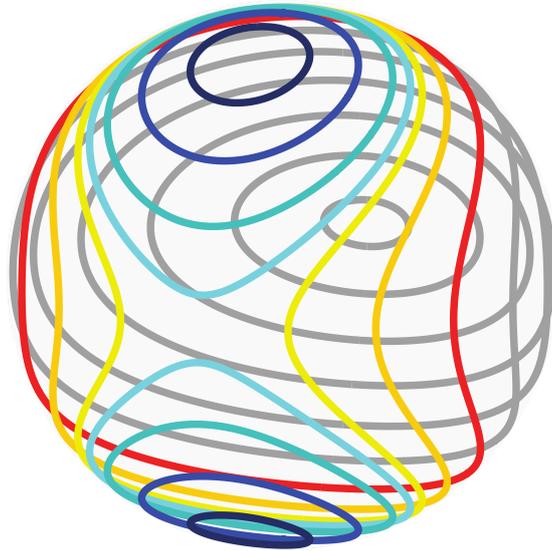


Fisher Information and entanglement of non-Gaussian spin states



Helmut Strobil
2015

Dissertation
submitted to the
Combined Faculties of the Natural Sciences and Mathematics
of the Ruperto-Carola-University of Heidelberg. Germany
for the degree of
Doctor of Natural Sciences

Put forward by
Helmut Strobel
born in: Ulm, Germany
Oral examination: 4 February 2016

Fisher Information and entanglement of non-Gaussian spin states

Referees: Prof. Dr. Markus K. Oberthaler
PD Dr. Carsten Klempt

Abstract

In this thesis, we study a novel method to extract the *Fisher information* of quantum states from direct measurements without the need for state reconstruction. Our method characterizes the distinguishability of experimental probability distributions of the collective spin. The Fisher information is obtained via the observed rate of change of their statistical distance as a function of an experimental control parameter, which constitutes a phase transformation of the quantum state.

The employed experimental system is a binary Bose-Einstein condensate of several hundred atoms. We use a combination of coherent collisional interaction and linear Rabi coupling of the two atomic states to generate collective non-classical spin states via quantum dynamics close to an unstable fixed point of the corresponding classical system. The fast redistribution of quantum uncertainty results in Gaussian spin-squeezed states for short evolution times which turn into non-Gaussian states on an experimentally feasible time scale.

For the generated non-Gaussian states we observe a Fisher information larger than the number of atoms in the detected ensemble, which is a signature of particle entanglement, in a regime where no spin-squeezing is present. We confirm the implied resource for quantum-enhanced measurements with the implementation of a model-free Bayesian protocol which obtains a sensitivity beyond the standard quantum limit in accordance with the extracted Fisher information.

Zusammenfassung

Diese Arbeit beschäftigt sich mit einer neuartigen Methode, die sogenannte *Fisher-Information* von Quantenzuständen aus direkten Messungen zu extrahieren, ohne dabei eine Zustands-Rekonstruktion zu benötigen. Unsere Methode charakterisiert die Unterscheidbarkeit der relativen Häufigkeitsverteilungen des kollektiven Spins. Die Fisher-Information wird dabei über die beobachtete Änderungsrate ihres statistischen Abstands als Funktion eines experimentellen Kontrollparameters erhalten, der eine Phasentransformation des Quantenzustands darstellt.

Unser experimentelles System ist ein Bose-Einstein-Kondensat aus mehreren hundert Atomen. Mit Hilfe einer Kombination von kohärenter Stoßwechselwirkung und linearer Rabi-Kopplung zweier atomarer Zustände erzeugen wir nichtklassische kollektive Spinzustände durch Quantendynamik in der Nähe eines instabilen Fixpunkts des korrespondierenden klassischen Systems. Die schnelle Umverteilung der Quantenunschärfe ergibt dabei auf kurzen Zeitskalen gaußförmige gequetschte Spinzustände, die sich auf experimentell zugänglichen Zeitskalen in nicht-gaußsche Zustände verformen.

In diesem Regime, in dem keine Spin-Quetschung mehr vorhanden ist, beobachten wir einen Wert der Fisher-Information, der die Atomzahl des detektierten Ensembles übersteigt, was eine Signatur von Teilchen-Verschränkung darstellt. Wir bestätigen die damit verbundene Möglichkeit quanten-verstärkter Messungen mit der experimentellen Umsetzung eines modell-freien bayesschen Protokolls, das im Einklang mit der extrahierten Fisher-Information eine Sensitivität jenseits des Standard-Quanten-Limits zeigt.

Contents

1	Introduction	1
2	Theoretical Concepts	5
2.1	Two-mode Bose Einstein condensate	5
2.1.1	Two-level atom	5
2.1.2	Collective spin picture	7
2.1.3	Collective rotations	9
2.1.4	Coherent spin states	9
2.1.5	Quasiprobabilities	11
2.1.6	Interaction and the two-mode BEC	13
2.1.7	Collective spin Hamiltonian	15
2.1.8	Mean field limit and classical trajectories	16
2.1.9	Mapping to a rigid pendulum	19
2.1.10	Quantum evolution from the classically unstable fixed point	22
2.2	Fisher Information and the Cramér-Rao bound	24
2.3	Maximum likelihood and Bayesian estimation	26
2.4	Fisher information extraction	29
2.4.1	The concept of statistical distance	29
2.4.2	Hellinger distance	31
2.4.3	Extraction procedure	32
2.4.4	Hellinger distance of experimental probability distributions	33
2.4.5	Jackknife resampling and Monte-Carlo simulation	36
2.5	Quantum Fisher Information	40
2.6	Time evolution of Fisher information	42
2.7	Fisher Information and entanglement	45
2.7.1	Entanglement and symmetric states	46
3	Experimental system	49
3.1	The route to BEC	49
3.2	Pseudospin system	50
3.2.1	Interaction	51
3.2.2	Linear Coupling	51
3.2.3	Detuning	52
3.2.4	Detection	52
3.3	Optical traps	53
3.3.1	Lattice	53
3.3.2	Transversal confinement	54

3.4	Resonance Shifts	55
3.4.1	Magnetic field	55
3.4.2	AC electric fields of the dipole traps	57
3.4.3	Interaction shifts	58
3.5	Characterization of the interaction parameters	58
3.6	Gradients of the parameters	63
3.6.1	Trapping	63
3.6.2	Magnetic offset field	64
3.6.3	Coupling fields	64
3.7	Atom loss	65
3.7.1	Two-body spin relaxation	66
3.7.2	Feshbach loss	67
4	State preparation and readout	71
4.1	Spin manipulation sequence	71
4.1.1	Towards the right parameters	72
4.1.2	Keeping track of the detuning	74
4.1.3	Spin-echo pulse	74
4.2	Readout and data analysis	76
5	State tomography	77
5.1	Quantum state reconstruction	77
5.2	Squeezing tomography and two-mode model	81
6	Fisher information extraction	86
6.1	Hellinger distances and extraction method	86
6.2	Gaussian spin-squeezed states	88
6.3	Transition to non-Gaussian states	90
7	Phase estimation	93
7.1	Gaussian performance	93
7.2	Bayesian phase sensitivity	95
8	Outlook	99
A	Supplementary	100
A.1	Quantum Fisher Information	100
A.2	Coupled basis and singlet states	103
A.3	SU(2) interferometer and rotations on the Bloch-sphere	104
A.4	Hyperfine structure in a magnetic field	105
B	Additional characterizations and technicalities	106
B.1	Characterization of the optical dipole traps	106
B.2	Magnetic field stabilization	110
B.3	Microwave and radio frequency system	115
B.4	Spectral sensitivity of pulse sequences	118
	Bibliography	125

