orientation is dependent on the coupling between the experimental parameters. Determining the orientation of said plane for all linearly independent beams yields a single point where all planes intersect, representing the optimal point of compensation, see figure 9.3 (b). This situation is exactly the situation described in section 9.2.1, but with $k = 3$. In principle, three datapoints are needed to define a plane in 3D, however, we find that using only three points yields an accuracy of $734\,\mathrm{V\,m^{-1}}$, which is larger than the typical range of the compensation drift, which averages to $\approx 600\,\mathrm{V\,m^{-1}}$. This is due to the quality of the fitted plane parameters. By using $n = 8$ datapoints per plane and therefore a total of $a_{\mathrm{plane}} \cdot n \cdot k = 24$ datapoints, the accuracy becomes $\sigma_{\mathrm{plane}} = 53\,\mathrm{V\,m^{-1}}$, which is sufficiently good for our day to day use and will represent the figure of merit for the following improvements.

9.3.4 PCA enhanced plane model

In order to apply PCA for a reduction in required data for compensating the stray electric fields, it is necessary to find a description of the planes in the plane model. The idea is to explain the variation of the plane parameters via principal components with large weights. Finally, we want to identify the orientation and position of a plane with different noise characteristics by means of the variation found with PCA. The specific implementation of the PCA method in the context of the plane model was developed by Pascal Kobel and will be detailed in his PhD thesis [224]. The position of a plane in 3D can be described by

$$a \cdot x + b \cdot y + 1 \cdot z - M = 0, \tag{9.1}$$

where a and b are coefficients describing the orientation and M is the offset in z-direction at $(x, y) = (0, 0)$. The input data for the PCA then consists of three 3-dimensional vectors of shape (a_i, b_i, M_i), where $i \in \{1, 2, 3\}$ is the beam number. This will yield three eigenvectors per beam that can be used to reconstruct each of the three planes by finding a linear combination of the eigenvectors that matches three independent points on the plane and transforming back into the (a_i, b_i, M_i) basis. To reduce the dimensionality of the problem, instead of using all three principal components for the plane, we only use the first principal component, which in our case already represents 89 % of the variance in the dataset. Let (j, k, l) be the eigenvalues corresponding to the eigenvectors $(\mathbf{J}, \mathbf{K}, \mathbf{L})$ with $j > k > l$ for beam i. Any plane can then be described by a linear combination of $\mathbf{J}, \mathbf{K}$ and $\mathbf{L}$. However, since most of the variance is along the first eigenvector $\mathbf{J}$, the planes in the dataset can be described reasonably well by just the group of planes corresponding to $\alpha \cdot \mathbf{J}$, where α is chosen such that the input point lies on the plane. This reduces the number of points needed to find the plane for each beam to just one datapoint. In total, this results in $a_{\mathrm{PCA}} \cdot n \cdot k = 3$ datapoints for one compensation measurement, assuming the principal components describe the noise profile reasonably well. Indeed, we find that the PCA enhanced model with an input of one point per beam achieves an accuracy of $\sigma_{\mathrm{PCA}} = 42\,\mathrm{V\,m^{-1}}$, which is smaller than the reference $\sigma_{\mathrm{plane}} = 53\,\mathrm{V\,m^{-1}}$. This concludes, that the accuracy of the method is not limited by the method itself, but rather by the accuracy of the measurement used to determine the principal components and that the PCA can successfully reduce the amount of data required with $a_{\mathrm{PCA}} = 1/n = 1/8$.

9.3.5 Artificial neural network model

In order to use PCA, one has to employ a model to find the principal component representation of the solutions. In this case, the model was given by the assumption that the solutions per beam form a plane in parameter space. Sometimes, using a model is complicated or no model exists for a given problem. Here, we demonstrate the use of a fully connected neural network to attain a solution without any prior knowledge of the system. The network architecture consists of an input layer of variable size, one hidden layer with 16 neurons and an output layer with three neurons, corresponding to the C_i settings of the compensation point, see figure 9.4. For the input layer, we test two configurations: Firstly, we use 9 neurons, corresponding to the C_i's of one point per plane and hence $a_{\mathrm{ANN},\,9} \cdot n \cdot k = 3$. Secondly we use 4 neurons, where the first three neurons correspond to a point on one plane and the fourth neuron identifies the beam used to generate the plane, resulting in a total of $a_{\mathrm{ANN},\,4} \cdot n \cdot k = 1$ datapoints. We train the network on subsets of datapoints from seven compensation runs by using backpropagation with Adam optimiser (see section 3.4.2) with the target labels being the corresponding compensation point. We find, that the 9-neuron neural network achieves comparably good results to the PCA enhanced plane model with $\sigma_{\mathrm{ANN},\,9} = 34\,\mathrm{V\,m^{-1}}$. For the 4-neuron network, the achieved accuracy drops to $\sigma_{\mathrm{ANN},\,4} = 271\,\mathrm{V\,m^{-1}}$. This shows, that without any prior knowledge of the system the neural network can perform equally well as the PCA enhanced model and is therefore limited by the measurement error and not by the intrinsical error of the method. Using only one datapoint as input reduces the accuracy but depending on the needed accuracy might still constitute an alternative, as it can speed up the compensation by an additional factor of $n = 3$.

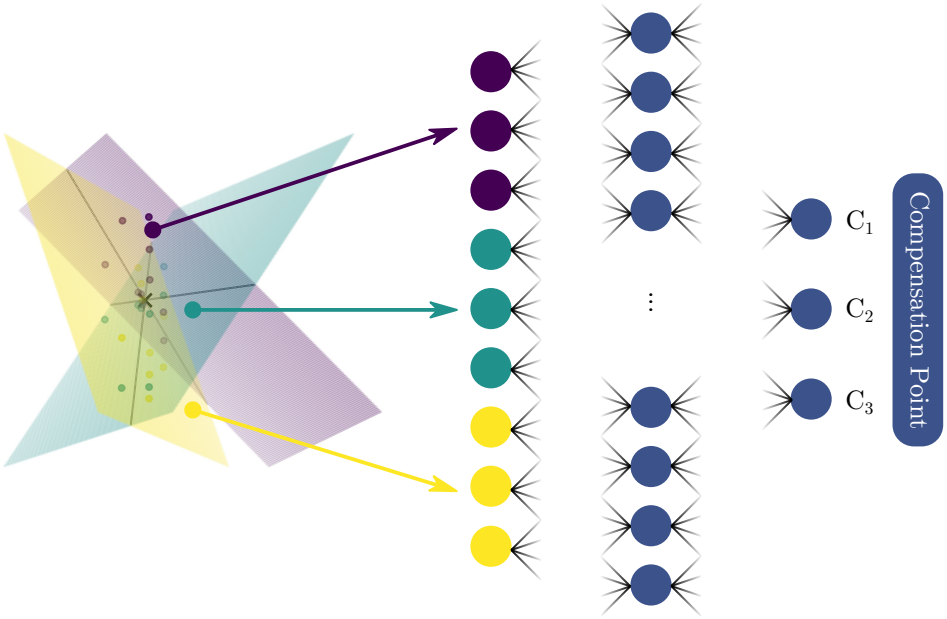

Figure 9.4: A neural network as empirical model. The 9 neuron neural network for the prediction of the compensation point. The network receives one point per beam as input and predicts the control parameters of the fully compensated point. In case of the 4 neuron neural network, the input layer consists of 4 neurons: One datapoint (=three neurons) and an identifier neuron that receives the information which beam was used.

9.3.6 Results

A summary of the results presented above is given in figure 9.5. Here, we compare the accuracy of different methods given a fixed number of measurements. The diameter of the circles indicates the deviation of the spread σ, normalised to the unmodified plane model with an accuracy of $54\,\mathrm{V\,m^{-1}}$. In summary, it was shown that machine learning approaches can speed up compensation measurements exponentially. This speeds up data acquisition and might even facilitate active stabilisation within one experimental cycle, that was not possible before. Although machine learning approaches usually need large amounts of data, we only use seven compensation measurements to train our models. This is due to the fact that the used models are kept simple. Since the introduced methodology is easy to implement, we expect that this ansatz can be used in a variety of different experimental settings.

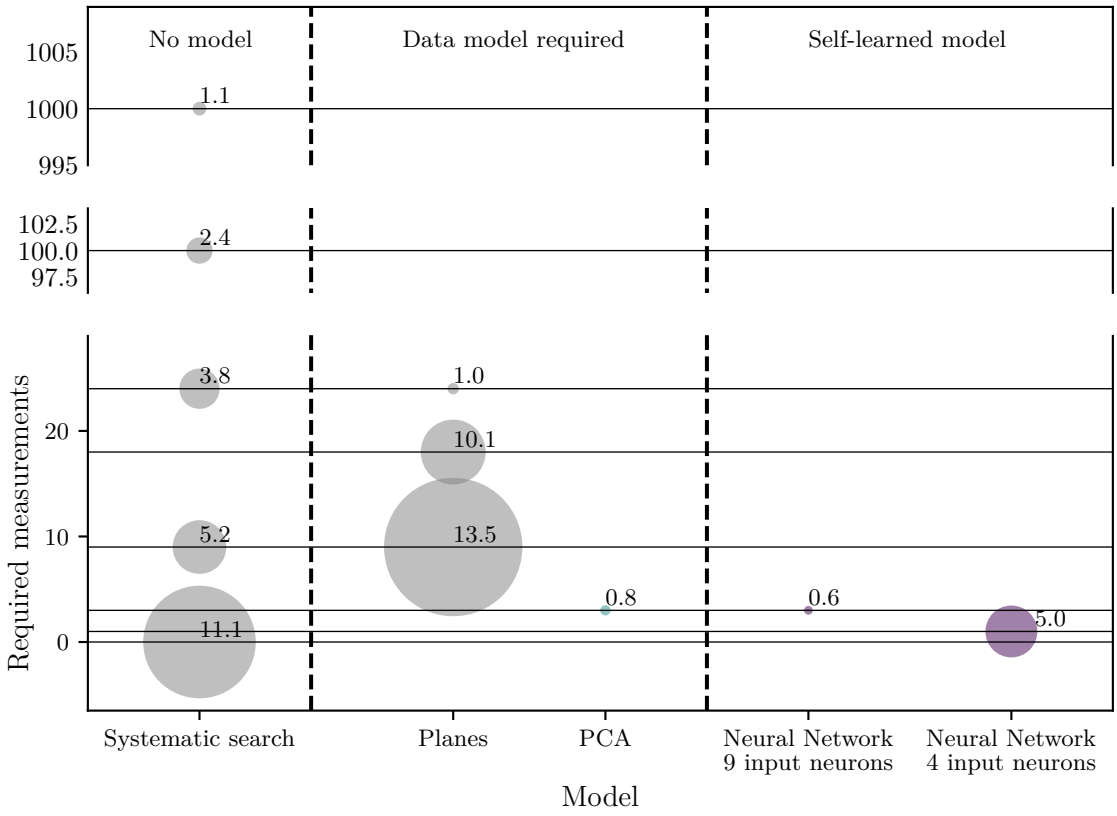

Figure 9.5: Accuracy of different models for given number of measurements. The size of the datapoint indicates the standard deviation of the spread σ at a given number of measurements. The number indicates the accuracy relative to the Doppler correlation measurement.

CHAPTER 10

Summary and Outlook

This chapter will briefly summarise the contents and results of this thesis. Based on the work that has been done, we will present three ways forward for future work. We will show that, although foreign at first glance, the two main topics of this thesis – non-equilibrium measurements and machine learning in physics – can be brought together in the future. As already briefly mentioned in chapter 8, one way to use the knowledge and experience gained during this work is by detecting oscillations in the condensate fraction directly from time-of-flight pictures without a rapid ramp that can potentially obstruct experimental signatures. Experimentally, two projects have already been started. Namely, the realisation of the 2D-3D crossover with a trap created by a blue detuned TEM01 beam. With this trap and with the techniques developed during this work, the question of the existence of the Higgs mode in the 2D-3D crossover may be answered in the future. Finally, in many cases a homogeneous system simplifies calculations and the interpretation of measurements. This is because many quantities that have a spatial dependence due to the harmonic potential are constant throughout the trap. A hollow beam setup based on axicons has already been prepared and implemented into the experimental setup and awaits further characterisation and alignment.

10.1 Summary

This thesis contains two main topics, which, at first glance, are disconnected. Namely, non-equilibrium physics in Fermi gases and machine learning in physics. Here, the results of the investigations into each of these topics are recalled and summarised, before attempting to bring both together for future and ongoing projects.

10.1.1 Non-equilibrium in ultracold Fermi gases

Non-equilibrium phenomena have been at the heart of several investigations in the past years. In context of Fermi gases, the question at the heart of these investigations is how a Fermi gas reacts to sudden changes of the interaction. On the BCS side of the Feshbach resonance, where particle-hole symmetry is roughly fulfilled, a stable Higgs mode characterised through oscillations in the absolute value of the order parameter $|\Delta|$ at a frequency of $2|\Delta|/h$ is

predicted. In chapter 7 this mode was observed by spectroscopic measurements. The mode was excited by a modulation of the order parameter, achieved by a weak modulation of the state population by applying radio-frequency radiation. Tailoring the detuning of the rf frequency and the rf power, the modulation frequency was scanned while keeping the amplitude constant. Theoretical investigations have confirmed that this kind of rf dressing indeed is able to excite oscillations in the order parameter. For certain modulation frequencies we observed a decrease in the condensate fraction, signaling the presence of the mode. We compared the corresponding frequency of the mode to theoretical and experimental values of the gap parameter and found good agreement. Furthermore, we observed a broadening of the mode towards the BEC side, caused by the gradual loss of particle hole symmetry required for a stable Higgs mode.