List of Figures

1.1	Phase estimation in the linear two-mode interferometer	1
2.1	Spin 1/2 Bloch sphere picture	6
2.2	Phase space picture of the weakly coupled interacting two-mode BEC	17
2.3	Classical phase space with finite detuning	18
2.4	Rigid pendulum mapping of the Bose Josephson Junction mean-field limit	20
2.5	Phase space picture of the ideal quantum evolution.	23
2.6	Geometrical picture of probability distributions and statistical distance	31
2.7	Multinomial sampling distribution	33
2.8	Monte-Carlo simulation of the Hellinger distance method I	36
2.9	Monte-Carlo simulation of the Hellinger distance method II	37
2.10	Monte-Carlo simulation of the Hellinger distance method III	38
2.11	Quantum parameter estimation setting	40
2.12	Time evolution of Fisher information and spin squeezing	44
2.13	Entanglement criteria – Fisher information vs. depth of entanglement	45
3.1	Geometry of the experiment and physical system	50
3.2	Hyperfine structure and magnetic field shifts	56
3.3	Scheme of plasma- and π -oscillations in the Rabi regime	59
3.4	Characterization of the nonlinearity	60
3.5	Interaction shift of the two-photon resonance	61
3.6	Influence of preparation error and detuning	62
3.7	Two-body spin relaxation loss	66
3.8	Losses due to the proximity of the Feshbach resonance	67
3.9	Exponential lifetime for different initial conditions	69
4.1	Experimental pulse sequence	72
4.2	Action of the spin echo pulse	75
5.1	Experimental probability distributions of the tomographic characterization	78
5.2	Gallery of reconstructed quantum states for 25 ms of evolution time	80
5.3	Gallery of squeezed states for 15 ms of evolution time	81
5.4	Squeezing parameters of the quantum state tomography	82
5.5	Comparison of the two-mode model with experimental distributions	83
6.1	Extraction of the Fisher information with the Hellinger distance method	87
6.2	Geometric picture for the evaluation of the mean spin length	89
6.3	Fisher information of spin-squeezed states	90

6.4	Fisher information at the transition to non-Gaussian states	92
7.1	Gaussian sensitivity of a non-Gaussian state	94
7.2	Bayesian phase estimation scheme	96
7.3	Bayesian phase estimation sensitivity	97
A.1	Linear SU(2) interferometer and rotations on the Bloch sphere	104
B.1	Characterization of the waveguide trap	107
B.2	Transversal trapping frequencies and trap depths of the waveguide	108
B.3	Trap frequency of the optical lattice potential	109
B.4	Principle of the magnetic offset field stabilization	110
B.5	Noise of the magnetic field sensor	111
B.6	Temperature drift of the magnetic field stabilisation	113
B.7	Magnetic field stability obtained with the active control	114
B.8	Block circuit of the microwave and radio frequency system	116
B.9	Universal spectral transfer functions of pulse sequences	118
B.10	Collected specifications and characterizations of the microwave sources	122
B.11	Collection of Allan deviations and examples of inferred angle fluctuations	123

Publication list

- **Rabi flopping induces spatial demixing dynamics**
E. Nicklas, H. Strobel, T. Zibold, C. Gross, B. A. Malomed,
P. G. Kevrekidis and M. K. Oberthaler
Phys. Rev. Lett. **107**, 193001 (2011)
- **Atomic homodyne detection of continuous variable entangled twin-atom states**
C. Gross, H. Strobel, E. Nicklas, T. Zibold, N. Bar-Gill, G. Kurizki and M. K. Oberthaler
Nature **480**, 219–223 (2011)
- **Optimized absorption imaging of mesoscopic atomic clouds**
W. Muessel, H. Strobel, M. Joos, E. Nicklas, I. Stroescu, J. Tomkovič,
D. B. Hume and M. K. Oberthaler
Appl. Phys. B **113**, 69–73 (2013)
- **Accurate Atom Counting in Mesoscopic Ensembles**
D. B. Hume, I. Stroescu, M. Joos, W. Muessel, H. Strobel, and M. K. Oberthaler
Phys. Rev. Lett. **111**, 253001 (2013)
- **Fisher information and entanglement of non-Gaussian spin states**
H. Strobel, W. Muessel, D. Linnemann, T. Zibold, D. B. Hume,
L. Pezzè, A. Smerzi, and M. K. Oberthaler
Science **345**, 424–427 (2014)
- **Scalable Spin Squeezing for Quantum-Enhanced Magnetometry
with Bose-Einstein Condensates**
W. Muessel, H. Strobel, D. Linnemann, D. B. Hume, and M. K. Oberthaler
Phys. Rev. Lett. **113**, 103004 (2014)
- **Twist-and-turn spin squeezing in Bose-Einstein condensates**
W. Muessel, H. Strobel, D. Linnemann, T. Zibold, B. Juliá-Díaz, and M. K. Oberthaler
Phys. Rev. A **92**, 023603 (2015)
- **Observation of Scaling in the Dynamics of a Strongly Quenched Quantum Gas**
E. Nicklas, M. Karl, M. Höfer, A. Johnson, W. Muessel, H. Strobel, J. Tomkovič,
T. Gasenzer, and M. K. Oberthaler
Phys. Rev. Lett. **115**, 245301 (2015)
- **Nonlinear dressed states at the miscibility-immiscibility threshold**
E. Nicklas, W. Muessel, H. Strobel, P. G. Kevrekidis, and M. K. Oberthaler
Phys. Rev. A **92**, 053614 (2015)
- **Observing the emergence of chaos in a many-particle quantum system**
J. Tomkovič, W. Muessel, H. Strobel, S. Löck, P. Schlagheck,
R. Ketzmerick, and M. K. Oberthaler
arXiv:1509.01809 (2015) (submitted for publication)

1 Introduction

Accurate measurements of physical quantities such as time, (gravitational) acceleration and rotation or electromagnetic fields are important for both fundamental science as well as technology. Atom interferometers [1–4] are especially well-suited for these purposes, since they provide high sensitivity due to the possibility of putting atoms into motional or internal quantum superposition states, while the employed sensors (the atoms) are perfectly reproducible in their physical properties and can be isolated from the environment to a large degree. The availability of laser cooling and trapping [5] and the achievement of Bose-Einstein condensation [6] has advanced this field considerably, since the enhanced coherence of the cold ensembles enables high precision spectroscopy and long interrogation times. However, the finite number of particles that can be probed in a single experimental realization imposes fundamental limits on the precision of such devices.

The prototypical metrological device is the linear interferometer with two arms as shown in Fig. 1.1 in its optical implementation in Mach-Zehnder configuration. In an atom interferometer, the two optical beamsplitters are replaced by matter-wave splitters. An important analogue of the Mach-Zehnder interferometer is the Ramsey method of separated oscillatory fields [7], which is an invaluable tool for high precision spectroscopy.

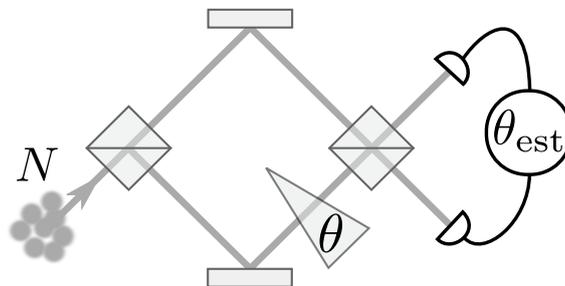


Figure 1.1 The linear two-mode interferometer (illustrated here in optical Mach-Zehnder configuration) is a model system of (quantum) metrology. The quantity of interest is the relative phase shift θ between the two arms of the interferometer. The estimation of this phase shift from the measured output statistics is fundamentally limited by the granularity of the input state with N particles.

Single particles entering one input port of the interferometer will be put into a quantum superposition between path 1 ($|\uparrow\rangle$) and 2 ($|\downarrow\rangle$). After a possible phase shift θ between the two paths inside the interferometer, which is the quantity of interest, the two paths are recombined and thus θ is mapped onto a probability p for the particle to exit one of the two output ports. Upon measurement, the most information that can be obtained from one particle is the output port it has taken (in which of the detectors it is detected). This is one classical bit of information. If the phase shift is fixed, this procedure can be repeated N times and the probability and thus the phase shift estimated from the relative frequency of the two

detectors. The statistics of this process is binomial like N coin tosses with the probability p for heads and $1 - p$ for tails and thus the precision on p is limited by the binomial uncertainty $\Delta p = \sqrt{p(1-p)/N}$, which leads for $p = \cos^2(\theta/2)$ via the Gaussian error propagation $\Delta p = |\cos(\theta/2) \sin(\theta/2)| \Delta \theta = \sqrt{p(1-p)} \Delta \theta$ to the famous shot-noise or standard quantum limit (SQL)

$$\Delta \theta_{\text{SQL}} = \frac{1}{\sqrt{N}} \quad (1.1)$$

for the estimation of the phase θ . This situation does not change if N independent particles are inside the interferometer at the same time.

Better sensitivity in specific ranges of θ can be achieved by introducing quantum correlations (entanglement) between the N particles. In this case, the entangled sensors experience the phase shift in parallel [8] and the fundamental limit of this approach is given by the Heisenberg limit $\Delta \theta_{\text{HL}} = 1/N$, which can be only attained with maximally entangled states. Generally, the obtainable precision is given by the Cramér-Rao bound

$$\Delta \theta_{\text{est}} \geq \frac{1}{\sqrt{F}} \quad (1.2)$$

where F is the Fisher information, which characterizes the sensitivity of the state on the parameter θ . If $F > N$ for a specific quantum state and phase transformation, it contains entanglement [9] and can be used to surpass the standard quantum limit.

The important class of Gaussian (spin-)squeezed states [10, 11] achieves this by reducing the output fluctuations, such that $\Delta \theta = \xi_R \Delta \theta_{\text{SQL}}$ with the metrological spin squeezing parameter $\xi_R < 1$ [10]. They have been produced to date in a variety of systems (see our publications [12, 13] and references therein) and are already employed in precision experiments, for example in advanced (optical) gravitational-wave detectors [14, 15], for which the use of squeezed states of light was first proposed [16].

In contrast, non-Gaussian states can allow for sensitive measurements beyond the standard quantum limit even though they feature increased fluctuations of the observable. They are an interesting pathway for metrology, since they can in principle provide sensitivity beyond what can be achieved with squeezed states in the same system [9]. Without disentangling operations before detection [17, 18], which are a challenge to implement, probabilistic methods like Bayesian analysis or the method of maximum likelihood have to be employed for phase estimation [19, 20] to benefit from the enhanced sensitivity.

In this thesis, we will discuss experiments using a Bose-Einstein condensate (BEC) with two internal spin states, in which coherent collisional interaction builds up entanglement between the particles. We enhance this process by the addition of linear coupling, which leads to an interesting situation with an inverted mathematical pendulum as its classical counterpart. This system allows to study the transition from Gaussian spin squeezed to non-Gaussian states on an experimentally feasible time scale. Since full quantum state reconstruction is intractable for several hundred particles, we have developed a statistical method to extract the Fisher information from measured experimental probability distributions, which serves to verify entanglement in the non-Gaussian regime, where no spin squeezing is present. To demonstrate that the resource can be employed in a phase estimation scenario, we characterize the sensitivity of a model-independent Bayesian method. These results are summarized in our publication [21].

This thesis is structured in the following way:

- In Chapter 2 we summarize the theoretical concepts. We start with the description of a two-mode BEC and its pseudospin representation and discuss the classical limit, which provides an intuitive picture of the quantum dynamics with the help of quasiprobability distributions. We show that the classical limit for zero external detuning can be mapped to a rigid pendulum. We discuss the concept of Fisher information and its role in statistical estimation and introduce our extraction method based on statistical distances, which we examine analytically and with the help of Monte-Carlo simulations. After a short discussion of the Quantum Fisher information, we study the ideal time evolution of spin squeezing and Fisher information with realistic parameters in the framework of the two-mode model and show the influence of Gaussian noise. We further give a summary of the relation between Fisher information and entanglement and a short discussion about the role of entanglement in symmetric states.
- In Chapter 3 we detail the experimental system and the implementation of the spin Hamiltonian. We characterize the optical traps and show the influence of external fields and the collisional interaction on the employed two-photon resonance. We show how the nonlinearity, which is enhanced with the help of a Feshbach resonance, can be directly characterized. We further discuss gradients along the employed BEC array and how we deal with them to obtain consistent conditions. We further show an analysis of the atom loss during the spin evolution.
- In Chapter 4 we describe the experimental sequence, the steps towards the right parameters and how we ensure their stability during long-term measurements. We show the nontrivial action of the employed spin-echo pulse and give details about readout and data analysis.
- In Chapter 5 we show the measurement of experimental probability distributions and the results of quantum state tomography. We discuss the results of maximum-likelihood reconstruction of the density matrices and the limitations of this approach. The quasiprobability distributions of the reconstructed states give an intuitive picture of the systematic behavior with the total atom number. We compare the results of a two-mode model with a quantitative variance analysis and non-Gaussian experimental probability distributions.
- In Chapter 6 we present the experimental Fisher information extraction using our statistical distance method and give an intuitive picture of the procedure. We test the method in the Gaussian regime with spin squeezed states and show corresponding results in the non-Gaussian regime.
- In Chapter 7 we discuss the limits of the phase estimation sensitivity for the non-Gaussian states in a Gaussian protocol and characterize the sensitivity of a model-free Bayesian phase estimation procedure, which can be used to harness the non-Gaussian resource in accordance with the extracted Fisher information.
- After the Outlook in Chapter 8 we give in Appendix A supplementary theoretical details and the parameters of the hyperfine structure as reference. In Appendix B we show

further characterizations of the employed dipole traps and provide technical details about the magnetic field stabilization and the microwave and radio frequency system. Furthermore we derive the influence of spectral phase noise on the performance of linear pulse sequences.

2 Theoretical Concepts

In this chapter we provide insights into the theoretical concepts relevant for this thesis. Although the generation of entangled states with Bose-Einstein condensates is a top-down approach without access to the individual particles, we begin the discussion with a microscopic view, which provides deeper insights into the mechanisms and their interpretation. We further discuss the concept of Fisher information which provides to date the most general understanding of advanced metrological applications and the role of entanglement therein. We will show its connection to the concept of statistical distance, which provides the basis of our method for experimental characterization of non-Gaussian quantum states.

2.1 Two-mode Bose Einstein condensate

Our systems of choice are Bose-Einstein condensates of ^{87}Rb with atom numbers in the range of 200 to 600 in tight optical traps, such that spatial excitation degrees of freedom are effectively frozen out. After their production in one energy eigenstate of the electronic ground state hyperfine manifold, we address the transition to a second substate with microwave and radio frequency magnetic fields and in this way realize an ensemble of two-level systems. The high density of the BEC induces significant collisional interactions, which in the case of spin dependency allows to redistribute the collective quantum uncertainty, which leads to entanglement between the atoms. The motional BEC state with negligible thermal fraction causes these collisions to be coherent and reversible for long periods of time. The intuitive microscopic picture of this is, that for the majority of collisional events of two atoms, initial and final states are the identical (two) spatial modes of the BEC.

An important consequence of the BEC state is, that the individual constituent atoms become indistinguishable even in principle. Since they share the same motional wavefunction, the description of the internal degrees of freedom has to be symmetric on particle exchange to yield a valid bosonic total wavefunction. This somewhat obscures the notion of entanglement, because the division into subsystems is ambiguous. We will give a practical perspective on this problem.

2.1.1 Two-level atom

The two-state system is the most generic setting in quantum mechanics and is considered most important as quantum bit (qubit) for technological applications [22]. In our case, this will be realized with two levels $|a\rangle$ and $|b\rangle$ of the ground state hyperfine manifold. It is convenient

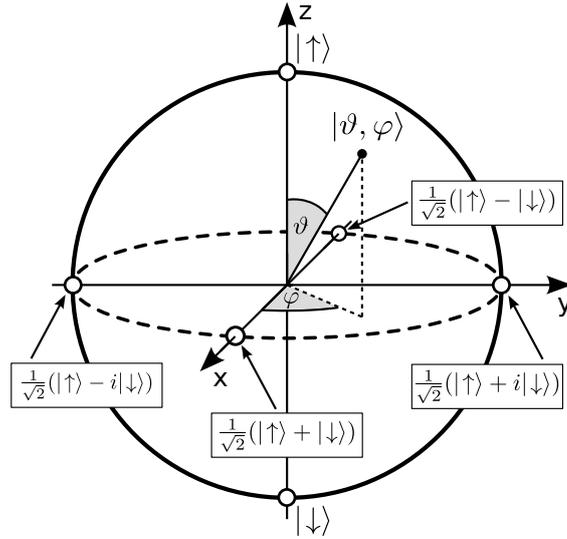


Figure 2.1 Bloch sphere picture of a single atom with two internal states $|\uparrow\rangle$ and $|\downarrow\rangle$. All possible pure state superpositions can be parametrized by two angles (ϑ, φ) and span the surface of the unit sphere

to map it to a spin-1/2 with the two states $|a\rangle = |\downarrow\rangle$ (spin down) and $|b\rangle = |\uparrow\rangle$ (spin up). All possible superpositions of these two states can be parametrized by

$$|\vartheta, \varphi\rangle = \cos \frac{\vartheta}{2} |\uparrow\rangle + \sin \frac{\vartheta}{2} e^{i\varphi} |\downarrow\rangle \quad (2.1)$$

with the two angles ϑ and φ ranging from 0 to π and 0 to 2π , respectively. With that, the Hilbert space has a one-to-one mapping to the surface of a sphere with radius 1/2, the so-called Bloch sphere. It provides an intuitive picture of unitary operations, which all translate to rotations of the Bloch vector [23]. An operator basis in this space is given by (together with the identity matrix) the Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (2.2)$$

which are the generators of rotations around the respective axes and act on the representation $|\downarrow\rangle \rightarrow (0 \ 1)^T$, $|\uparrow\rangle \rightarrow (1 \ 0)^T$ which is the eigenbasis of σ_z . They obey the SU(2) commutation relation $[\sigma_x, \sigma_y] = 2i\sigma_z$ and their cyclic permutations. Another pair of useful operators are the raising and lowering operators $\sigma_{\pm} = \frac{1}{2}(\sigma_x \pm i\sigma_y)$, acting on the basis states in the following way:

$$\sigma_+ |\uparrow\rangle = 0 \quad \sigma_+ |\downarrow\rangle = |\uparrow\rangle \quad \sigma_- |\uparrow\rangle = |\downarrow\rangle \quad \sigma_- |\downarrow\rangle = 0. \quad (2.3)$$

They obey the commutation relations $[\sigma_+, \sigma_-] = 2\sigma_z$ and $[\sigma_z, \sigma_{\pm}] = \pm\sigma_{\pm}$. In turn, these can be used to define the operators

$$\frac{\sigma_x}{2} = \frac{1}{2}(\sigma_+ + \sigma_-) \quad \text{and} \quad \frac{\sigma_y}{2} = \frac{1}{2i}(\sigma_+ - \sigma_-). \quad (2.4)$$

Together with the commutation relations, they are especially useful to construct the basis for coupled spins.