In order to observe the non-equilibrium behaviour of an ultracold Fermi gas directly, it has to be subjected to a sudden quench of the interaction. However, these quenches are subject to experimental difficulties, mostly stemming from the required short timescale of the change on the order of the Fermi time $\tau_F = h/E_F \approx 4.7\,\mu s$. Two experimental approaches to sudden quenches were presented in chapter 8. The first approach presented a quench that was achieved by a fast radio-frequency transfer of one of the states in the two-component Fermi gas to a different state. As the Feshbach resonances of both mixtures have different properties this constitutes an effective change in interaction. The achieved timescale for the full transfer was $28(2)\,\mu s$, which is in the order of the quasi-particle excitation time. The response of the system to a quench from strongly to weakly interacting showed an initial fast decay of the condensate fraction, followed by a squeezed exponential approach to a steady state value which is higher than the lowest achieved value. The width of the condensate part of the cloud showed an initial reshuffling of quasiparticle excitations on the order of the quasiparticle relaxation time. This was followed by an oscillation that can be explained by an excitation of the monopole mode in the gas. The observed oscillation frequency agrees well with theoretical predictions for a hydrodynamic gas. By comparing the results to the predictions of a theoretical model, it was possible to verify the observed response of the gas. The second approach implemented the quenches by means of fast magnetic field quenches. For this purpose, we designed and tested a new miniature coil setup including switching circuit. The new setup was able to achieve quench times on the order of $t_{\mathrm{quench}} = 3\,\mu s < \tau_F$. The response of the system to a quench from weakly to strongly interacting in the long time limit $t \gg \tau_F$ showed oscillations in the density and condensate fraction of the cloud with an interaction dependent time delay between them. For both approaches no oscillations in the condensate fraction were observed in the short time limit $t < 1/\nu_{\mathrm{Higgs}}$ yet.

10.1.2 Machine learning in physics

Machine Learning has manifested itself as a strong tool in physics in the past years. In this thesis, it has been used in several ways for investigations. Firstly, to measure the critical temperature and phase diagram in the BEC-BCS crossover regime in chapter 5. In section 5.2 we trained a neural network to detect the condensate fraction directly from time of flight pictures. This was not possible before, as the signature of a Fermi condensate becomes practically undetectable towards unitarity and the BCS side of the Feshbach resonance. So

far, to determine the condensate fraction a rapid projection onto tightly bound molecules has been used. While this technique is still used to generate the labels for the training process, at the time of inference the network does not require the projection anymore. This can be especially valuable in cases where the projection itself can obstruct experimental quantities. We used the network to find a clearer signature for the onset of interaction and, consequently, the critical temperature by setting the condensate fraction to zero for thermal clouds. This generates a sharp onset of the condensate that cannot be achieved by a common bimodal fit. With this, we measured the BEC-BCS Crossover phase diagram with high resolution. We found very good agreement with theory and experiment around unitarity and towards the BEC side, however, we observed a slowed decay of the critical temperature towards the weakly attractive regime.
Although independent of the rapid ramp at the time of inference, the supervised learning method still relies on labelled example data generated from a rapid ramp sequence. In chapter 6, we used an autoencoder network to generate a low-dimensional representation of the data that was used to measure the phase diagram with the supervised neural network. We showed that the representation successfully distinguishes between interactions and temperatures. By analysing the shape of the different interaction clusters it was possible to extract the critical temperature from the dataset. We found very good agreement with theory and experiment throughout the range of investigated interactions. It should be noted that the only input into the autoencoder were the time-of-flight pictures and that all recognised features were detected from first principles without any additional information or input.
Finally, in chapter 9, we used two machine learning approaches – principal component analysis and fully connected neural networks – to improve the performance of compensation measurements in a general physics context. For this, we created empirical models of the experiment setup that are able to determine the correct set of parameters to compensate disturbances with exponentially less data than the general case of brute force search. Both empirical models, for principal component analysis and the neural network, were tested in the context of stray field compensation for a single trapped ytterbium ion in a Paul trap and achieved an accuracy only limited by measurement error with a fraction of data needed compared to the full compensation procedure.

10.2 Outlook

This section presents three possible future routes of investigation. All suggested future endeavors draw from the results of this thesis. With this, I hope to convince the reader of having performed not only the results summarised in the previous section, but also the ground work for future investigations.

10.2.1 Coherent detection of the Higgs mode with a neural network

Although the spectroscopic method presented in chapter 7 presents a detection of the Higgs mode within a fermionic BCS superfluid, it is still not sensitive to the coherence of the mode. A more direct way to detect the mode would be to observe oscillations with frequency $2|\Delta|/h$ in the order parameter Δ after a sudden change in interaction strength. As Δ cannot be observed directly, the observable of choice is the condensate fraction which is directly related.

In chapter 8, a magnetic field coil was developed that is able to change the interaction strength within the timescale that is necessary to excite the Higgs mode. So far, in the short time limit $t \approx \tau_{\mathrm{higgs}}$ no oscillation of the condensate fraction has been observed. One possible reason for this could be a dampening of the oscillation amplitude during the rapid ramp. Theoretical investigations have shown that the mode might become strongly damped towards the BEC regime [189]. This is supported by the observed strong broadening of the mode signature in the spectroscopic measurement towards the BEC side of the Feshbach resonance shown in chapter 7 and ref [1]. Assuming an excitation of the Higgs mode in the BCS regime, ramping to the BEC side during rapid ramp might therefore obscure any oscillations in the condensate fraction and consequently prevent a direct observation of the mode.
The neural network based detection of the condensate fraction presented in section 5.2 presents a way around this problem. Although the neural network is trained by labels generated through the rapid ramp technique, after training, it is possible to detect the condensate fraction without rapid ramp from time of flight pictures. After a magnetic field quench the interaction therefore would not have to be changed in order to convert pairs to tightly bound molecules and the extraction of the condensate fraction could take place at the magnetic field corresponding to the final interaction strength. This might improve the amplitude of the oscillation and, with an accurate measure of the condensate fraction through the network, might enable a direct observation of the Higgs mode.

10.2.2 The Higgs mode in the 2D-3D crossover

In two-dimensional systems, long range order is prohibited by thermal fluctuations at any finite temperature [240, 241]. As a consequence, superfluidity like in the 3D case investigated in this thesis, cannot exist. However, below a critical temperature, unpaired vortices and anti-vortices can transition to vortex-anti-vortex pairs and form a *quasi-condensate*. This second order topological phase transition is called Berezinski-Kosterlitz-Thouless transition after their discoverers [242–244], a work awarded with the Nobel prize in physics in 2016. There is, however, no associated broken symmetry and therefore no corresponding spatially uniform order parameter like the gap Δ in 3D. The BKT-transition has been observed in a trapped atomic gas [245], two-dimensional ^{4}He films [246], josephson-junctions [247] and two-dimensional atomic hydrogen [248]. The question, whether a stable Higgs mode can exist in a system that does not exhibit spontaneous symmetry breaking is a field of active research [249–251]. Fermi gases provide the experimental means to investigate a continuous crossover from 3D to 2D systems [252]. Together with the spectroscopic detection method introduced in chapter 7 or a direct measurement based on the quenches introduced in chapter 8, possibly in combination with the neural network approach from chapter 5 (see also section 10.2.1 within this chapter), might provide a way to study the evolution of the Higgs mode in the dimensional crossover.

During this work, we already implemented the required tools into the experimental system. To reach the 2D regime, the atomic cloud has to be squeezed along one direction, i.e. the excitation energy associated with the trap frequency along one direction has to become large as compared to the energy scale of the atomic sample. This will effectively *freeze out* any dynamics along this direction. A trap which is capable to squeeze the cloud is generated by a green detuned high power laser beam. Compared to the gaussian dipole trap beams, the

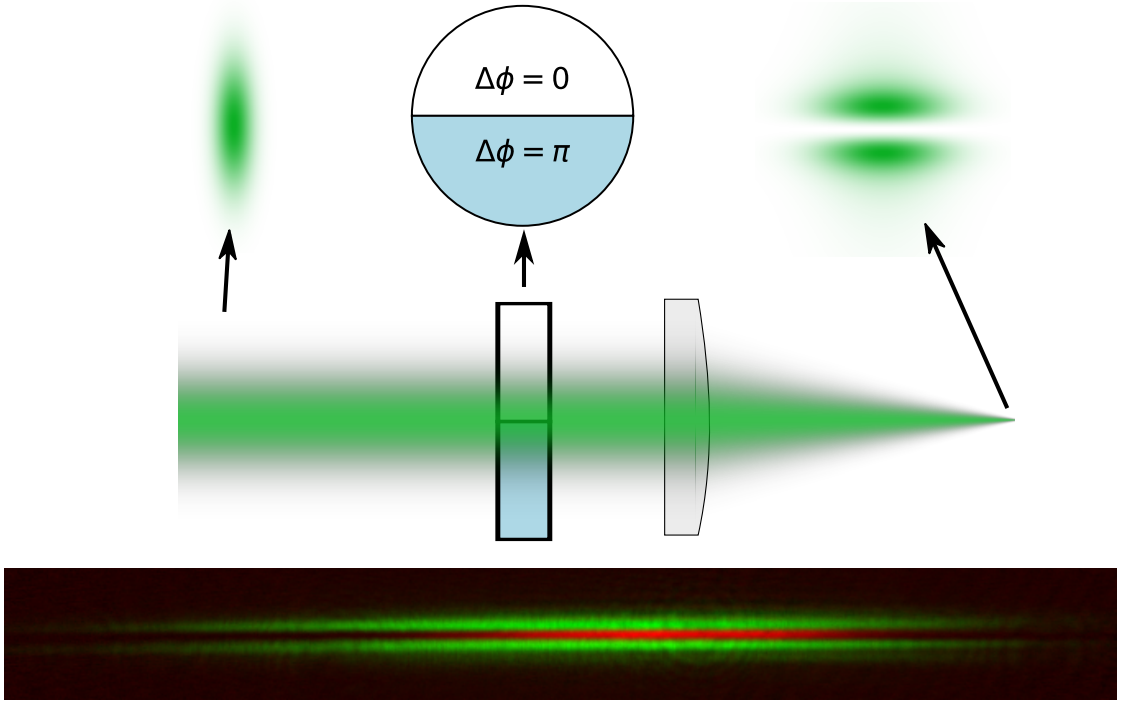

Figure 10.1: TEM01 beam concept. The TEM01 beam is created with a waveplate that creates a phase difference of π between two halves of the incoming beam. The beam is focused such that the the cloud is squeezed between the two light sheets. The bottom picture shows a false color picture of the intensity distribution of the TEM01 beam (green) and the shape of the atom cloud (red). The atom cloud and beam were imaged separately and superimposed.

green beam is transferred to a higher transversal mode, namely TEM01. This is achieved by a waveplate that introduces a phase difference of $\Delta\Phi = \pi$ between two halves of the beam. The beam is then elongated and focused onto the atom position. A sketch of the beam preparation and an image of a trapped cloud within this trap can be seen in figure 10.1, more information on the initial setup and testing can be found in the masters thesis of Andreas Kell [253]. The dimensional crossover can be verified by observation of the scaling of the transversal cloud size as shown in figure 10.2. The size of the cloud along the squeezed direction is constant, as long as only the ground state is populated. Increasing the atom number increases the Fermi energy of the system, occupying more energy states, up to a point where a higher transversal state is populated and the scaling becomes linear.

Having all experimental prerequisites built into the system, future investigations might focus on the presence of the Higgs mode in the 2D-3D crossover. This might help the understanding of the nature of 2D superfluidity and the BKT-transition. In order to go deeper into the 2D regime, the setup can be extended by a lattice laser that was set up by Andreas Kell during his masters thesis [253].

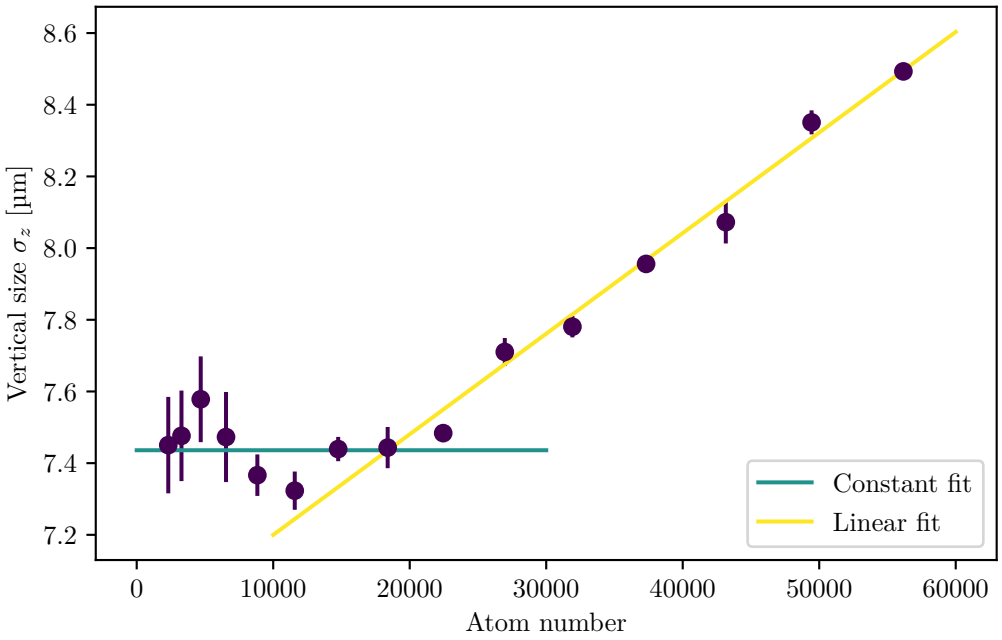

Figure 10.2: Scaling in the 2D-3D crossover. The vertical size of the cloud does not scale with atom number, as long as no higher states than the ground state are populated in the squeezed direction. If the atom number increases above the critical point, where the Fermi energy is comparable to the energy difference and the dynamics is not frozen out, the cloud size scales linearly. The plot shows the cloud size in vertical direction versus the atom number together with a constant and linear fit indicating the two regimes.