2.1.2 Collective spin picture

For our N atoms, or equivalently N elementary spins $\mathbf{s} = (\sigma_x/2, \sigma_y/2, \sigma_z/2)$, the Hilbert space is given by the direct product $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_N$ of the two-dimensional \mathcal{H}_k of the individual spin-1/2 systems which has the exponentially large dimension 2^N .

When the individual qubits are not resolved, this complexity can be reduced by considering the collective spin operators

$$\hat{\mathcal{J}}_\alpha = \sum_{k=1}^N s_\alpha^{(k)} = \frac{1}{2} \sum_{k=1}^N \sigma_\alpha^{(k)}, \quad \alpha = x, y, z. \quad (2.5)$$

This has the advantage, that the z -component $\hat{\mathcal{J}}_z = (n_\uparrow - n_\downarrow)/2$ corresponds to our measurement observable, when counting the number of spin-ups and spin-downs of the ensemble, implementing a projective measurement. The length

$$\sqrt{\langle \hat{\mathcal{J}}^2 \rangle} = \sqrt{\langle \hat{\mathcal{J}}_x^2 + \hat{\mathcal{J}}_y^2 + \hat{\mathcal{J}}_z^2 \rangle} = \sqrt{J(J+1)} \quad (2.6)$$

of this spin can take all values from 0 to $\sqrt{N/2(N/2+1)}$ if N is even and $\sqrt{3}/2$ up to the same maximal value if N is odd. However, the value of J does not uniquely identify a subspace of \mathcal{H} . This is most obvious, when summing the dimensions of J -subspaces (here for N even):

$$\sum_{J=0}^{N/2} (2J+1) = \frac{N}{2} \left(\frac{N}{2} + 2 \right) + 1 \ll 2^N \quad (2.7)$$

The reason is, that except for the case $J = N/2$ there exists a multitude of orthogonal subspaces with the same J [24]. A special and important case is realized, if the individual qubits are indistinguishable even in principle. This is the case, when the spatial wave functions of the physical qubits perfectly overlap as in our case of the two-mode BEC, or all interactions are sufficiently fast and nonlocal as is the case for ions sharing collective trap-oscillations [25, 26], or atoms, which interact collectively via the light field inside an optical cavity [27, 28]. A symmetric initial state can for example be realized by preparing all qubits in $|\downarrow\rangle$ and applying a preparation pulse sufficiently fast and homogeneous over the ensemble to prepare each spin in the same superposition. The state space for indistinguishable qubits is spanned by the eigenstates of $\hat{\mathcal{J}}_z$ (any axis is equally suitable) with the largest spin length, which are the fully symmetric, permutationally invariant Dicke states $|J, m\rangle = |N/2, m\rangle$. We will also use the notation $|k, N-k\rangle$ with $k = m + N/2$ to simplify expressions. The spin length has to be maximal for indistinguishable qubits, as can be seen with the explicit form of the Dicke state in first quantization

$$|k, N-k\rangle = \frac{1}{\sqrt{\binom{N}{k}}} \sum_{j,l} \hat{P}_{jl} |\underbrace{\uparrow\uparrow \dots \uparrow}_k \underbrace{\downarrow\downarrow \dots \downarrow}_{N-k}\rangle, \quad (2.8)$$

where \hat{P}_{jl} exchanges qubit j with qubit l and the sum runs over all distinguishable permutations. Since the permutation operator commutes with $\hat{\mathcal{J}}_z^2$ one can evaluate

$$\hat{\mathcal{J}}_z^2 |k, N-k\rangle = \frac{1}{\sqrt{\binom{N}{k}}} \sum_{j,l} \hat{P}_{jl} \hat{\mathcal{J}}_z^2 |\underbrace{\uparrow\uparrow \dots \uparrow}_k \underbrace{\downarrow\downarrow \dots \downarrow}_{N-k}\rangle = \left(k - \frac{N}{2}\right)^2 |k, N-k\rangle \quad (2.9)$$

Similarly, one obtains $\langle \hat{\mathcal{J}}_x^2 \rangle = \langle \hat{\mathcal{J}}_y^2 \rangle = N/2 + k(N - k)$, which leads to $\langle \hat{\mathcal{J}}^2 \rangle = N/2(N/2 + 1)$ and thus to the maximal $J = N/2$. For the example $N = 3$ the Dicke states are

$$\begin{aligned} |3, 0\rangle &= |\uparrow\uparrow\uparrow\rangle \\ |2, 1\rangle &= \frac{1}{\sqrt{3}} (|\downarrow\uparrow\uparrow\rangle + |\uparrow\downarrow\uparrow\rangle + |\uparrow\uparrow\downarrow\rangle) \\ |1, 2\rangle &= \frac{1}{\sqrt{3}} (|\uparrow\downarrow\downarrow\rangle + |\downarrow\uparrow\downarrow\rangle + |\downarrow\downarrow\uparrow\rangle) \\ |0, 3\rangle &= |\downarrow\downarrow\downarrow\rangle \end{aligned} \quad (2.10)$$

As reference, in appendix A.2, we list also the non-symmetric basis states for $N = 3$ and the singlet states (zero collective spin length) of $N = 4$. In analogy to the single qubit, the collective raising and lowering operators

$$\hat{\mathcal{J}}_{\pm} = \frac{1}{2} \sum_{k=1}^N \hat{\sigma}_{\pm}^{(k)} \quad (2.11)$$

can be defined, which inherit the commutation relations $[\hat{\mathcal{J}}_+, \hat{\mathcal{J}}_-] = 2\hat{\mathcal{J}}_z$ and $[\hat{\mathcal{J}}_z, \hat{\mathcal{J}}_{\pm}] = \pm\hat{\mathcal{J}}_{\pm}$. They can be used to construct the symmetric basis beginning with $|0, N\rangle$. One finds

$$\begin{aligned} \hat{\mathcal{J}}_+ |k, N - k\rangle &= \sqrt{(k + 1)(N - k)} |k + 1, N - k - 1\rangle \\ \hat{\mathcal{J}}_- |k, N - k\rangle &= \sqrt{k(N - k + 1)} |k - 1, N - k + 1\rangle. \end{aligned}$$

In this fully symmetric subspace, the collective spin operators can also be represented in second quantization using the generic harmonic oscillator operators a^\dagger and b^\dagger with the commutation relations $[a, a^\dagger] = 1$, $[b, b^\dagger] = 1$ and $[a^\dagger, b^\dagger] = [a, b] = 0$ leading to

$$\hat{J}_+ = b^\dagger a \quad (2.12)$$

$$\hat{J}_- = a^\dagger b \quad (2.13)$$

$$\hat{J}_x = \frac{1}{2}(\hat{J}_+ + \hat{J}_-) = \frac{1}{2}(b^\dagger a + a^\dagger b) \quad (2.14)$$

$$\hat{J}_y = \frac{1}{2i}(\hat{J}_+ - \hat{J}_-) = \frac{1}{2i}(b^\dagger a - a^\dagger b) \quad (2.15)$$

$$\hat{J}_z = \frac{1}{2}(b^\dagger b - a^\dagger a) \quad (2.16)$$

This is the Schwinger boson representation [29, 30] with built-in symmetrization among the individual qubits. It can be efficiently written as the Schwinger-Jordan map

$$\mathbf{J} = \begin{pmatrix} b^\dagger & a^\dagger \\ \sigma & \end{pmatrix} \frac{\sigma}{2} \begin{pmatrix} b \\ a \end{pmatrix}. \quad (2.17)$$

\hat{J}_z describes the occupation number difference between mode a and mode b . Strictly speaking, our observable is $\hat{\mathcal{J}}_z$, but we will drop this distinction for simplicity.