10.2.3 Homogeneous Fermi gases and the gap in imbalanced Fermi gases

Measuring the Higgs mode amounts to measuring the amplitude of the gap parameter Δ. Being an important quantity throughout the crossover, the behaviour of Δ has remained experimentally elusive. Imbalanced Fermi gases provide an especially interesting prospect for investigation. Superfluidity in imbalanced mixtures has been investigated [254, 255], however the gap has not been measured in the case of imbalance. It is expected to decrease, as with increasing imbalance different regions of the Fermi surface partake in the pairing [256, 257], however, whether it vanishes completely has not been measured and could be interesting the context of Fulde-Ferrell-Larkin-Ovchinnikov (FFLO) finite-momentum superconductivity, where different subregions of the Fermi surfaces contribute in the pairing. In the presence of a harmonic trap, strongly interacting imbalanced Fermi gases undergo phase separation [258] which complicates the gap measurements. Additionally, the non-uniform spatial density leads to varying energy and length scales throughout the trap, complicating the observation of exotic phases of matter such as the aforementioned FFLO superconductivity that only exist in a narrow range of parameters. To circumvent the undesired effects of a harmonic trapping potential, homogeneous atom traps have been developed and used in 2D and 3D Fermi gases [144, 145]. A realisation of these traps based on three axicons [145, 259, 260]

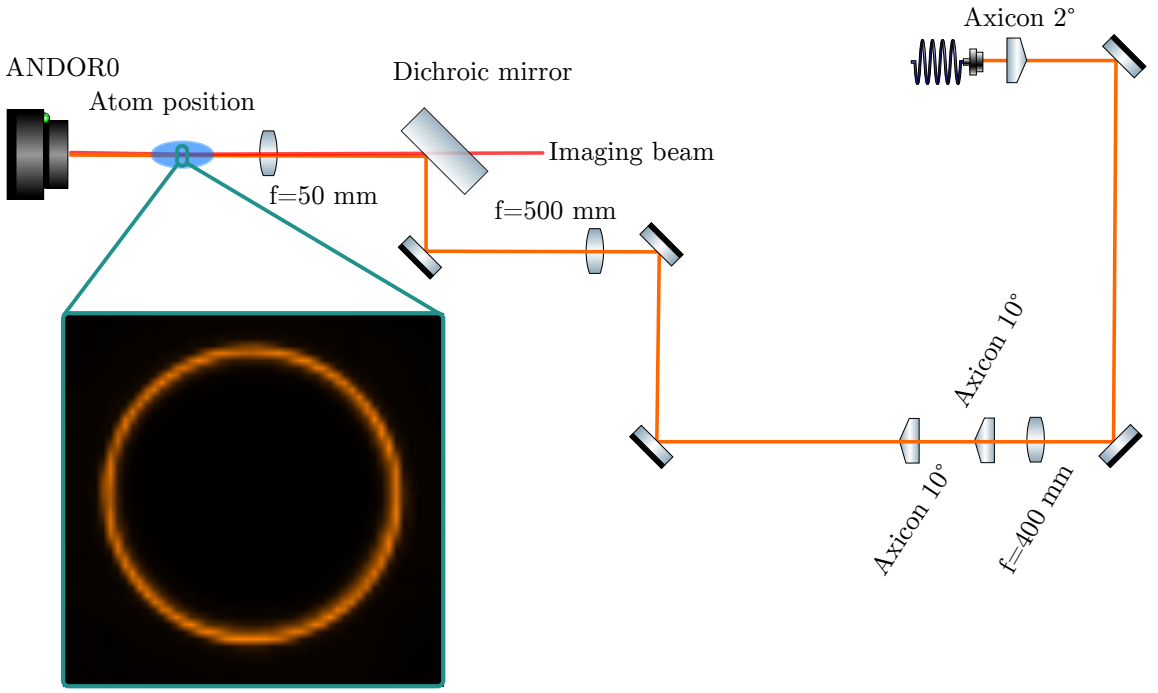

Figure 10.3: Axicon beam concept. The beam for the creation of the hollow beam is indicated in orange. The setup uses three axicons and is based on the setup by Hueck et al. [145]. The inset shows a camera picture of the hollow beam.

has been built and tested for our experiment. The axicons are used to create a hollow beam with sharp edges at a wavelength of 589 nm. Together with the light sheets created from the TEM01 beam described in section 10.2.2, a homogeneous trap can be built. The experimental setup is shown in figure 10.3, together with a camera picture of the hollow beam. The beam is already aligned to the atom position and has to be characterised in detail for future measurements. Together with the techniques developed in chapters 7 and 8, the homogeneous trap might provide a way to measure the superfluid gap in more exotic systems such as the imbalanced Fermi gas and understand the nature of pairing in these systems. Furthermore, the homogeneous trap will be a valuable tool for all future measurements that aim to observe phenomena that are obstructed by spatially varying density, as given in a harmonic trap.

Bibliography

[1] A. Behrle et al., *Higgs mode in a strongly interacting fermionic superfluid*, Nature Physics **14** (2018) 781–785, URL: https://doi.org/10.1038/s41567-018-0128-6 (cit. on pp. iv, 105, 160).

[2] T. Harrison et al., *Decay and revival of a transient trapped Fermi condensate*, 2020, arXiv: 2007.11466 [cond-mat.quant-gas] (cit. on pp. iv, 121, 123).

[3] P. Kobel, M. Link and M. Köhl, *Exponentially improved detection and correction of errors in experimental systems using neural networks*, 2020, arXiv: 2005.09119 [quant-ph] (cit. on pp. iv, 145).

[4] M. Gall, *Quantum simulation of the bilayer Hubbard model using ultracold atoms*, PhD thesis: Universität Bonn, 2020 (cit. on p. v).

[5] K. Onnes, *Research Notebooks*, tech. rep., Kammerlingh Onnes Archive, 1987 (cit. on p. 1).

[6] K. Onnes, *The resistance of pure mercury at helium temperatures*, Comm. Leiden **119b** (1911) (cit. on pp. 1, 18).

[7] K. Onnes, *The disappearance of resistivity of mercury*, Comm. leiden (1911) (cit. on pp. 1, 18).

[8] J. Bardeen, L. N. Cooper and J. R. Schrieffer, *Theory of Superconductivity*, Phys. Rev. **108** (5 1957) 1175–1204, URL: https://link.aps.org/doi/10.1103/PhysRev.108.1175 (cit. on pp. 1, 19, 21).

[9] J. Bardeen, L. N. Cooper and J. R. Schrieffer, *Microscopic Theory of Superconductivity*, Phys. Rev. **106** (1 1957) 162–164, URL: https://link.aps.org/doi/10.1103/PhysRev.106.162 (cit. on pp. 1, 19, 21, 23).

[10] M. Randeria and E. Taylor, *Crossover from Bardeen-Cooper-Schrieffer to Bose-Einstein Condensation and the Unitary Fermi Gas*, Annual Review of Condensed Matter Physics **5** (2014) 209–232, eprint: https://doi.org/10.1146/annurev-conmatphys-031113-133829, URL: https://doi.org/10.1146/annurev-conmatphys-031113-133829 (cit. on pp. 2, 29).

[11] C. A. R. Sá de Melo, M. Randeria and J. R. Engelbrecht,
Crossover from BCS to Bose superconductivity: Transition temperature and time-dependent Ginzburg-Landau theory, Phys. Rev. Lett. **71** (19 1993) 3202–3205, URL: https://link.aps.org/doi/10.1103/PhysRevLett.71.3202
(cit. on pp. 2, 26, 28–30).

[12] S. Bose, *Plancks Gesetz und Lichquantenhypothese*, Z. Phys **26** (1924) 178
(cit. on p. 3).

[13] A. Einstein, *Sitzungsbericht der Preussischen Akademie der Wissenschaften, Physikalisch-mathematische Klasse*, tech. rep., 1925 (cit. on p. 3).

[14] A. Einstein, *Sitzungsbericht der Preussischen Akademie der Wissenschaften, Physikalisch-mathematische Klasse*, tech. rep., 1924 (cit. on p. 3).

[15] C. A. Regal et al., *Creation of ultracold molecules from a Fermi gas of atoms*, Nature **424** (2003) 47–50, URL: https://doi.org/10.1038/nature01738
(cit. on pp. 3, 71).

[16] M. Greiner, C. A. Regal and D. S. Jin,
Emergence of a molecular Bose–Einstein condensate from a Fermi gas,
Nature **426** (2003) 537–540, URL: https://doi.org/10.1038/nature02199
(cit. on pp. 3, 71).

[17] C. A. Regal, M. Greiner and D. S. Jin,
Observation of Resonance Condensation of Fermionic Atom Pairs,
Phys. Rev. Lett. **92** (4 2004) 040403,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.92.040403
(cit. on pp. 3, 67, 72).

[18] M. W. Zwierlein et al.,
Condensation of Pairs of Fermionic Atoms near a Feshbach Resonance,
Phys. Rev. Lett. **92** (12 2004) 120403,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.92.120403
(cit. on pp. 3, 67, 72, 94, 95).

[19] A. Bulgac, J. E. Drut and P. Magierski,
Quantum Monte Carlo simulations of the BCS-BEC crossover at finite temperature,
Phys. Rev. A **78** (2 2008) 023625,
URL: https://link.aps.org/doi/10.1103/PhysRevA.78.023625
(cit. on pp. 3, 6, 72, 73, 94, 95, 102, 117, 119).

[20] I. Boettcher, J. M. Pawlowski and S. Diehl,
Ultracold atoms and the Functional Renormalization Group,
Nuclear Physics B - Proceedings Supplements **228** (2012) 63 –135, "Physics at all scales: The Renormalization Group" Proceedings of the 49th Internationale Universitätswochen für Theoretische Physik, ISSN: 0920-5632, URL: http://www.sciencedirect.com/science/article/pii/S0920563212001612
(cit. on pp. 3, 28, 72).

[21] L. Pisani et al.,
Entanglement between pairing and screening in the Gorkov-Melik-Barkhudarov correction to the critical temperature throughout the BCS-BEC crossover,
Physical Review B **97** (2018), ISSN: 2469-9969,
URL: http://dx.doi.org/10.1103/PhysRevB.97.014528
(cit. on pp. 3, 72, 73, 94, 102).

[22] M. J. H. Ku et al., *Revealing the Superfluid Lambda Transition in the Universal Thermodynamics of a Unitary Fermi Gas*, Science **335** (2012) 563–567,
ISSN: 1095-9203, URL: http://dx.doi.org/10.1126/science.1214987
(cit. on pp. 3, 72, 73, 85, 93, 95, 102, 104).

[23] G. Carleo et al., *Machine learning and the physical sciences*,
Rev. Mod. Phys. **91** (4 2019) 045002,
URL: https://link.aps.org/doi/10.1103/RevModPhys.91.045002
(cit. on pp. 4, 34, 73).

[24] J. Carrasquilla and R. G. Melko, *Machine learning phases of matter*,
Nature Physics **13** (2017) 431–434, ISSN: 1745-2481,
URL: http://dx.doi.org/10.1038/nphys4035 (cit. on pp. 4, 73).

[25] K. Ch'ng et al., *Machine Learning Phases of Strongly Correlated Fermions*,
Physical Review X **7** (2017), ISSN: 2160-3308,
URL: http://dx.doi.org/10.1103/PhysRevX.7.031038 (cit. on pp. 4, 73).

[26] P. Broecker et al.,
Machine learning quantum phases of matter beyond the fermion sign problem,
Scientific Reports **7** (2017), ISSN: 2045-2322,
URL: http://dx.doi.org/10.1038/s41598-017-09098-0 (cit. on pp. 4, 73).

[27] P. Broecker, F. F. Assaad and S. Trebst,
Quantum phase recognition via unsupervised machine learning, 2017,
arXiv: 1707.00663 [cond-mat.str-el] (cit. on pp. 4, 73).

[28] D.-L. Deng, X. Li and S. Das Sarma, *Machine learning topological states*,
Physical Review B **96** (2017), ISSN: 2469-9969,
URL: http://dx.doi.org/10.1103/PhysRevB.96.195145 (cit. on pp. 4, 73).

[29] Y. Zhang and E.-A. Kim, *Quantum Loop Topography for Machine Learning*,
Physical Review Letters **118** (2017), ISSN: 1079-7114,
URL: http://dx.doi.org/10.1103/PhysRevLett.118.216401 (cit. on pp. 4, 73).

[30] P. Zhang, H. Shen and H. Zhai,
Machine Learning Topological Invariants with Neural Networks,
Physical Review Letters **120** (2018), ISSN: 1079-7114,
URL: http://dx.doi.org/10.1103/PhysRevLett.120.066401 (cit. on pp. 4, 73).

[31] B. S. Rem et al., *Identifying quantum phase transitions using artificial neural networks on experimental data*, Nature Physics **15** (2019) 917–920,
URL: https://doi.org/10.1038/s41567-019-0554-0 (cit. on pp. 4, 34, 73, 77).

[32] R. Sooryakumar and M. V. Klein, *Raman Scattering by Superconducting-Gap Excitations and Their Coupling to Charge-Density Waves*, Phys. Rev. Lett. **45** (8 1980) 660–662, URL: https://link.aps.org/doi/10.1103/PhysRevLett.45.660 (cit. on pp. 5, 106).

[33] R. Matsunaga et al., *Higgs Amplitude Mode in the BCS Superconductors* $Nb_{1-x}Ti_xN$ *Induced by Terahertz Pulse Excitation*, Phys. Rev. Lett. **111** (5 2013) 057002, URL: https://link.aps.org/doi/10.1103/PhysRevLett.111.057002 (cit. on pp. 5, 106).

[34] D. Sherman et al., *The Higgs mode in disordered superconductors close to a quantum phase transition*, Nature Physics **11** (2015) 188–192, URL: https://doi.org/10.1038/nphys3227 (cit. on pp. 5, 106).

[35] S. Y. Chang et al., *Quantum Monte Carlo studies of superfluid Fermi gases*, Phys. Rev. A **70** (4 2004) 043602, URL: https://link.aps.org/doi/10.1103/PhysRevA.70.043602 (cit. on pp. 6, 117, 119).

[36] A. Gezerlis and J. Carlson, *Strongly paired fermions: Cold atoms and neutron matter*, Phys. Rev. C **77** (3 2008) 032801, URL: https://link.aps.org/doi/10.1103/PhysRevC.77.032801 (cit. on pp. 6, 117, 119).

[37] Q. Chen, *Effect of the particle-hole channel on BCS–Bose-Einstein condensation crossover in atomic Fermi gases*, Scientific Reports **6** (2016) 25772, URL: https://doi.org/10.1038/srep25772 (cit. on pp. 6, 117, 119).

[38] R. Haussmann et al., *Thermodynamics of the BCS-BEC crossover*, Phys. Rev. A **75** (2 2007) 023610, URL: https://link.aps.org/doi/10.1103/PhysRevA.75.023610 (cit. on pp. 6, 28, 73, 94, 95, 117, 119).

[39] P. Pieri, L. Pisani and G. C. Strinati, *BCS-BEC crossover at finite temperature in the broken-symmetry phase*, Phys. Rev. B **70** (9 2004) 094508, URL: https://link.aps.org/doi/10.1103/PhysRevB.70.094508 (cit. on pp. 6, 117, 119).

[40] A. Schirotzek et al., *Determination of the Superfluid Gap in Atomic Fermi Gases by Quasiparticle Spectroscopy*, Phys. Rev. Lett. **101** (14 2008) 140403, URL: https://link.aps.org/doi/10.1103/PhysRevLett.101.140403 (cit. on pp. 6, 117).

[41] S. Hoinka et al., *Goldstone mode and pair-breaking excitations in atomic Fermi superfluids*, Nature Physics **13** (2017) 943–946, URL: https://doi.org/10.1038/nphys4187 (cit. on pp. 6, 117, 119).

[42] D. Pekker and C. Varma, *Amplitude/Higgs Modes in Condensed Matter Physics*, Annual Review of Condensed Matter Physics **6** (2015) 269–297, eprint: https://doi.org/10.1146/annurev-conmatphys-031214-014350, URL: https://doi.org/10.1146/annurev-conmatphys-031214-014350 (cit. on pp. 6, 105–107).

[43] E. A. Yuzbashyan et al., *Quantum quench phase diagrams of an s-wave BCS-BEC condensate*, Phys. Rev. A **91** (3 2015) 033628, URL: https://link.aps.org/doi/10.1103/PhysRevA.91.033628 (cit. on pp. 6, 121–123, 134, 135).

[44] C. Jarzynski, *Diverse phenomena, common themes*, Nature Physics **11** (2015) 105–107, URL: https://doi.org/10.1038/nphys3229 (cit. on pp. 6, 121).

[45] D. Fausti et al., *Light-Induced Superconductivity in a Stripe-Ordered Cuprate*, Science **331** (2011) 189–191, ISSN: 0036-8075, eprint: https://science.sciencemag.org/content/331/6014/189.full.pdf, URL: https://science.sciencemag.org/content/331/6014/189 (cit. on pp. 6, 121).

[46] M. Mitrano et al., *Possible light-induced superconductivity in K3C60 at high temperature*, Nature **530** (2016) 461–464, URL: https://doi.org/10.1038/nature16522 (cit. on pp. 6, 121).

[47] P. B. Wigley et al., *Fast machine-learning online optimization of ultra-cold-atom experiments*, Scientific Reports **6** (2016) 25890, URL: https://doi.org/10.1038/srep25890 (cit. on pp. 7, 34).

[48] A. D. Tranter et al., *Multiparameter optimisation of a magneto-optical trap using deep learning*, Nature Communications **9** (2018) 4360, URL: https://doi.org/10.1038/s41467-018-06847-1 (cit. on pp. 7, 34).

[49] A. J. Barker et al., *Applying machine learning optimization methods to the production of a quantum gas*, Machine Learning: Science and Technology **1** (2020) 015007, URL: https://doi.org/10.1088%2F2632-2153%2Fab6432 (cit. on pp. 7, 34).

[50] Y. Wu et al., *Active Learning Approach to Optimization of Experimental Control*, 2020, arXiv: 2003.11804 [cond-mat.quant-gas] (cit. on pp. 7, 34).

[51] W. Ketterle and M. W. Zwierlein, "Making, probing and understanding ultracold Fermi gases", *Ultracold Fermi Gases, Proceedings of the International School of Physics "Enrico Fermi"*, ed. by M. Inguscio, W. Ketterle and C. Salomon, IOS Press, Amsterdam, 2006 (cit. on pp. 9, 13, 15, 17, 19, 20, 22, 23, 26, 27, 31, 67, 72, 74, 75).

[52] I. Bloch, J. Dalibard and W. Zwerger, *Many-body physics with ultracold gases*, Rev. Mod. Phys. **80** (3 2008) 885–964, URL: https://link.aps.org/doi/10.1103/RevModPhys.80.885 (cit. on pp. 9, 19).

[53] S. Giorgini, L. P. Pitaevskii and S. Stringari, *Theory of ultracold atomic Fermi gases*, Rev. Mod. Phys. **80** (4 2008) 1215–1274, URL: https://link.aps.org/doi/10.1103/RevModPhys.80.1215 (cit. on pp. 9, 17).

[54] J. T. M. Walraven, *Quantum Gases*, lecture notes, 2018 (cit. on pp. 9, 13–15).

[55] E. Fermi, *Zur Quantelung des idealen einatomigen Gases*, Zeitschrift für Physik **36** (1926) 902–912, URL: https://doi.org/10.1007/BF01400221 (cit. on p. 10).

[56] M. Köhl, *Atomic, Molecular and Optical Physics*, lecture notes (cit. on pp. 13, 15, 16, 19, 20, 23, 87).

[57] L. Landau and E. Lifshitz, *Quantum Mechanics: Non-relativistic Theory*, Pergamon Press, New York, 1987 (cit. on pp. 13–15).

[58] U. Fano, *Effects of Configuration Interaction on Intensities and Phase Shifts*, Phys. Rev. **124** (6 1961) 1866–1878, URL: https://link.aps.org/doi/10.1103/PhysRev.124.1866 (cit. on p. 15).

[59] H. Feshbach, *Unified theory of nuclear reactions*, Annals of Physics **5** (1958) 357 –390, ISSN: 0003-4916, URL: http://www.sciencedirect.com/science/article/pii/0003491658900071 (cit. on p. 15).

[60] C. Chin et al., *Feshbach resonances in ultracold gases*, Rev. Mod. Phys. **82** (2 2010) 1225–1286, URL: https://link.aps.org/doi/10.1103/RevModPhys.82.1225 (cit. on pp. 15, 19).

[61] M. Bartenstein et al., *Precise Determination of* ^{6}Li *Cold Collision Parameters by Radio-Frequency Spectroscopy on Weakly Bound Molecules*, Phys. Rev. Lett. **94** (10 2005) 103201, URL: https://link.aps.org/doi/10.1103/PhysRevLett.94.103201 (cit. on p. 17).

[62] G. Zürn et al., *Precise Characterization of* ^{6}Li *Feshbach Resonances Using Trap-Sideband-Resolved RF Spectroscopy of Weakly Bound Molecules*, Phys. Rev. Lett. **110** (13 2013) 135301, URL: https://link.aps.org/doi/10.1103/PhysRevLett.110.135301 (cit. on p. 17).

[63] G. Zürn et al., *Precise Characterization of* ^{6}Li *Feshbach Resonances Using Trap-Sideband-Resolved RF Spectroscopy of Weakly Bound Molecules*, Phys. Rev. Lett. **110** (13 2013) 135301, URL: https://link.aps.org/doi/10.1103/PhysRevLett.110.135301 (cit. on p. 18).

[64] D. S. Petrov, C. Salomon and G. V. Shlyapnikov,
Weakly Bound Dimers of Fermionic Atoms, Phys. Rev. Lett. **93** (9 2004) 090404,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.93.090404
(cit. on pp. 18, 19).

[65] S. Jochim et al., *Bose-Einstein Condensation of Molecules*,
Science **302** (2003) 2101–2103, ISSN: 0036-8075,
eprint: https://science.sciencemag.org/content/302/5653/2101.full.pdf,
URL: https://science.sciencemag.org/content/302/5653/2101
(cit. on pp. 18, 19, 71).

[66] L. N. Cooper, *Bound Electron Pairs in a Degenerate Fermi Gas*,
Phys. Rev. **104** (4 1956) 1189–1190,
URL: https://link.aps.org/doi/10.1103/PhysRev.104.1189 (cit. on p. 19).

[67] R. Grimm, *Ultracold Fermi gases in the BEC-BCS crossover: a review from the Innsbruck perspective*, 2007, arXiv: cond-mat/0703091 [cond-mat.other]
(cit. on p. 19).

[68] D. S. Petrov,
Three-body problem in Fermi gases with short-range interparticle interaction,
Phys. Rev. A **67** (1 2003) 010703,
URL: https://link.aps.org/doi/10.1103/PhysRevA.67.010703 (cit. on p. 19).

[69] D. S. Petrov, C Salomon and G. V. Shlyapnikov,
Diatomic molecules in ultracold Fermi gases—novel composite bosons,
Journal of Physics B: Atomic, Molecular and Optical Physics **38** (2005) S645–S660,
URL: https://doi.org/10.1088%2F0953-4075%2F38%2F9%2F014 (cit. on p. 19).

[70] N. N. Bogoljubov, *On a new method in the theory of superconductivity*,
Il Nuovo Cimento (1955-1965) **7** (1958) 794–805,
URL: https://doi.org/10.1007/BF02745585 (cit. on p. 22).

[71] J. G. Valatin, *Comments on the theory of superconductivity*,
Il Nuovo Cimento (1955-1965) **7** (1958) 843–857,
URL: https://doi.org/10.1007/BF02745589 (cit. on p. 22).

[72] L. Gor'kov and T. Melik-Barkhudarov,
Contribution to the Theory of Superfluidity in an Imperfect Fermi Gas,
JETP **13** (1960) 1018 (cit. on pp. 23, 24, 27, 73, 94, 95).

[73] M. Marini, F. Pistolesi and G. C. Strinati, *Evolution from BCS superconductivity to Bose condensation: analytic results for the crossover in three dimensions*,
The European Physical Journal B - Condensed Matter and Complex Systems **1**
(1998) 151–159, URL: https://doi.org/10.1007/s100510050165 (cit. on p. 24).

[74] M. Drechsler and W. Zwerger,
Crossover from BCS-superconductivity to Bose-condensation,
Annalen der Physik **1** (1992) 15–23 (cit. on pp. 26, 28).

[75] P. Nozières and S. Schmitt-Rink, *Bose condensation in an attractive fermion gas: From weak to strong coupling superconductivity*,
Journal of Low Temperature Physics **59** (1985) 195–211,
URL: https://doi.org/10.1007/BF00683774 (cit. on pp. 28, 30).

[76] W. Ketterle, D. S. Durfee and D. M. Stamper-Kurn,
Making, probing and understanding Bose-Einstein condensates, 1999,
arXiv: cond-mat/9904034 [cond-mat] (cit. on p. 28).

[77] V. A. Kashurnikov, N. V. Prokof'ev and B. V. Svistunov,
Critical Temperature Shift in Weakly Interacting Bose Gas,
Phys. Rev. Lett. **87** (12 2001) 120402,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.87.120402
(cit. on p. 28).

[78] P. Arnold and G. Moore,
BEC Transition Temperature of a Dilute Homogeneous Imperfect Bose Gas,
Phys. Rev. Lett. **87** (12 2001) 120401,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.87.120401
(cit. on p. 28).

[79] A. Bulgac, J. E. Drut and P. Magierski,
Spin 1/2 *Fermions in the Unitary Regime: A Superfluid of a New Type*,
Phys. Rev. Lett. **96** (9 2006) 090404,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.96.090404
(cit. on p. 28).

[80] E. Burovski et al.,
Critical Temperature and Thermodynamics of Attractive Fermions at Unitarity,
Phys. Rev. Lett. **96** (16 2006) 160402,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.96.160402
(cit. on p. 28).