2.1.3 Collective rotations

Rotations of the qubit state $|\vartheta, \varphi\rangle$ on the Bloch sphere can be experimentally implemented by phase-controlled pulses of electromagnetic fields whose frequency is tuned close to resonance of the energy difference between $|\downarrow\rangle$ and $|\uparrow\rangle$. In our case, relaxation of $|\uparrow\rangle$ due to spontaneous processes is negligible on the experimental timescale, such that the single-qubit Hamiltonian that can be implemented in this way reads in rotating wave approximation

$$\hat{H}_{\text{rot}}^{(k)} = \frac{\Omega^{(k)}}{2} (\sin(\phi)\sigma_x^{(k)} + \cos(\phi)\sigma_y^{(k)}) + \frac{\delta_r^{(k)}}{2}\sigma_z^{(k)} \quad (2.18)$$

The direction of the rotation axis can be controlled by the phase ϕ of the radiation field and the detuning $\delta_r^{(k)}/2\pi = f_{\text{radiation}} - f_{\text{atom}}$ from atomic resonance. A qubit initially in state $|\downarrow\rangle$, performs Rabi oscillations with a frequency $\Omega^{(k)}/2\pi$ proportional to the amplitude of the applied field and $\phi = 0$ by convention. For $\delta_r = 0$ and $\Omega\tau_{\text{pulse}} = \pi/2$, that is a $\pi/2$ -pulse, the qubit is prepared in the state $|-\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle - |\downarrow\rangle)$, which is an equal superposition lying on the equator of the Bloch sphere and the starting point for most of our experiments. If the applied field is close to homogeneous over the ensemble, that is $\Omega_k \approx \Omega$ for every $k = 1 \dots N$ and detuning (magnetic field) gradients are sufficiently weak ($\delta_r^{(k)} \approx \delta_r$), the rotations can be described by the collective spin operators and the Hamiltonian reads

$$\hat{H}_{\text{rot}} = \sum_{k=1}^N \hat{H}_{\text{rot}}^{(k)} = \Omega (\sin(\phi)\hat{\mathcal{J}}_x + \cos(\phi)\hat{\mathcal{J}}_y) + \delta_r\hat{\mathcal{J}}_z \quad (2.19)$$

2.1.4 Coherent spin states

An important class of symmetric states are coherent spin states [31, 32] (CSS). They can be produced by collective rotations of the states $|N, 0\rangle$ or $|0, N\rangle$ and correspond to the preparation of each qubit in the same superposition:

$$|\vartheta, \varphi\rangle_N = |\vartheta, \varphi\rangle^{\otimes N} = \exp(i\varphi\hat{J}_z) \exp(i\vartheta\hat{J}_y) |N, 0\rangle \quad (2.20)$$

An explicit expansion in the Dicke state basis is given by

$$\begin{aligned} |\vartheta, \varphi\rangle_N &= \sum_{k=0}^N \binom{N}{k}^{\frac{1}{2}} \cos(\vartheta/2)^k \sin(\vartheta/2)^{N-k} \exp(-ik\varphi) |k, N-k\rangle \\ &= \sum_{k=0}^N \left[\binom{N}{k} p^k (1-p)^{N-k} \right]^{\frac{1}{2}} \exp(-ik\varphi) |k, N-k\rangle \end{aligned} \quad (2.21)$$

with $p = \cos^2(\vartheta/2)$. In the limiting case of $p = 0$ one defines $|\pi, \varphi\rangle_N = |0, N\rangle$ and likewise $|0, \varphi\rangle_N = |N, 0\rangle$ for $p = 1$. The CSS feature a binomial probability distribution in J_z , which can be intuitively understood, since the qubits are independently prepared with a probability p to be in the excited state $|\uparrow\rangle$. Upon measurement, each qubit is projected to $|\uparrow\rangle$ with probability p and $|\downarrow\rangle$ with probability $(1-p)$. The resulting distribution of the ensemble is therefore binomial like the one of N independent coin tosses with probability p for heads and $(1-p)$

for tails (unfair coin for $p \neq 1/2$). Continuing the example of $N = 3$, the coherent states on the equator of the Bloch sphere are explicitly

$$\left| \frac{\pi}{2}, \varphi \right\rangle = \frac{1}{\sqrt{8}} (|0, 3\rangle + \sqrt{3} e^{-i\varphi} |1, 2\rangle + \sqrt{3} e^{-2i\varphi} |2, 1\rangle + e^{-3i\varphi} |3, 0\rangle) \quad (2.22)$$

The coherent states are non-orthogonal since their overlap is given by

$$\begin{aligned} \langle \vartheta', \varphi' | \vartheta, \varphi \rangle &= \left[\cos \frac{\vartheta'}{2} \cos \frac{\vartheta}{2} e^{i(\varphi' - \varphi)} + \sin \frac{\vartheta'}{2} \sin \frac{\vartheta}{2} \right]^N \\ |\langle \vartheta', \varphi' | \vartheta, \varphi \rangle| &= \left(\frac{1 + \mathbf{n} \cdot \mathbf{n}'}{2} \right)^{\frac{N}{2}} = \cos \left(\frac{\alpha}{2} \right)^N, \end{aligned} \quad (2.23)$$

which is generally nonzero. Here, \mathbf{n} and \mathbf{n}' are the unit vectors of the mean spin directions and α is the angle between them, which is explicitly $\cos(\alpha) = \cos(\vartheta) \cos(\vartheta') + \sin(\vartheta) \sin(\vartheta') \cos(\varphi - \varphi')$. The overlap is strictly zero only if the mean spin directions are diametrically opposite ($\mathbf{n}' = -\mathbf{n}$). In the limit of big N , its absolute value can be approximated through

$$\cos \left(\frac{\alpha}{2} \right)^N \approx \left(1 - \frac{\alpha^2}{8} \right)^N \approx \exp \left[-\frac{N\alpha^2}{8} \right] = \exp \left[-\frac{1}{2} \frac{\alpha^2}{\sigma_\alpha^2} \right] \quad (2.24)$$

with $\sigma_\alpha = 2/\sqrt{N}$. The CSS become more and more localized around the mean spin direction with increasing N and even orthogonal in the classical limit $N \rightarrow \infty$. They form an overcomplete basis with the completeness relation

$$\frac{N+1}{4\pi} \int d\Omega |\vartheta, \varphi\rangle \langle \vartheta, \varphi| = 1, \quad (2.25)$$

where $d\Omega = \sin \vartheta d\vartheta d\varphi$ is the solid angle element. Due to the overcompleteness, off-diagonal elements $|\vartheta', \varphi'\rangle \langle \vartheta, \varphi|$ are unnecessary in the description.

A general mixed state is representable via

$$\hat{\rho} = \int d\Omega P(\vartheta, \varphi) |\vartheta, \varphi\rangle \langle \vartheta, \varphi|. \quad (2.26)$$

$P(\vartheta, \varphi)$ is called the P representation (Glauber–Sudarshan P in optics) and can be viewed as a quasiprobability distribution in the phase space (ϑ, φ) , in which a single CSS is a two-dimensional point-like delta-function [33]

$$P_{\text{CSS}}(\vartheta, \varphi) = \delta(\vartheta - \vartheta_0) \delta(\sin(\vartheta)[\varphi - \varphi_0]). \quad (2.27)$$

Here, the $\sin(\vartheta)$ cures the ambiguity of φ at the poles of the sphere, such that it is well defined everywhere.

The P function can be used to define non-classicality: A state that cannot be represented as a mixture of coherent states where P is consistent with a classical probability measure has no classical counterpart and is therefore genuinely quantum [34].

2.1.5 Quasiprobabilities

For visualization purposes, the state of N indistinguishable qubits can be displayed on a generalized Bloch sphere with the help of a quasiprobability distribution. Strictly speaking, such a procedure is a transformation to a semiclassical phase-space picture, which attaches a quasiprobability density value to each coherent state direction (ϑ, φ) . To make the scaling properties with atom number intuitive, one often uses the spin length $N/2$ as radius of the sphere, which is a legitimate but arbitrary choice.

Coming from the discussion of the coherent states and the attached definition of non-classicality, the most natural function to consider for this purpose seems to be the P -function. However, this has obvious complications, since the representation of a coherent spin state is already a delta-distribution. Only a minor redistribution of its quantum uncertainty around the mean spin direction (for example due to squeezing) leads to a P -function which is more singular than a delta function [35]. For a more intuitive view of this singularity, it is instructive to consider the limit where N is sufficiently big and the quantum state localized around the mean spin direction, such that the curvature of the sphere can be neglected. Then, a Holstein-Primakoff approximation [36] can be employed, which keeps two rescaled orthogonal spin operators perpendicular to the mean spin direction and drops the third direction (approximately constant). To simplify the notation, it is assumed that the mean spin direction is the z -direction. Then, the rescaled operators read

$$\hat{j}_{x,y} = \frac{1}{\sqrt{N}} \hat{J}_{x,y}, \quad \hat{j}_{\pm} = \frac{1}{\sqrt{N}} \hat{J}_{\pm} \quad (2.28)$$

and a general quadrature operator $\hat{j}_{\varphi} = (\hat{j}_{+} e^{i\varphi} + \hat{j}_{-} e^{-i\varphi})$ is introduced, which parametrizes all linear combinations of $\hat{j}_x = \hat{j}_0$ and $\hat{j}_y = \hat{j}_{\pi/2}$. This is the same type of approximation used for the analysis of balanced homodyne detection in optics, where the local oscillator field is treated classically [37] and the photocurrent difference has the same operational form $\Delta_{\text{photo}}/\sqrt{N_{\text{LO}}} = (\hat{a}^{\dagger} e^{i\varphi} + \hat{a} e^{-i\varphi})$. This enables one to apply the full framework of quasiprobability theory developed mainly for optical systems in one harmonic oscillator mode. Of course, this approximation becomes questionable when the quantum uncertainty of the spin state gets more extended, but is very useful to develop an intuition. The P -function in this framework is still a function of two variables, which is combined to a complex number $\alpha = |\alpha| e^{i\varphi} = \alpha_r + i\alpha_i$ and can be written in the form

$$P(\alpha) = \frac{1}{\pi^2} \int_{\mathbb{C}} d^2\beta \exp[i(\alpha_i\beta_r - \alpha_r\beta_i)] \Phi(\beta) \quad (2.29)$$

The characteristic function $\Phi(\beta)$ of the complex argument $\beta = \beta_r + i\beta_i$ is the two dimensional Fourier transform of $P(\alpha)$ [34] and can be written as [36]

$$\Phi(\beta) = \langle \exp[\beta \hat{j}_{-} - \beta^{*} \hat{j}_{+}] \rangle \exp[|\beta|^2/2] \quad (2.30)$$

Here, $\langle \cdot \rangle$ denotes the quantum mechanical expectation value. We know that the P -function of the coherent state is a δ -function, which is the Fourier transform of $\Phi(\beta) = 1$. Thus, the term in expectation value brackets is $\exp(-|\beta|^2/2)$ in this case. A similarly intuitive example is the squeezed state, whose characteristic function reads [38]

$$\Phi(\beta) = \exp \left[-\frac{\beta_r^2}{2} V_x - \frac{\beta_i^2}{2} V_y \right] \exp[|\beta|^2/2]. \quad (2.31)$$

V_x and V_y are the variances in the two orthogonal directions with $V_x V_y \geq 1$ due to the Heisenberg uncertainty product $\Delta \hat{J}_x \Delta \hat{J}_y \geq \langle \hat{J}_z \rangle / 2$. If one of them is smaller than one, the characteristic function is exponentially rising and its Fourier transform is more singular than the δ -function. In order to obtain a well-behaved function for all quantum states, appropriate filtering $\Phi_{\text{filt}}(\beta) = \Phi(\beta)F(\beta)$ has to be performed on the characteristic function [38] leading to different quasiprobability representations. A special class of such filters is given by

$$F(\beta, s) = \exp[-(1-s)|\beta|^2/2] \quad (2.32)$$

This leads to the most prominent ones, namely the Wigner function ($s = 0$) and the Husimi Q -function ($s = -1$), which can be nicely summarized in this way. If the filter is nonzero everywhere, the filtering procedure is theoretically reversible and the full information about the quantum state is preserved. If one wants to assign non-classical effects to negativities of the filtered quasiprobabilities, care has to be taken, that the filter itself does not introduce these negativities. This can be fulfilled, if the filter has a non-negative Fourier transform. A simple problematic example is a hard cutoff at a specific value of $|\beta|$, which Fourier transforms to an Airy-Disk with negativities. Recently, properly defined "nonclassicality filters" were introduced [38] and applied to show negative quasiprobabilities of optical and atomic squeezed states [35, 36]. Since $\Phi(\beta)/\exp[|\beta|^2/2]$ is Gaussian for the squeezed state, the observed structure in this case is mainly given by the Fourier transform of the filter (which is also bell-shaped with variable width) multiplied with $\exp[|\beta|^2/2]$, convoluted with the experimental Gaussian distribution.

Although the Holstein-Primakoff approximation becomes questionable for small atom numbers, it allows to develop a good intuition about quasiprobability distributions and their appropriate formulation in the spherical phase space of the atomic two-level system.

Husimi Q representation One particularly useful quasiprobability measure is the Husimi- Q function, which is easily computable and defined as

$$Q(\vartheta, \varphi) = \frac{N+1}{4\pi} \langle \vartheta, \varphi | \hat{\rho} | \vartheta, \varphi \rangle. \quad (2.33)$$

As already discussed, it forms a filtered version of the P representation, which can be observed, when inserting Eq. 2.26 in 2.33, yielding

$$\begin{aligned} Q(\vartheta, \varphi) &= \frac{N+1}{4\pi} \int d\Omega' P(\vartheta', \varphi') |\langle \vartheta', \varphi' | \vartheta, \varphi \rangle|^2 \\ &\approx \frac{N+1}{4\pi} \int d\Omega' P(\vartheta', \varphi') \exp[-N\alpha^2/4]. \end{aligned}$$

The approximation is valid for $N \gg 1$. Inserting P_{CSS} of the coherent spin state (Eq. 2.27), its Husimi representation $Q(\vartheta, \varphi) \propto \exp[-1/2(\alpha_0^2/\sigma_{\alpha_0}^2)]$ in the big- N limit is isotropically Gaussian around the mean spin direction (ϑ_0, φ_0) with an angular width $\sigma_{\alpha_0} = \sqrt{2/N}$. Compared with its angular uncertainty $1/\sqrt{N}$, this is a factor of $\sqrt{2}$ broader. Since the coherent spin state was a point-like δ -function in the P representation, this is the minimal structure size that can be observed with a positive P representation. As seen before, smaller

structures are only possible with negative values of P or points, at which it is more singular than a δ -function. The Husimi distribution is zero or positive everywhere and thus non-classical effects are less obvious due to the smoothing effect. Single Dicke states for example, whose quasiprobability distributions are ring-shaped around the Bloch sphere [39] feature a width in ϑ -direction on the order of $1/\sqrt{N}$ in the Q representation, which is a relatively weak effect in view of their highly non-classical character.

2.1.6 Interaction and the two-mode BEC

The discussion so far has not considered the spatial wavefunction of the qubits. Since they are embedded in the BEC, they spatially overlap and interact via binary collisions. The interaction part of the Hamiltonian in second quantization reads [40]

$$H_{\text{int}} = \int d^3x \left(\frac{g_{aa}}{2} \Psi_a^\dagger \Psi_a^\dagger \Psi_a \Psi_a + \frac{g_{bb}}{2} \Psi_b^\dagger \Psi_b^\dagger \Psi_b \Psi_b + g_{ab} \Psi_a^\dagger \Psi_b^\dagger \Psi_a \Psi_b \right) \quad (2.34)$$

where the field operator $\Psi^\dagger(\mathbf{x})$ creates a particle in spin state a at position \mathbf{x} and $g_{ij} = 4\pi\hbar^2 a_{ij}/m$ quantify the s-wave interaction strength with the scattering lengths a_{ij} of the two components among themselves and each other. For ultracold gases, scattering via partial waves with higher angular momentum are strongly suppressed, such that this is a good approximation. For repulsive interactions (positive scattering length) the two components are miscible if $g_{aa}g_{bb} \geq g_{ab}^2$, and tend to maximize their overlap [41]. For harmonic traps with high trapping frequencies, spatial dynamics are frozen out, and the field operators can be approximated as $\Psi_\alpha^\dagger(\mathbf{x}) = \psi_\alpha(\mathbf{x})\alpha^\dagger$ with a spatial mode function $\psi_\alpha(\mathbf{x})$ and the harmonic oscillator creation operators α^\dagger ($\alpha = a, b$). The mode functions can be found as the ground state of the Gross-Pitaevskii equation [41], assuming, that they are stationary and do not depend on the relative occupation number. With this two-mode approximation, the interaction Hamiltonian turns into

$$\begin{aligned} H_{\text{int}} &= X_{aa} a^\dagger a^\dagger a a + X_{bb} b^\dagger b^\dagger b b + 2X_{ab} a^\dagger b^\dagger a b \\ &= X_{aa} \hat{N}_a (\hat{N}_a - 1) + X_{bb} \hat{N}_b (\hat{N}_b - 1) + 2X_{ab} \hat{N}_a \hat{N}_b \end{aligned} \quad (2.35)$$

with the interaction constants

$$X_{\alpha\beta} = \frac{g_{\alpha\beta}}{2} \int d^3x |\psi_\alpha(\mathbf{x})|^2 |\psi_\beta(\mathbf{x})|^2. \quad (2.36)$$