[81] L. Salasnich, N. Manini and A. Parola,
Condensate fraction of a Fermi gas in the BCS-BEC crossover,
Phys. Rev. A **72** (2 2005) 023621,
URL: https://link.aps.org/doi/10.1103/PhysRevA.72.023621
(cit. on pp. 29, 30).

[82] C. Campbell, "Condensed Matter Theories", ed. by J. Clark and P. Panat, vol. 12,
Nova Science, New York, 1997 131 (cit. on p. 29).

[83] A. Perali, P. Pieri and G. C. Strinati, *Shrinking of a condensed fermionic cloud in a trap approaching the Bose-Einstein condensation limit*,
Phys. Rev. A **68** (3 2003) 031601,
URL: https://link.aps.org/doi/10.1103/PhysRevA.68.031601 (cit. on p. 30).

[84] J. R. Engelbrecht, M. Randeria and C. A. R. Sáde Melo,
BCS to Bose crossover: Broken-symmetry state,
Phys. Rev. B **55** (22 1997) 15153–15156,
URL: https://link.aps.org/doi/10.1103/PhysRevB.55.15153 (cit. on p. 30).

[85] W. S. McCulloch and W. Pitts,
A logical calculus of the ideas immanent in nervous activity,
The bulletin of mathematical biophysics **5** (1943) 115–133,
URL: https://doi.org/10.1007/BF02478259 (cit. on p. 33).

[86] D. Hebb, *The organization of behavior. A neuropsychological theory.*
Erlbaum Books, 1949 (cit. on p. 33).

[87] F. Rosenblatt, *A Perceiving and Recognizing Automaton*, tech. rep.,
Cornell Aeronautical Laboratory, 1957 (cit. on pp. 33, 38).

[88] M. Minsky and S. A. Papert,
Perceptrons: An Introduction to Computational Geometry, Expanded Edition,
MIT Press, 1987 (cit. on p. 33).

[89] S. Linnainmaa, *The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors*,
MA thesis: University of Helsinki, 1970 (cit. on p. 33).

[90] P. Werbos,
Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences,
PhD thesis: Harvard University, 1974 (cit. on p. 33).

[91] D. E. Rumelhart, G. E. Hinton and R. J. Williams,
Learning representations by back-propagating errors, Nature **323** (1986) 533–536,
URL: https://doi.org/10.1038/323533a0 (cit. on pp. 34, 42).

[92] D. E. Rumelhart, G. E. Hinton and R. J. Williams,
Parallel Distributed Processing: Explorations in the Microstructure of Cognition,
vol. Volume 1: Foundations, MIT Press, 1986 (cit. on pp. 34, 42).

[93] K. Hornik, M. Stinchcombe and H. White,
Multilayer feedforward networks are universal approximators,
Neural Networks **2** (1989) 359 –366, ISSN: 0893-6080,
URL: http://www.sciencedirect.com/science/article/pii/0893608089900208
(cit. on pp. 34, 39, 150).

[94] G. Cybenko, *Approximation by superpositions of a sigmoidal function*,
Mathematics of Control, Signals and Systems **2** (1989) 303–314,
URL: https://doi.org/10.1007/BF02551274 (cit. on pp. 34, 39, 150).

[95] Y. LeCun et al., *Backpropagation Applied To Handwritten Zip Code Recognition*,
Neural Computation **1** (1989) 541–551 (cit. on pp. 34, 40).

[96] Y. LeCun et al., *Handwritten Digit Recognition with a Back-Propagation Network*,
Proc. Advances in Neural Information Processing Systems 2 (1990) 396–404
(cit. on pp. 34, 40).

[97] A. Krizhevsky, I. Sutskever and G. Hinton,
ImageNet Classification with Deep Convolutional Neural Networks,
Neural Information Processing Systems **25** (2012) (cit. on p. 34).

[98] V. Mnih et al., *Human-level control through deep reinforcement learning*, Nature **518** (2015) 529–533, URL: https://doi.org/10.1038/nature14236 (cit. on p. 34).

[99] D. Silver et al., *Mastering the game of Go with deep neural networks and tree search*, Nature **529** (2016) 484–489, URL: https://doi.org/10.1038/nature16961 (cit. on p. 34).

[100] O. Vinyals et al., *Grammar as a Foreign Language*, 2014, arXiv: 1412.7449 [cs.CL] (cit. on p. 34).

[101] R. Jozefowicz et al., *Exploring the Limits of Language Modeling*, 2016, arXiv: 1602.02410 [cs.CL] (cit. on p. 34).

[102] G. Torlai et al., *Neural-network quantum state tomography*, Nature Physics **14** (2018) 447–450, URL: https://doi.org/10.1038/s41567-018-0048-5 (cit. on p. 34).

[103] E. P. L. van Nieuwenburg, Y.-H. Liu and S. D. Huber, *Learning phase transitions by confusion*, Nature Physics **13** (2017) 435–439, URL: https://doi.org/10.1038/nphys4037 (cit. on pp. 34, 73).

[104] G. Carleo and M. Troyer, *Solving the quantum many-body problem with artificial neural networks*, Science **355** (2017) 602–606, ISSN: 0036-8075, eprint: https://science.sciencemag.org/content/355/6325/602.full.pdf, URL: https://science.sciencemag.org/content/355/6325/602 (cit. on pp. 34, 73).

[105] M. A. Nielsen, *Neural Networks and Deep Learning*, Determination Press, 2015 (cit. on pp. 34, 43, 44).

[106] X. Glorot, A. Bordes and Y. Bengio, *Deep Sparse Rectifier Neural Networks*, Proceedings of the 14th International Conference on Artificial Intelligence and Statistics **15** (2011) 315–323 (cit. on p. 38).

[107] D. H. Hubel and T. N. Wiesel, *Receptive fields and functional architecture of monkey striate cortex*, The Journal of Physiology **195** (1968) 215–243, eprint: https://physoc.onlinelibrary.wiley.com/doi/pdf/10.1113/jphysiol.1968.sp008455, URL: https://physoc.onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.1968.sp008455 (cit. on p. 40).

[108] K. Fukushima, *Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position*, Biological Cybernetics **36** (1980) 193–202, URL: https://doi.org/10.1007/BF00344251 (cit. on p. 40).

[109] Y. LeCun, Y. Bengio and G. Hinton, *Deep learning*, Nature **521** (2015) 436–444, URL: https://doi.org/10.1038/nature14539 (cit. on pp. 40–42).

[110] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, `http://www.deeplearningbook.org`, MIT Press, 2016 (cit. on pp. 41, 42, 45, 46, 48, 103).

[111] Zhou and Chellappa, "Computation of optical flow using a neural network", *IEEE 1988 International Conference on Neural Networks*, 1988 71–78 vol.2 (cit. on p. 41).

[112] Y. Lecun, "Une procedure d'apprentissage pour reseau a seuil asymmetrique (A learning scheme for asymmetric threshold networks)", English (US), *Proceedings of Cognitiva 85, Paris, France*, 1985 599–604 (cit. on p. 42).

[113] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, 2014, arXiv: `1412.6980 [cs.LG]` (cit. on pp. 44, 45, 77, 189, 190).

[114] V. Bushaev, *Adam - latest trends in deep learning optimization.* (Cit. on p. 45).

[115] G. E. Hinton et al., *Improving neural networks by preventing co-adaptation of feature detectors*, 2012, arXiv: `1207.0580 [cs.NE]` (cit. on p. 45).

[116] N. Srivastava et al., *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*, Journal of Machine Learning Research **15** (2014) 1929–1958 (cit. on p. 45).

[117] O. Simeone, *A Very Brief Introduction to Machine Learning With Applications to Communication Systems*, 2018, arXiv: `1808.02342 [cs.IT]` (cit. on p. 46).

[118] K. Pearson, *On Lines and Planes of Closest Fit to Systems of Points in Space*, Philosophical Magazine **2** (1901) 559–572 (cit. on p. 46).

[119] H. Hotelling, *Relations Between Two Sets of Variates*, Biometrika **28** (1936) 321–377 (cit. on p. 46).

[120] L. I. Smith, *A tutorial on Principal Components Analysis* (cit. on p. 46).

[121] I. Jolliffe, *Principal Component Analysis*, 2nd edition, Springer Verlag, 2002 (cit. on p. 46).

[122] M. A. Kramer, *Nonlinear principal component analysis using autoassociative neural networks*, AIChE Journal **37** (1991) 233–243, eprint: `https://aiche.onlinelibrary.wiley.com/doi/pdf/10.1002/aic.690370209`, URL: `https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/aic.690370209` (cit. on p. 48).

[123] G. E. Hinton and R. R. Salakhutdinov, *Reducing the Dimensionality of Data with Neural Networks*, Science **313** (2006) 504–507, ISSN: 0036-8075, eprint: `https://science.sciencemag.org/content/313/5786/504.full.pdf`, URL: `https://science.sciencemag.org/content/313/5786/504` (cit. on pp. 48–50, 103).

[124] H. Bourlard,
Auto-Association by Multilayer Perceptrons and Singular Value Decomposition,
(2000), URL: http://infoscience.epfl.ch/record/82601 (cit. on p. 49).

[125] D. Chicco, P. Sadowski and P. Baldi,
"Deep Autoencoder Neural Networks for Gene Ontology Annotation Predictions",
Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, BCB '14, Association for Computing Machinery, 2014
533–540, ISBN: 9781450328944, URL: https://doi.org/10.1145/2649387.2649442
(cit. on p. 49).

[126] E. Plaut,
From Principal Subspaces to Principal Components with Linear Autoencoders, 2018,
arXiv: 1804.10253 [stat.ML] (cit. on p. 49).

[127] P. Vincent et al., *Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion*,
Journal of Machine Learning Research **11** (2010) 3371–3408 (cit. on p. 49).

[128] Y. LeCun, C. Cortez and C. C. Burges, *The MNIST Handwritten Digit Database*,
1998, URL: yann.lecun.com (cit. on pp. 49, 50).

[129] A. B. Behrle, *Driving a Strongly Interacting Superfluid out of Equilibrium*,
PhD thesis: Universität Bonn, 2017 (cit. on pp. 51, 54, 57, 63, 105, 123).

[130] T. J. Harrison,
Measuring the Gap and investigating Non-Equilibrium in the BEC-BCS Crossover,
PhD thesis: Universität Bonn, 2017 (cit. on pp. 51, 54–57, 60, 63, 105, 123).

[131] C. A. Stan and W. Ketterle,
Multiple species atom source for laser-cooling experiments,
Review of Scientific Instruments **76** (2005) 063113,
eprint: https://doi.org/10.1063/1.1935433,
URL: https://doi.org/10.1063/1.1935433 (cit. on p. 51).

[132] T. Leuteritz, *High Resolution Imaging of an Ultracold Fermi Gas*,
Master thesis: Universität Bonn, 2016 (cit. on p. 52).

[133] S. Crane and C. R. Ekstrom, "Nonmagnetic UHV optical viewports", *Proceedings of the 2005 IEEE International Frequency Control Symposium and Exposition, 2005.*
2005 297–300 (cit. on p. 53).

[134] M. E. Gehm, *Properties of 6Li*, tech. rep., 2003 (cit. on pp. 53, 55).

[135] D. A. Steck, *Sodium D Line Data*, tech. rep., Los Alamos National Laboratory, 2003
(cit. on pp. 53, 55).

[136] S. C. Bell et al., *A slow atom source using a collimated effusive oven and a single-layer variable pitch coil Zeeman slower*,
Review of Scientific Instruments **81** (2010) 013105,
eprint: https://doi.org/10.1063/1.3276712,
URL: https://doi.org/10.1063/1.3276712 (cit. on p. 54).

[137] C. J. Foot, *Atomic Physics*, Oxford University Press, 2005 (cit. on p. 54).

[138] W. Ketterle et al.,
High densities of cold atoms in a dark spontaneous-force optical trap,
Phys. Rev. Lett. **70** (15 1993) 2253–2256,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.70.2253 (cit. on p. 54).

[139] M.-O. Mewes et al., *Simultaneous magneto-optical trapping of two lithium isotopes*,
Physical Review A **61** (1999) (cit. on p. 55).

[140] M. Okano et al., *Simultaneous magneto-optical trapping of lithium and ytterbium atoms towards production of ultracold polar molecules*,
Applied Physics B **98** (2010) 691–696,
URL: https://doi.org/10.1007/s00340-009-3728-0 (cit. on p. 55).

[141] J. Andrijauskas, *Optical dipole trap laser setup in a new quantum gas experiment*,
Master thesis: Universität Bonn, 2015 (cit. on p. 57).

[142] R. Grimm, M. Weidemüller and Y. B. Ovchinnikov,
Optical dipole traps for neutral atoms, 1999,
arXiv: physics/9902072 [physics.atom-ph] (cit. on p. 57).

[143] C. Pethick and H. Smith, *Bose-Einstein Condensation in Dilute Gases*,
Cambridge University Press, 2008 (cit. on p. 58).

[144] B. Mukherjee et al., *Homogeneous Atomic Fermi Gases*,
Phys. Rev. Lett. **118** (12 2017) 123401,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.118.123401
(cit. on pp. 59, 162).

[145] K. Hueck et al., *Two-Dimensional Homogeneous Fermi Gases*,
Phys. Rev. Lett. **120** (6 2018) 060402,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.120.060402
(cit. on pp. 59, 162, 163).