\hat{N}_a and \hat{N}_b denote the occupation number operators of the two modes. Substituting $\hat{N}_a = N/2 - \hat{J}_z$ and $\hat{N}_b = N/2 + \hat{J}_z$ and neglecting constant terms (global energy shifts) yields

$$\begin{aligned} H_{\text{int}} &= (N - 1)(X_{bb} - X_{aa})\hat{J}_z + (X_{aa} + X_{bb} - 2X_{ab})\hat{J}_z^2 \\ &= \delta_{\text{int}}\hat{J}_z + \chi\hat{J}_z^2 \end{aligned}$$

The interaction thus yields a linear term, which leads to an additional phase evolution like the previous detuning from atomic resonance (density- or meanfield-shift) and the nonlinear term, which generates squeezing and entanglement and is called one-axis twisting Hamiltonian [11]. It constitutes a phase evolution which itself depends on the relative occupation numbers of the

two modes. This is a very non-local operation among the individual qubits, as can be seen, when writing this Hamiltonian in the individual particle spins as

$$\hat{J}_z^2 = \left(\frac{1}{2} \sum_{j=1}^N \sigma_z^{(j)} \right)^2 = \frac{N}{4} + \frac{1}{2} \sum_{j < k}^N \sigma_z^{(j)} \sigma_z^{(k)}, \quad (2.37)$$

which is valid in the symmetric subspace. It is the interaction part of the fully connected Ising model and shows that for $\chi \neq 0$ every qubit directly interacts nonlocally with every other. The phase evolution of a single spin thus depends on the population imbalance of all others.

The time evolution under this Hamiltonian starting from a coherent spin state on the equator of the Bloch sphere can be intuitively grasped as redistribution (shearing) of the quantum uncertainty parallel to the equator, since every constituent J_z -eigenstate $|k, N - k\rangle$ acquires a different phase during the evolution time t_{evo} . With the expansion of the coherent spin states (Eq. 2.21) this reads for example

$$\begin{aligned} |\Psi(t_{\text{evo}})\rangle &= \exp \left[-i\chi t_{\text{evo}} \hat{J}_z^2 \right] \left| \frac{\pi}{2}, 0 \right\rangle_N \\ &= \sum_{k=0}^N \sqrt{\binom{N}{k} \frac{1}{2^N}} \exp \left[-i\chi t_{\text{evo}} \left(k - \frac{N}{2} \right)^2 \right] |k, N - k\rangle \end{aligned}$$

The evolution results in squeezed states up to a timescale $\sim 1/\chi\sqrt{N}$ and culminates in a phase cat state (equal superposition of two coherent states diametrically opposite on the equator of the Bloch sphere) for half of the recurrence time $t_{\text{rec}} = \pi/\chi$. This fragile state however is far beyond what can be currently reached in the BEC system, but impressive results have been achieved in cavity QED systems [42, 43] and few trapped atomic ions [44, 45]. In the quantum information context, the quantum evolution up to $t_{\text{rec}}/2$ is known as Mølmer-Sørensen [26] entangling quantum gate, realizable in ion systems via collective trap oscillations with direct driving [46] or geometric phase imprints [47, 48].

Further insight into this interaction can be gained by rotating the coordinate system by 90 degrees, such that the initial state is the coherent spin state $|0, 0\rangle_N$ (corresponding to all atoms in $|\uparrow\rangle$) and \hat{J}_z^2 turns into \hat{J}_x^2 interaction. This situation is usually found in discussions on the generation of the squeezed vacuum state [49] in quantum optics or the Mølmer-Sørensen interaction. In second quantization, one can expand

$$\chi \hat{J}_x^2 = \frac{\chi}{4} (b^\dagger a + a^\dagger b)^2 = \frac{\chi}{4} (b^\dagger b^\dagger a a + a^\dagger a^\dagger b b + 2\hat{N}_a(N - \hat{N}_a) + N). \quad (2.38)$$

One observes, that from this perspective the action of \hat{J}_x^2 is pair production of excitations of the initially empty mode ψ_a . For short evolution time and big atom numbers N , one can approximate the operator of the highly populated mode with a c-number ($b^2 \approx N \exp[i2\phi_b] = \zeta$) and $\hat{N}_a \approx 0$. In quantum optics, this is called the undepleted pump approximation [50]. It leads to

$$\chi \hat{J}_x^2 \approx \frac{\chi}{4} (\zeta^* a^2 + \zeta a^{\dagger 2}). \quad (2.39)$$

The corresponding unitary time evolution operator $\exp[-it\chi(\zeta^* a^2 + \zeta a^{\dagger 2})/4]$ is exactly the squeezing operator found in textbook discussions of squeezed light [51, 52]. The analogy is

good, as long as the local approximation of the sphere as flat surface over the extension of the quantum state is justified. For longer evolution time, the spherical topology of the phase space becomes important and the 'undepleted pump approximation' becomes invalid as the state begins to wrap around the Bloch sphere.

Rabi coupling A similar treatment for external Rabi coupling [40], which leads to interconversion of the two components, yields an effective Rabi frequency

$$\tilde{\Omega} = \Omega \int d^3x \psi_a^*(\mathbf{x})\psi_b(\mathbf{x}) \leq \Omega. \quad (2.40)$$

However, this is only a valid description when the interconversion is adiabatic, meaning that the spatial mode has to be able to adapt faster than the Rabi drive is changing its occupation. This is because the Rabi coupling microscopically constitutes a single particle addressing. For example, if all atoms are initially prepared in ψ_a , which would not perfectly overlap with the other equilibrium mode ψ_b , the application of very strong Rabi coupling for $\tau_{\text{pulse}} = \pi/\Omega$ would lead to a transfer of all atoms to $|\uparrow\rangle$ but the spatial mode would still be ψ_a . The Rabi frequency in this limit would correspond to the bare Ω , not to $\tilde{\Omega}$ and the subsequent time evolution would be an oscillatory motion of the atomic cloud in the trapping potential around the equilibrium position of the mode ψ_b . In our case, the optical traps are spin-independent to a very good approximation and interaction effects on the shape of the two modes negligible, such that $\psi_a \approx \psi_b$ and $\tilde{\Omega} \approx \Omega$. When this single spatial mode approximation becomes invalid, the interplay of strong Rabi coupling and interaction leads to demixing effects also in the initially miscible situation, which is counterintuitive but can be understood as demixing of dressed states [41], which we also studied and published in [53].

2.1.7 Collective spin Hamiltonian

In the two-mode approximation with the choice $\phi = 3\pi/2$ for the phase of the Rabi coupling (Eq. 2.19) the relevant spin Hamiltonian becomes

$$\hat{H} = \chi \hat{J}_z^2 - \Omega \hat{J}_x + \delta \hat{J}_z \quad (2.41)$$

with the detuning $\delta = \delta_r + \delta_{\text{int}}$ being a combination of the detuning of the coupling radiation from the bare atomic resonance and the interaction induced shift. This is a special case of the Lipkin-Meshkov-Glick Hamiltonian [54] studied in nuclear physics. The corresponding time evolution it induces features very rich phenomena depending on the relative strength of the three ingredients. Its analogy to the Josephson phenomenology of two weakly coupled superfluids will be shown in the next section. In qubit operator basis it reads

$$\hat{H} - \text{const.} = \frac{\chi}{2} \sum_{j < k} \sigma_z^{(j)} \sigma_z^{(k)} - \frac{\Omega}{2} \sum_j \sigma_x^{(j)} + \frac{\delta}{2} \sum_j \sigma_z^{(j)}, \quad (2.42)$$

which is the Hamiltonian of the fully connected Ising model [55] in a combined transverse and axial magnetic field.