[146] P. A. Murthy et al.,
Matter-wave Fourier optics with a strongly interacting two-dimensional Fermi gas,
Phys. Rev. A **90** (4 2014) 043611,
URL: https://link.aps.org/doi/10.1103/PhysRevA.90.043611
(cit. on pp. 62, 113, 125).

[147] G. Reinaudi et al.,
Strong saturation absorption imaging of dense clouds of ultracold atoms,
Opt. Lett. **32** (2007) 3143–3145,
URL: http://ol.osa.org/abstract.cfm?URI=ol-32-21-3143 (cit. on pp. 63, 65).

[148] L. A. Miller, *Ultracold Fermions in Two-Dimensional Optical Lattices: Quantum Simulation of the Hubbard Model*, PhD thesis: Cambridge University, 2016
(cit. on p. 63).

[149] E. Cocchi, *Analogue Quantum Simulation of the Two-Dimensional Hubbard Model with Ultracold Fermions*, PhD thesis: Cambridge University, 2016 (cit. on p. 63).

[150] A. Perali, P. Pieri and G. C. Strinati,
Extracting the Condensate Density from Projection Experiments with Fermi Gases,
Phys. Rev. Lett. **95** (1 2005) 010407,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.95.010407
(cit. on p. 67).

[151] E. Altman and A. Vishwanath, *Dynamic Projection on Feshbach Molecules: A Probe of Pairing and Phase Fluctuations*, Phys. Rev. Lett. **95** (11 2005) 110404,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.95.110404
(cit. on p. 67).

[152] E. A. Yuzbashyan et al., *Nonequilibrium cooper pairing in the nonadiabatic regime*,
Phys. Rev. B **72** (22 2005) 220503,
URL: https://link.aps.org/doi/10.1103/PhysRevB.72.220503 (cit. on p. 67).

[153] M. W. Zwierlein et al., *Formation Dynamics of a Fermion Pair Condensate*,
Phys. Rev. Lett. **94** (18 2005) 180401,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.94.180401
(cit. on p. 68).

[154] J. Cubizolles et al.,
Production of Long-Lived Ultracold Li_2 *Molecules from a Fermi Gas*,
Phys. Rev. Lett. **91** (24 2003) 240401,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.91.240401
(cit. on p. 71).

[155] K. E. Strecker, G. B. Partridge and R. G. Hulet,
Conversion of an Atomic Fermi Gas to a Long-Lived Molecular Bose Gas,
Phys. Rev. Lett. **91** (8 2003) 080406,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.91.080406
(cit. on p. 71).

[156] S. Jochim et al.,
Pure Gas of Optically Trapped Molecules Created from Fermionic Atoms,
Phys. Rev. Lett. **91** (24 2003) 240402,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.91.240402
(cit. on p. 71).

[157] M. W. Zwierlein et al., *Observation of Bose-Einstein Condensation of Molecules*,
Phys. Rev. Lett. **91** (25 2003) 250401,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.91.250401
(cit. on p. 71).

[158] M. Bartenstein et al.,
Crossover from a Molecular Bose-Einstein Condensate to a Degenerate Fermi Gas,
Phys. Rev. Lett. **92** (12 2004) 120401,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.92.120401
(cit. on p. 71).

[159] T. Bourdel et al.,
Experimental Study of the BEC-BCS Crossover Region in Lithium 6,
Phys. Rev. Lett. **93** (5 2004) 050401,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.93.050401
(cit. on p. 71).

[160] G. B. Partridge et al., *Molecular Probe of Pairing in the BEC-BCS Crossover*,
Phys. Rev. Lett. **95** (2 2005) 020404,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.95.020404
(cit. on p. 71).

[161] M. W. Zwierlein et al.,
Direct observation of the superfluid phase transition in ultracold Fermi gases,
Nature **442** (2006) 54–58, URL: https://doi.org/10.1038/nature04936
(cit. on p. 72).

[162] R. B. Diener and T.-L. Ho,
Projecting Fermion Pair Condensates into Molecular Condensates, 2004,
arXiv: cond-mat/0404517 [cond-mat.supr-con] (cit. on p. 72).

[163] A. V. Avdeenkov and J. L. Bohn, *Pair wave functions in atomic Fermi condensates*,
Phys. Rev. A **71** (2 2005) 023609,
URL: https://link.aps.org/doi/10.1103/PhysRevA.71.023609 (cit. on p. 72).

[164] Q. Chen et al., *Understanding the superfluid phase diagram in trapped Fermi gases*,
Phys. Rev. A **73** (4 2006) 041601,
URL: https://link.aps.org/doi/10.1103/PhysRevA.73.041601 (cit. on p. 72).

[165] G. Baym et al., *The Transition Temperature of the Dilute Interacting Bose Gas*,
Phys. Rev. Lett. **83** (9 1999) 1703–1706,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.83.1703
(cit. on pp. 73, 94, 95).

[166] Baym, G., Blaizot, J.-P. and Zinn-Justin, J.,
The transition temperature of the dilute interacting Bose gas for N internal states,
Europhys. Lett. **49** (2000) 150–155,
URL: https://doi.org/10.1209/epl/i2000-00130-3 (cit. on pp. 73, 94, 95).

[167] G. Baym et al., *Bose-Einstein transition in a dilute interacting gas*, The European
Physical Journal B - Condensed Matter and Complex Systems **24** (2001) 107–124,
URL: https://doi.org/10.1007/s100510170028 (cit. on pp. 73, 94, 95).

[168] E. Burovski et al., *Critical Temperature Curve in BEC-BCS Crossover*,
Physical Review Letters **101** (2008), ISSN: 1079-7114,
URL: http://dx.doi.org/10.1103/PhysRevLett.101.090402
(cit. on pp. 73, 94, 95, 102).

[169] M. Buchanan, *The power of machine learning*, Nature Physics **15** (2019) 1208–1208,
URL: https://doi.org/10.1038/s41567-019-0737-8 (cit. on p. 73).

[170] P. Huembeli, A. Dauphin and P. Wittek,
Identifying quantum phase transitions with adversarial neural networks,
Phys. Rev. B **97** (13 2018) 134109,
URL: https://link.aps.org/doi/10.1103/PhysRevB.97.134109 (cit. on p. 73).

[171] L. Wang, *Discovering phase transitions with unsupervised learning*,
Physical Review B **94** (2016), ISSN: 2469-9969,
URL: http://dx.doi.org/10.1103/PhysRevB.94.195105 (cit. on p. 73).

[172] X. Gao and L.-M. Duan,
Efficient representation of quantum many-body states with deep neural networks,
Nature Communications **8** (2017), ISSN: 2041-1723,
URL: http://dx.doi.org/10.1038/s41467-017-00705-2 (cit. on p. 73).

[173] G. Torlai et al.,
Integrating Neural Networks with a Quantum Simulator for State Reconstruction,
Physical Review Letters **123** (2019), ISSN: 1079-7114,
URL: http://dx.doi.org/10.1103/PhysRevLett.123.230504 (cit. on p. 73).

[174] M. Abadi et al.,
TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems,
Software available from tensorflow.org, 2015, URL: https://www.tensorflow.org/
(cit. on pp. 77, 189, 190).

[175] J. Kinast et al., *Heat Capacity of a Strongly Interacting Fermi Gas*,
Science **307** (2005) 1296–1299, ISSN: 0036-8075,
eprint: https://science.sciencemag.org/content/307/5713/1296.full.pdf,
URL: https://science.sciencemag.org/content/307/5713/1296 (cit. on p. 78).

[176] E. D. Kuhnle et al.,
Temperature Dependence of the Universal Contact Parameter in a Unitary Fermi Gas,
Phys. Rev. Lett. **106** (17 2011) 170402,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.106.170402
(cit. on p. 78).

[177] A. Shrikumar, P. Greenside and A. Kundaje,
Learning Important Features Through Propagating Activation Differences, 2017,
arXiv: 1704.02685 [cs.CV] (cit. on pp. 81, 82).

[178] L. H. Gilpin et al.,
Explaining Explanations: An Overview of Interpretability of Machine Learning, 2018,
arXiv: 1806.00069 [cs.AI] (cit. on p. 81).

[179] R. S. Lous et al.,
Thermometry of a deeply degenerate Fermi gas with a Bose-Einstein condensate,
Phys. Rev. A **95** (5 2017) 053627,
URL: https://link.aps.org/doi/10.1103/PhysRevA.95.053627 (cit. on p. 85).

[180] T.-L. Ho,
Universal Thermodynamics of Degenerate Quantum Gases in the Unitarity Limit,
Phys. Rev. Lett. **92** (9 2004) 090402,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.92.090402
(cit. on p. 88).

[181] A. Perali et al.,
BCS-BEC Crossover at Finite Temperature for Superfluid Trapped Fermi Atoms,
Phys. Rev. Lett. **92** (22 2004) 220404,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.92.220404
(cit. on p. 94).

[182] R. Salakhutdinov and G. Hinton, *Semantic hashing*,
International Journal of Approximate Reasoning **50** (2009) 969 –978, Special Section on Graphical Models and Information Retrieval, ISSN: 0888-613X, URL:
http://www.sciencedirect.com/science/article/pii/S0888613X08001813
(cit. on p. 100).

[183] J. Kombe, *Nonequilibrium Dynamics of Correlated Fermi Gases*,
PhD thesis: Universität Bonn, 2020 (cit. on pp. 105, 128).

[184] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*,
Phys. Rev. Lett. **13** (16 1964) 508–509,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508 (cit. on p. 106).

[185] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*,
Phys. Rev. Lett. **13** (9 1964) 321–323,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321 (cit. on p. 106).

[186] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble,
Global Conservation Laws and Massless Particles,
Phys. Rev. Lett. **13** (20 1964) 585–587,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.585 (cit. on p. 106).

[187] R. Shimano and N. Tsuji, *Higgs Mode in Superconductors*,
Annual Review of Condensed Matter Physics **11** (2020) 103–124, ISSN: 1947-5462,
URL: http://dx.doi.org/10.1146/annurev-conmatphys-031119-050813
(cit. on p. 107).

[188] D. Podolsky, A. Auerbach and D. P. Arovas,
Visibility of the amplitude (Higgs) mode in condensed matter,
Phys. Rev. B **84** (17 2011) 174522,
URL: https://link.aps.org/doi/10.1103/PhysRevB.84.174522 (cit. on p. 107).

[189] R. G. Scott et al.,
Rapid ramps across the BEC-BCS crossover: A route to measuring the superfluid gap,
Phys. Rev. A **86** (5 2012) 053604,
URL: https://link.aps.org/doi/10.1103/PhysRevA.86.053604
(cit. on pp. 107, 115, 131, 142, 160).

[190] Y. Barlas and C. M. Varma, *Amplitude or Higgs modes in d-wave superconductors*, Phys. Rev. B **87** (5 2013) 054503, URL: https://link.aps.org/doi/10.1103/PhysRevB.87.054503 (cit. on p. 107).

[191] B. Liu, H. Zhai and S. Zhang, *Evolution of the Higgs mode in a fermion superfluid with tunable interactions*, Phys. Rev. A **93** (3 2016) 033641, URL: https://link.aps.org/doi/10.1103/PhysRevA.93.033641 (cit. on p. 107).

[192] X. Han, B. Liu and J. Hu, *Observability of Higgs mode in a system without Lorentz invariance*, Phys. Rev. A **94** (3 2016) 033608, URL: https://link.aps.org/doi/10.1103/PhysRevA.94.033608 (cit. on p. 107).

[193] C. Rüegg et al., *Quantum Magnets under Pressure: Controlling Elementary Excitations in* $TlCuCl_3$, Phys. Rev. Lett. **100** (20 2008) 205701, URL: https://link.aps.org/doi/10.1103/PhysRevLett.100.205701 (cit. on p. 107).

[194] W. HALPERIN and E. VAROQUAUX, "CHAPTER 7 - Order-Parameter Collective Modes in Superfluid 3He", *Helium Three*, ed. by W. HALPERIN and L. PITAEVSKII, vol. 26, Modern Problems in Condensed Matter Sciences, Elsevier, 1990 353 –522, URL: http://www.sciencedirect.com/science/article/pii/B9780444874764500133 (cit. on p. 107).

[195] U. Bissbort et al., *Detecting the Amplitude Mode of Strongly Interacting Lattice Bosons by Bragg Scattering*, Phys. Rev. Lett. **106** (20 2011) 205303, URL: https://link.aps.org/doi/10.1103/PhysRevLett.106.205303 (cit. on p. 107).

[196] M. Endres et al., *The 'Higgs'amplitude mode at the two-dimensional superfluid/Mott insulator transition*, Nature **487** (2012) 454–458, URL: https://doi.org/10.1038/nature11255 (cit. on p. 107).

[197] T. M. Hoang et al., *Adiabatic quenches and characterization of amplitude excitations in a continuous quantum phase transition*, Proceedings of the National Academy of Sciences **113** (2016) 9475–9479, ISSN: 0027-8424, eprint: https://www.pnas.org/content/113/34/9475.full.pdf, URL: https://www.pnas.org/content/113/34/9475 (cit. on p. 107).

[198] J. Léonard et al., *Monitoring and manipulating Higgs and Goldstone modes in a supersolid quantum gas*, Science **358** (2017) 1415–1418 (cit. on p. 107).

[199] M. Greiner, C. A. Regal and D. S. Jin, *Probing the Excitation Spectrum of a Fermi Gas in the BCS-BEC Crossover Regime*, Phys. Rev. Lett. **94** (7 2005) 070403, URL: https://link.aps.org/doi/10.1103/PhysRevLett.94.070403 (cit. on p. 107).

[200] G. M. Bruun, *Low-Energy Monopole Modes of a Trapped Atomic Fermi Gas*, Phys. Rev. Lett. **89** (26 2002) 263002, URL: https://link.aps.org/doi/10.1103/PhysRevLett.89.263002 (cit. on p. 115).

[201] A. Korolyuk, J. J. Kinnunen and P. Törmä, *Density response of a trapped Fermi gas: A crossover from the pair vibration mode to the Goldstone mode*, Phys. Rev. A **84** (3 2011) 033623, URL: https://link.aps.org/doi/10.1103/PhysRevA.84.033623 (cit. on p. 115).

[202] A. Korolyuk, J. J. Kinnunen and P. Törmä, *Collective excitations of a trapped Fermi gas at finite temperature*, Phys. Rev. A **89** (1 2014) 013602, URL: https://link.aps.org/doi/10.1103/PhysRevA.89.013602 (cit. on p. 115).

[203] J. Tokimoto, S. Tsuchiya and T. Nikuni, *Higgs Mode in a Trapped Superfluid Fermi Gas*, Journal of Low Temperature Physics **187** (2017) 765–770, URL: https://doi.org/10.1007/s10909-017-1766-2 (cit. on p. 115).

[204] K. Gao et al., *A compact and fast magnetic coil for the interaction manipulation of quantum gases with Feshbach resonances*, Preliminary, 2020 (cit. on pp. 121, 131).

[205] E. Schrödinger, *What is Life?*, Cambridge University Press, 1944 (cit. on p. 121).

[206] M. C. Marchetti et al., *Hydrodynamics of soft active matter*, Rev. Mod. Phys. **85** (3 2013) 1143–1189, URL: https://link.aps.org/doi/10.1103/RevModPhys.85.1143 (cit. on p. 121).

[207] J. Prost, F. Jülicher and J.-F. Joanny, *Active gel physics*, Nature Physics **11** (2015) 111–117, URL: https://doi.org/10.1038/nphys3224 (cit. on p. 121).

[208] J. Eisert, M. Friesdorf and C. Gogolin, *Quantum many-body systems out of equilibrium*, Nature Physics **11** (2015) 124–130, ISSN: 1745-2481, URL: http://dx.doi.org/10.1038/nphys3215 (cit. on p. 121).

[209] A. Volkov and S. M. Kogan, *Collisionless relaxation of the energy gap in superconductors*, Journal of Experimental and Theoretical Physics **38** (1974) 1018 (cit. on p. 121).

[210] R. A. Barankov, L. S. Levitov and B. Z. Spivak, *Collective Rabi Oscillations and Solitons in a Time-Dependent BCS Pairing Problem*, Phys. Rev. Lett. **93** (16 2004) 160401, URL: https://link.aps.org/doi/10.1103/PhysRevLett.93.160401 (cit. on p. 121).

[211] R. A. Barankov and L. S. Levitov, *Synchronization in the BCS Pairing Dynamics as a Critical Phenomenon*, Phys. Rev. Lett. **96** (23 2006) 230403, URL: https://link.aps.org/doi/10.1103/PhysRevLett.96.230403 (cit. on pp. 121, 122).

[212] E. A. Yuzbashyan and M. Dzero,
Dynamical Vanishing of the Order Parameter in a Fermionic Condensate,
Phys. Rev. Lett. **96** (23 2006) 230404,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.96.230404
(cit. on pp. 121, 122).

[213] G. J. Milburn and C. A. Holmes,
Dissipative Quantum and Classical Liouville Mechanics of the Anharmonic Oscillator,
Phys. Rev. Lett. **56** (21 1986) 2237–2240,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.56.2237 (cit. on p. 121).

[214] E. A. Yuzbashyan, O. Tsyplyatyev and B. L. Altshuler, *Relaxation and Persistent Oscillations of the Order Parameter in Fermionic Condensates*,
Phys. Rev. Lett. **96** (9 2006) 097005,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.96.097005
(cit. on p. 122).

[215] M. J. Bhaseen et al.,
Holographic Superfluids and the Dynamics of Symmetry Breaking,
Phys. Rev. Lett. **110** (1 2013) 015301,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.110.015301
(cit. on p. 122).

[216] Y. Kagan, E. L. Surkov and G. V. Shlyapnikov,
Evolution of a Bose-condensed gas under variations of the confining potential,
Phys. Rev. A **54** (3 1996) R1753–R1756,
URL: https://link.aps.org/doi/10.1103/PhysRevA.54.R1753 (cit. on p. 127).

[217] Y. Castin and R. Dum, *Bose-Einstein Condensates in Time Dependent Traps*,
Phys. Rev. Lett. **77** (27 1996) 5315–5319,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.77.5315 (cit. on p. 127).

[218] M. A. Sentef et al., *Theory of light-enhanced phonon-mediated superconductivity*,
Phys. Rev. B **93** (14 2016) 144506,
URL: https://link.aps.org/doi/10.1103/PhysRevB.93.144506 (cit. on p. 128).

[219] J. Kombe et al., *Finite-duration interaction quench in dilute attractively interacting Fermi gases: Emergence of preformed pairs*, Phys. Rev. A **100** (1 2019) 013604,
URL: https://link.aps.org/doi/10.1103/PhysRevA.100.013604 (cit. on p. 128).

[220] URL: https://www.filamentworld.de/3d-druck-wissen/was-ist-pla/
(cit. on p. 133).

[221] J. T. Stewart, J. P. Gaebler and D. S. Jin,
Using photoemission spectroscopy to probe a strongly interacting Fermi gas,
Nature **454** (2008) 744–747, URL: https://doi.org/10.1038/nature07172
(cit. on p. 138).

[222] C. Chin et al., *Observation of the Pairing Gap in a Strongly Interacting Fermi Gas*,
Science **305** (2004) 1128–1130, ISSN: 0036-8075,
eprint: https://science.sciencemag.org/content/305/5687/1128.full.pdf,
URL: https://science.sciencemag.org/content/305/5687/1128 (cit. on p. 138).

[223] M. Punk and W. Zwerger, *Theory of rf-Spectroscopy of Strongly Interacting Fermions*, Phys. Rev. Lett. **99** (17 2007) 170404, URL: https://link.aps.org/doi/10.1103/PhysRevLett.99.170404 (cit. on p. 138).

[224] P. Kobel, PhD Thesis, work in progress (cit. on pp. 145, 151, 152).

[225] E. Piselli and A. Akroh, *New CERN Proton Synchrotron beam optimization tool*, (2018) TUPHA120. 5 p, URL: http://cds.cern.ch/record/2305963 (cit. on p. 145).

[226] M. Steiner, *A single ion inside a miniature cavity*, PhD thesis: University of Cambridge, 2014 (cit. on p. 151).

[227] H. M. Meyer, *A fibre-cavity based photonic interface for a single ion*, PhD thesis: University of Cambridge, 2014 (cit. on p. 151).

[228] M. Link, *A simplified laser system for trapping and cooling Ytterbium (+) ions*, MA thesis: Universität Bonn, 2015 (cit. on p. 151).

[229] R. Blatt and D. Wineland, *Entangled states of trapped atomic ions*, Nature **453** (2008) 1008–1015, ISSN: 1476-4687, URL: https://doi.org/10.1038/nature07125 (cit. on p. 152).

[230] J. Keller et al., *Precise determination of micromotion for trapped-ion optical clocks*, Journal of Applied Physics **118** (2015) 104501, eprint: https://doi.org/10.1063/1.4930037, URL: https://doi.org/10.1063/1.4930037 (cit. on p. 152).

[231] C. W. Chou et al., *Frequency Comparison of Two High-Accuracy* Al^+ *Optical Clocks*, Phys. Rev. Lett. **104** (7 2010) 070802, URL: https://link.aps.org/doi/10.1103/PhysRevLett.104.070802 (cit. on p. 152).

[232] T. Schneider, E. Peik and C. Tamm, *Sub-Hertz Optical Frequency Comparisons between Two Trapped* $^{171}Yb^+$ *Ions*, Phys. Rev. Lett. **94** (23 2005) 230801, URL: https://link.aps.org/doi/10.1103/PhysRevLett.94.230801 (cit. on p. 152).

[233] D. J. Berkeland et al., *Minimization of ion micromotion in a Paul trap*, Journal of Applied Physics **83** (1998) 5025–5033, eprint: https://doi.org/10.1063/1.367318, URL: https://doi.org/10.1063/1.367318 (cit. on p. 152).

[234] T. F. Gloger et al., *Ion-trajectory analysis for micromotion minimization and the measurement of small forces*, Phys. Rev. A **92** (4 2015) 043421, URL: https://link.aps.org/doi/10.1103/PhysRevA.92.043421 (cit. on p. 152).

[235] Y. Ibaraki, U. Tanaka and S. Urabe, *Detection of parametric resonance of trapped ions for micromotion compensation.*, Appl. Phys. B **105** (2011) 219–233, URL: https://doi.org/10.1007/s00340-011-4463-x (cit. on p. 152).

[236] S. Narayanan et al.,
Electric field compensation and sensing with a single ion in a planar trap,
Journal of Applied Physics **110** (2011) 114909,
eprint: https://doi.org/10.1063/1.3665647,
URL: https://doi.org/10.1063/1.3665647 (cit. on p. 152).

[237] U. Tanaka et al., *Micromotion compensation in a surface electrode trap by parametric excitation of trapped ions*, Appl. Phys. B **107** (2012) 907–912,
URL: https://doi.org/10.1007/s00340-013-5580-5 (cit. on p. 152).

[238] K. Pyka et al., *A high-precision segmented Paul trap with minimized micromotion for an optical multiple-ion clock*, Appl. Phys. B **114** (2014) 231–241,
URL: https://doi.org/10.1007/s00340-013-5580-5 (cit. on p. 152).

[239] M. Schubert et al., *Transient internal dynamics of a multilevel ion*,
Phys. Rev. A **52** (4 1995) 2994–3006,
URL: https://link.aps.org/doi/10.1103/PhysRevA.52.2994 (cit. on p. 152).

[240] N. D. Mermin and H. Wagner, *Absence of Ferromagnetism or Antiferromagnetism in One- or Two-Dimensional Isotropic Heisenberg Models*,
Phys. Rev. Lett. **17** (22 1966) 1133–1136,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.17.1133 (cit. on p. 160).

[241] P. C. Hohenberg, *Existence of Long-Range Order in One and Two Dimensions*,
Phys. Rev. **158** (2 1967) 383–386,
URL: https://link.aps.org/doi/10.1103/PhysRev.158.383 (cit. on p. 160).

[242] V. L. Berezinskiĭ, *Destruction of Long-range Order in One-dimensional and Two-dimensional Systems having a Continuous Symmetry Group I. Classical Systems*,
Soviet Journal of Experimental and Theoretical Physics **32** (1971) 493
(cit. on p. 160).

[243] V. L. Berezinskiĭ,
Destruction of Long-range Order in One-dimensional and Two-dimensional Systems Possessing a Continuous Symmetry Group. II. Quantum Systems,
Soviet Journal of Experimental and Theoretical Physics **34** (1972) 610
(cit. on p. 160).

[244] J. M. Kosterlitz and D. J. Thouless,
Ordering, metastability and phase transitions in two-dimensional systems,
Journal of Physics C: Solid State Physics **6** (1973) 1181–1203,
URL: https://doi.org/10.1088%2F0022-3719%2F6%2F7%2F010 (cit. on p. 160).

[245] Z. Hadzibabic et al.,
Berezinskii–Kosterlitz–Thouless crossover in a trapped atomic gas,
Nature **441** (2006) 1118–1121, URL: https://doi.org/10.1038/nature04851
(cit. on p. 160).

[246] D. J. Bishop and J. D. Reppy,
Study of the Superfluid Transition in Two-Dimensional ^{4}He *Films*,
Phys. Rev. Lett. **40** (26 1978) 1727–1730,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.40.1727 (cit. on p. 160).

[247] D. J. Resnick et al.,
Kosterlitz-Thouless Transition in Proximity-Coupled Superconducting Arrays,
Phys. Rev. Lett. **47** (21 1981) 1542–1545,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.47.1542 (cit. on p. 160).

[248] A. I. Safonov et al.,
Observation of Quasicondensate in Two-Dimensional Atomic Hydrogen,
Phys. Rev. Lett. **81** (21 1998) 4545–4548,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.81.4545 (cit. on p. 160).

[249] S. A. Weidinger and W. Zwerger, *Higgs mode and magnon interactions in 2D quantum antiferromagnets from Raman scattering*,
The European Physical Journal B **88** (2015) 237,
URL: https://doi.org/10.1140/epjb/e2015-60438-1 (cit. on p. 160).

[250] L. Salasnich, *Goldstone and Higgs Hydrodynamics in the BCS–BEC Crossover*,
Condensed Matter **2** (2017) 22, ISSN: 2410-3896,
URL: http://dx.doi.org/10.3390/condmat2020022 (cit. on p. 160).

[251] T. Hong et al., *Higgs amplitude mode in a two-dimensional quantum antiferromagnet near the quantum critical point*, Nature Physics **13** (2017) 638–642,
URL: https://doi.org/10.1038/nphys4182 (cit. on p. 160).

[252] P. Dyke et al., *Crossover from 2D to 3D in a Weakly Interacting Fermi Gas*,
Phys. Rev. Lett. **106** (10 2011) 105304,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.106.105304
(cit. on p. 160).

[253] A. Kell, *Optical Trap for Ultracold 2D Fermi Gas*, MA thesis: Universität Bonn, 2017 (cit. on p. 161).

[254] M. W. Zwierlein et al., *Fermionic Superfluidity with Imbalanced Spin Populations*,
Science **311** (2006) 492–496, ISSN: 0036-8075,
eprint: https://science.sciencemag.org/content/311/5760/492.full.pdf,
URL: https://science.sciencemag.org/content/311/5760/492 (cit. on p. 162).

[255] Y.-i. Shin et al.,
Phase diagram of a two-component Fermi gas with resonant interactions,
Nature **451** (2008) 689–693, URL: https://doi.org/10.1038/nature06473
(cit. on p. 162).

[256] A. Nahum and E. Bettelheim,
Dissipationless BCS dynamics with large branch imbalance,
Phys. Rev. B **78** (18 2008) 184510,
URL: https://link.aps.org/doi/10.1103/PhysRevB.78.184510 (cit. on p. 162).

[257] C. K. Chung and C. K. Law,
Pairing and coherence order parameters in a three-component ultracold Fermi gas,
Phys. Rev. A **82** (3 2010) 033620,
URL: https://link.aps.org/doi/10.1103/PhysRevA.82.033620 (cit. on p. 162).

[258] Y. Shin et al.,
Observation of Phase Separation in a Strongly Interacting Imbalanced Fermi Gas,
Physical Review Letters **97** (2006), ISSN: 1079-7114,
URL: http://dx.doi.org/10.1103/PhysRevLett.97.030401 (cit. on p. 162).

[259] J. H. McLeod, *The Axicon: A New Type of Optical Element*,
J. Opt. Soc. Am. **44** (1954) 592–597,
URL: http://www.osapublishing.org/abstract.cfm?URI=josa-44-8-592
(cit. on p. 162).

[260] I Manek, Y. Ovchinnikov and R Grimm,
Generation of a hollow laser beam for atom trapping using an axicon,
Optics Communications **147** (1998) 67 –70, ISSN: 0030-4018,
URL: http://www.sciencedirect.com/science/article/pii/S0030401897006457
(cit. on p. 162).

APPENDIX A

Appendix

A.1 Neural network architectures

A.1.1 Phase diagrams

Layer (type)	Output Shape	Parameters
Input	(None, 150, 170, 1)	0
2D Convolutional	(None, 150, 170, 30)	300
Max Pooling	(None, 75, 85, 30)	0
2D Convolutional	(None, 75, 85, 40)	30040
Max Pooling	(None, 37, 42, 40)	0
2D Convolutional	(None, 37, 42, 50)	50050
Max Pooling	(None, 18, 21, 50)	0
Flatten	(None, 18900)	0
Dense	(None, 600)	11340600
Dropout	(None, 600)	0
Dense	(None, 300)	180300
Dropout	(None, 300)	0
Dense	(None, 1)	301

Table A.1: Neural network architecture used for the phase diagrams. This network architecture was used for all phase diagrams in this thesis (Figures 5.17, 5.19, 5.20, A.1, A.2), but trained separately with different training data. The total number of optimised parameters is 11 601 591. The None in the output shapes describes the batch dimension. We used a batch size of 30 and trained for 15 epochs with Adam optimiser [113] and a learning rate of 1.5×10^{-4} with mean squared error as loss function. We realised this network with the TensorFlow library [174].

A.1.2 Autoencoder

Layer (type)	Output Shape	Parameters
Input	(None, 150, 170, 1)	0
2D Convolutional	(None, 150, 170, 32)	320
Max Pooling	(None, 75, 85, 32)	0
2D Convolutional	(None, 75, 85, 64)	0
Max Pooling	(None, 15, 17, 64)	0
Flatten	(None, 16320)	0
Dense	(None, 255)	4161855
Dense	(None, 500)	128000
Dense (Bottleneck)	(None, 2)	1002
Dense	(None, 500)	1500
Dense	(None, 255)	127755
Reshape	(None, 15, 17, 1)	0
2D Transpose Convolutional	(None, 75, 85, 64)	320
Batch Normalization	(None, 75, 85, 64)	256
2D Transpose Convolutional	(None, 150, 170, 32)	8224
Batch Normalization	(None, 150, 170, 32)	128
2D Convolutional	(150, 170, 1)	289

Table A.2: Autoencoder architecture. This network architecture was used in section 6 to create the low-dimensional representation of the phase diagram data shown in figure 6.1. The total number of parameters is 4 437 905 and the None in the output shape describes the batch dimension. We used a batch size of 40 and trained for 15 epochs with Adam optimiser [113] and a learning rate of 4×10^{-4} with mean squared error as loss function. The network was realised with the TensorFlow library [174].

A.2 Phase diagrams

Figures A.1 and A.2 show phase diagrams similar to the diagrams in figures 5.20 and 5.19 but with additional training data closer to the unitarity regime. This does not change the picture drawn in section 5.4.1, as the additional fields are still in the regime where the detection does not change as compared to the full phase diagram.

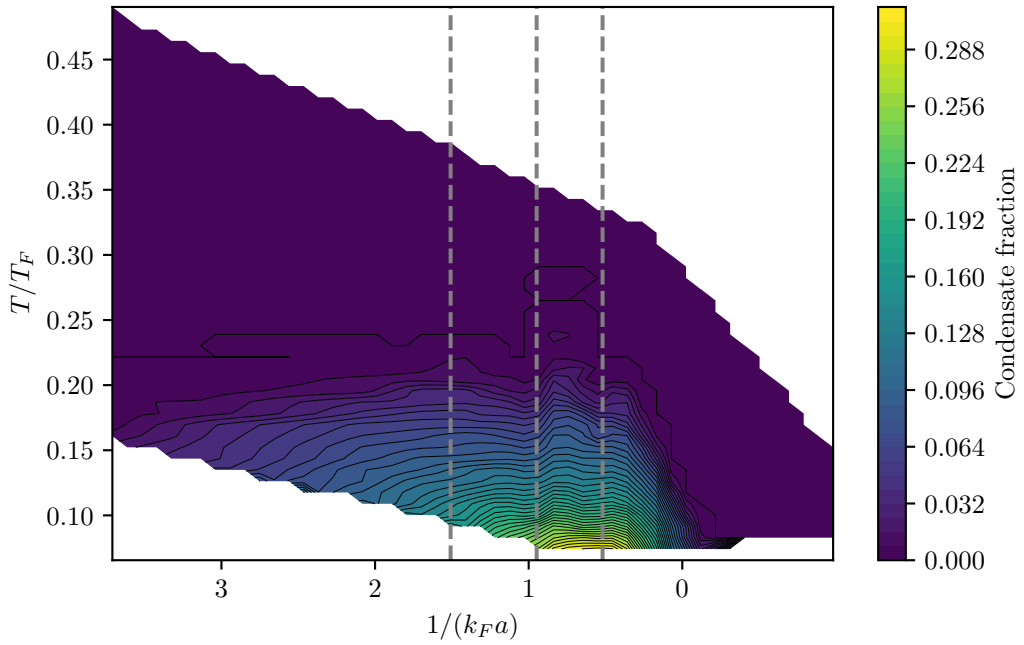

Figure A.1: Phase diagram with more training data on the BEC side. Compared to figure 5.20 we include training data closer to the unitarity regime. This increases the training data, but does not change the qualitative picture of the phase diagram.

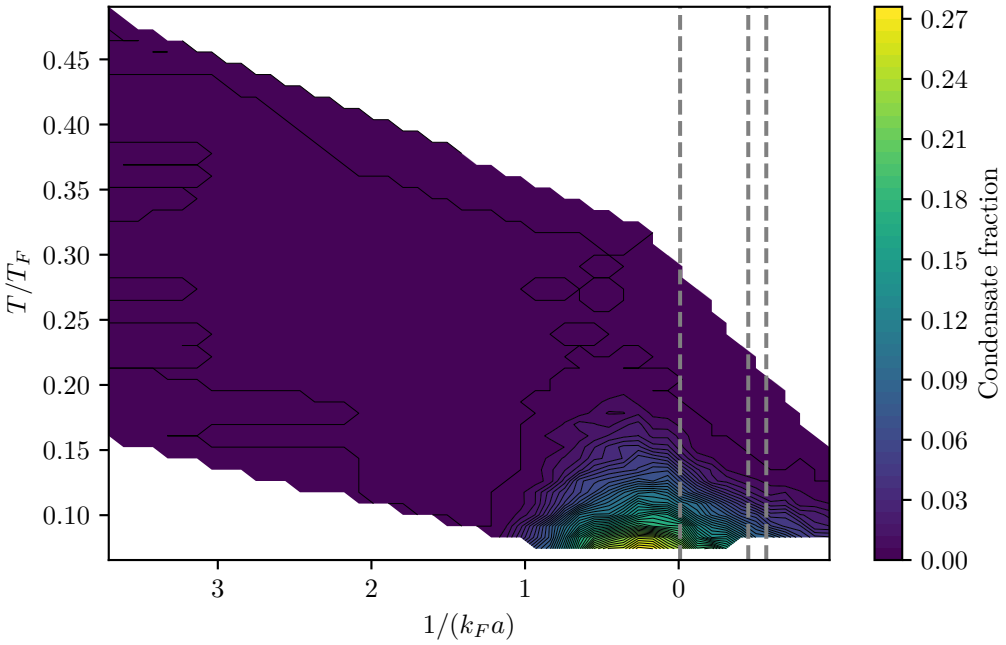

Figure A.2: Phase diagram with more training data on the BCS side. Compared to figure 5.19 we include training data closer to the unitarity regime. This increases the training data, but does not change the qualitative picture of the phase diagram.

A.3 Pictures of the quench coil setup

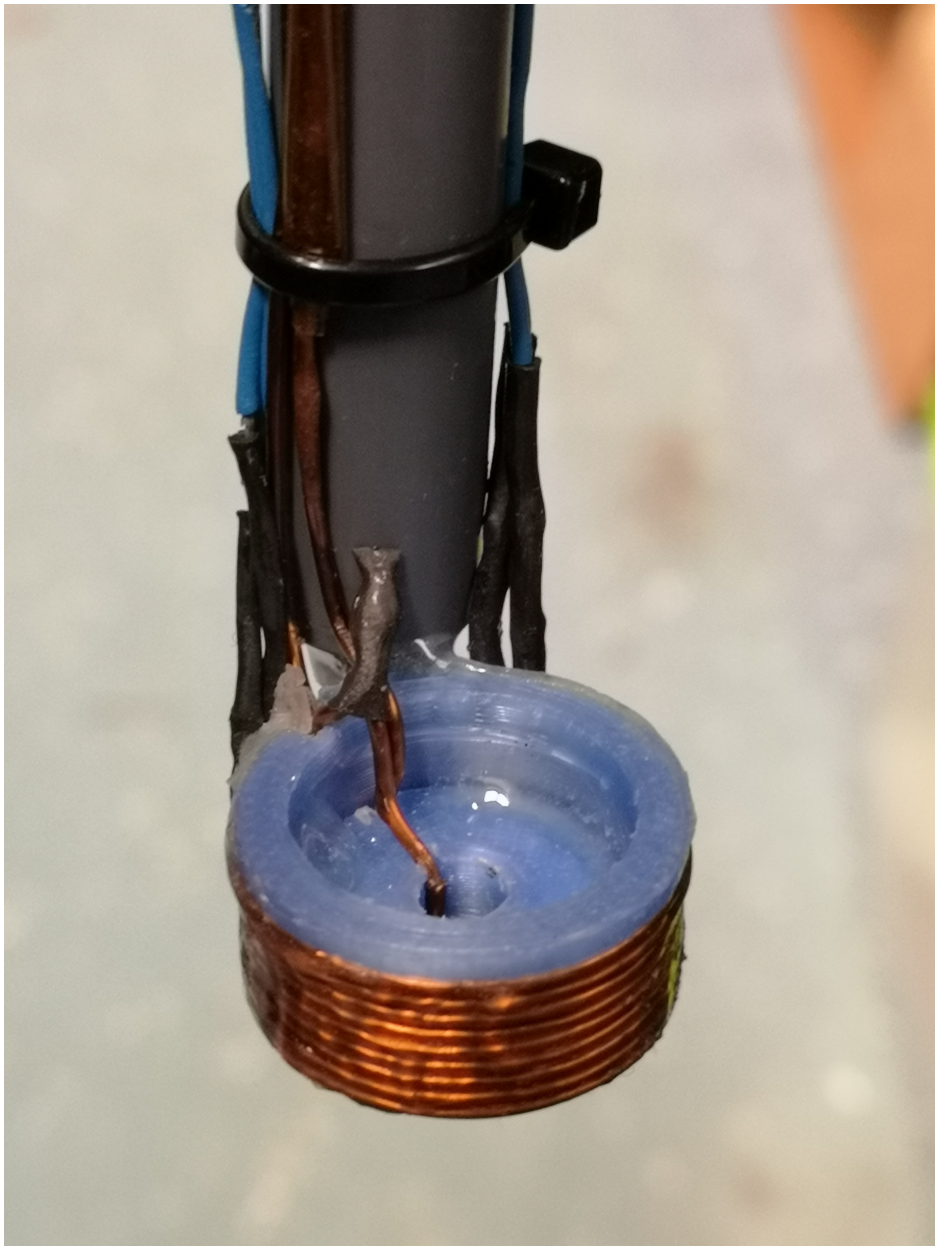

Figure A.3: Picture of the assembled quench coil setup.

List of Figures

List of Tables