2.1.8 Mean field limit and classical trajectories

In the limit $N \rightarrow \infty$ the finite quantum uncertainty $1/\sqrt{N}$ of the coherent state around the mean spin direction can be neglected and the operators replaced by their expectation value for a coherent spin state $|\vartheta, \varphi\rangle$, explicitly

$$\begin{aligned} (\hat{J}_x, \hat{J}_y, \hat{J}_z) &\rightarrow (\langle \hat{J}_x \rangle, \langle \hat{J}_y \rangle, \langle \hat{J}_z \rangle) = \frac{N}{2}(x, y, z) \\ &= \frac{N}{2}(\sin \vartheta \cos \varphi, \sin \vartheta \sin \varphi, \cos \vartheta) \\ &= \frac{N}{2}(\sqrt{1-z^2} \cos \varphi, \sqrt{1-z^2} \sin \varphi, z) \end{aligned} \quad (2.43)$$

The Hamiltonian in this limit becomes

$$H = \frac{N}{2} \left(\frac{N\chi}{2} z^2 - \Omega x + \delta z \right) \quad (2.44)$$

The equations of motion can be derived from the Ehrenfest theorem, which reads for the spin operators

$$\frac{d}{dt} \langle \hat{J}_\alpha \rangle = \frac{1}{i} \langle [\hat{J}_\alpha, \hat{H}] \rangle + \left\langle \frac{\partial \hat{J}_\alpha}{\partial t} \right\rangle. \quad (2.45)$$

The operators have no explicit time dependence, so only the result of the commutator is relevant. After the substitutions $\langle \hat{J}_\alpha \hat{J}_\beta \rangle = \langle \hat{J}_\alpha \rangle \langle \hat{J}_\beta \rangle$ this leads to the equations of motion

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} -N\chi zy - \delta y \\ N\chi zx + \delta x + \Omega z \\ -\Omega y \end{pmatrix}. \quad (2.46)$$

Their solutions are the classical trajectories, which are isopotential lines of Eq. 2.44. Owing to the spherical topology of the phase space, the equations of motion obtain the compactest form in the angle coordinates $(\vartheta, \varphi) = (\arccos(z), \arctan(y/x))$ with the time derivatives $\dot{\vartheta} = -\dot{z}/\sqrt{1-z^2}$ and $\dot{\varphi} = (\dot{y}x - y\dot{x})/(x^2 + y^2)$, where they read

$$\begin{pmatrix} \dot{\vartheta} \\ \dot{\varphi} \end{pmatrix} = \Omega \begin{pmatrix} -\sin \varphi \\ \Lambda \cos \vartheta + \cos \varphi / \tan \vartheta + \tilde{\delta} \end{pmatrix} \quad (2.47)$$

The Rabi frequency Ω determines the timescale of the dynamics and the dimensionless parameters $\Lambda = N\chi/\Omega$ and $\tilde{\delta} = \delta/\Omega$ the shape and symmetry properties of the phase space portrait. It is shown in Fig. 2.2 for the symmetric case $\delta = 0$. In the experimental situation, Λ is controlled by the Rabi coupling strength, so for fixed detuning, the parameter δ/Ω changes as well. This situation is shown in Fig. 2.3 for a constant detuning $\delta = 2\pi \times 2$ Hz. We reparametrize once more to obtain the equations of motion for the the population imbalance z and the relative phase φ between the two modes, which is mostly discussed in the literature [56–61] and reads

$$\begin{pmatrix} \dot{z} \\ \dot{\varphi} \end{pmatrix} = \Omega \begin{pmatrix} -\sqrt{1-z^2} \sin \varphi \\ \Lambda z + (z/\sqrt{1-z^2}) \cos \varphi + \tilde{\delta} \end{pmatrix}. \quad (2.48)$$

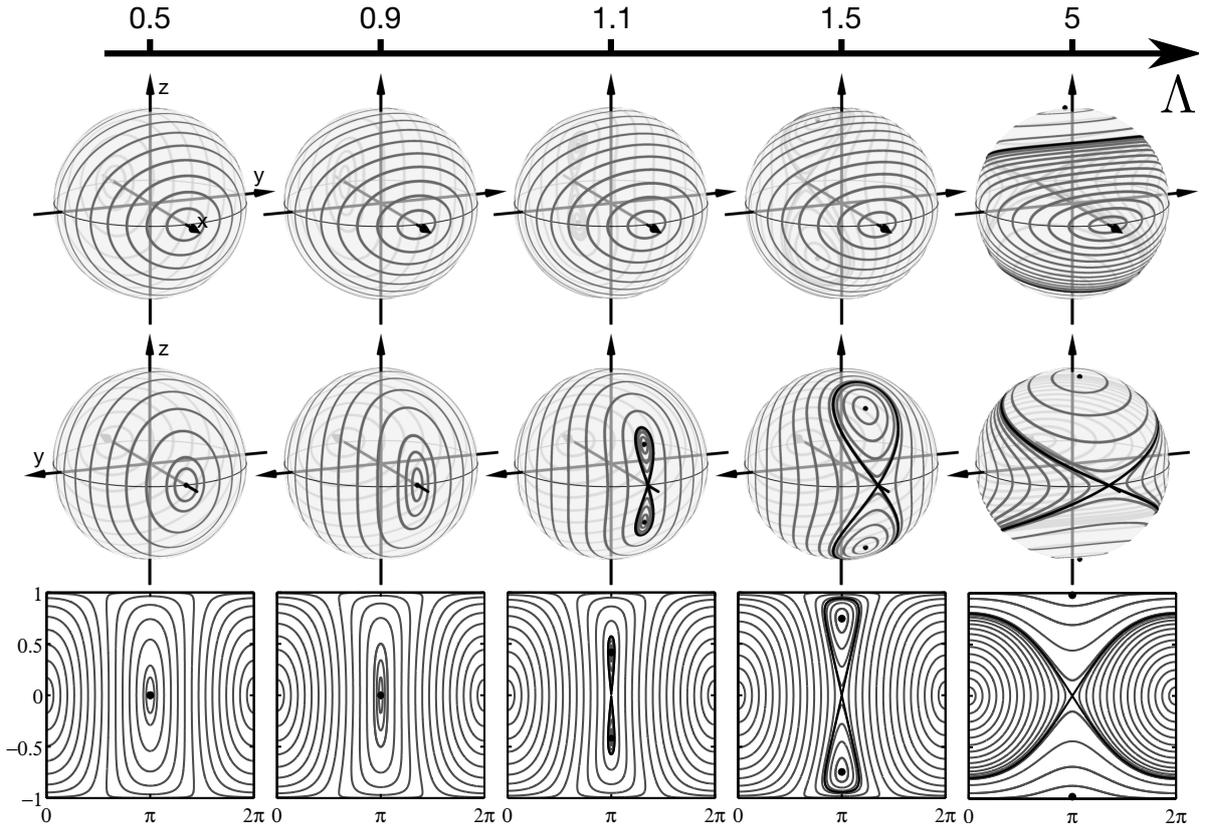


Figure 2.2 Phase space picture of the weakly coupled interacting two-mode BEC. In the classical limit $N \rightarrow \infty$ the quantum uncertainty can be neglected, which yields classical equations of motion for the mean spin direction. The resulting trajectories are shown on the sphere and as flat phase space portraits of the variable φ and its canonical momentum z for the resonant case $\delta = 0$ and various values of the system parameter $\Lambda = N\chi/\Omega$, which quantifies the relative strength of the interaction energy scale $N\chi$ and the Rabi coupling Ω . For small Λ , the dynamics is dominated by Ω , which results in slightly anharmonic oscillations with altered frequency (plasma- and π -oscillations). The phase space features two stable fixed points $F_0 = (0, 0)$ and $F_\pi = (0, \pi)$ on the x -axis, which is the rotation axis. At the critical value $\Lambda = 1$, the stable fixed point F_π bifurcates into two new stable fixed points $F_\pm = (\pm\sqrt{1 - 1/\Lambda^2}, \pi)$ and F_π itself becomes an unstable hyperbolic fixed point. The separatrix (black line) marks the border between trajectories with $\langle z \rangle \neq 0$ (macroscopic quantum self trapping) and those with $\langle z \rangle = 0$. For $\Lambda > 2$, the separatrix embraces the two poles, which results in possible trajectories with φ monotonically increasing (running phase modes).

The classical Hamiltonian in these coordinates becomes

$$H = \frac{N\Omega}{2} \left(\frac{\Lambda}{2} z^2 - \sqrt{1 - z^2} \cos \varphi + \tilde{\delta} z \right). \quad (2.49)$$

The dynamics of the system features a rich variety of phenomena, some of which are reminiscent of the behavior of a Josephson junction [57, 62] in superconducting materials (SJJ). In this case, the two modes are the macroscopically occupied wave functions of cooper pairs divided by a thin membrane, which provides a weak link of the two sides via tunneling.

Our system is the implementation of an internal bosonic Josephson junction (BJJ) in analogy to the external version of a BEC in a double well potential, where the weak link (Rabi coupling)

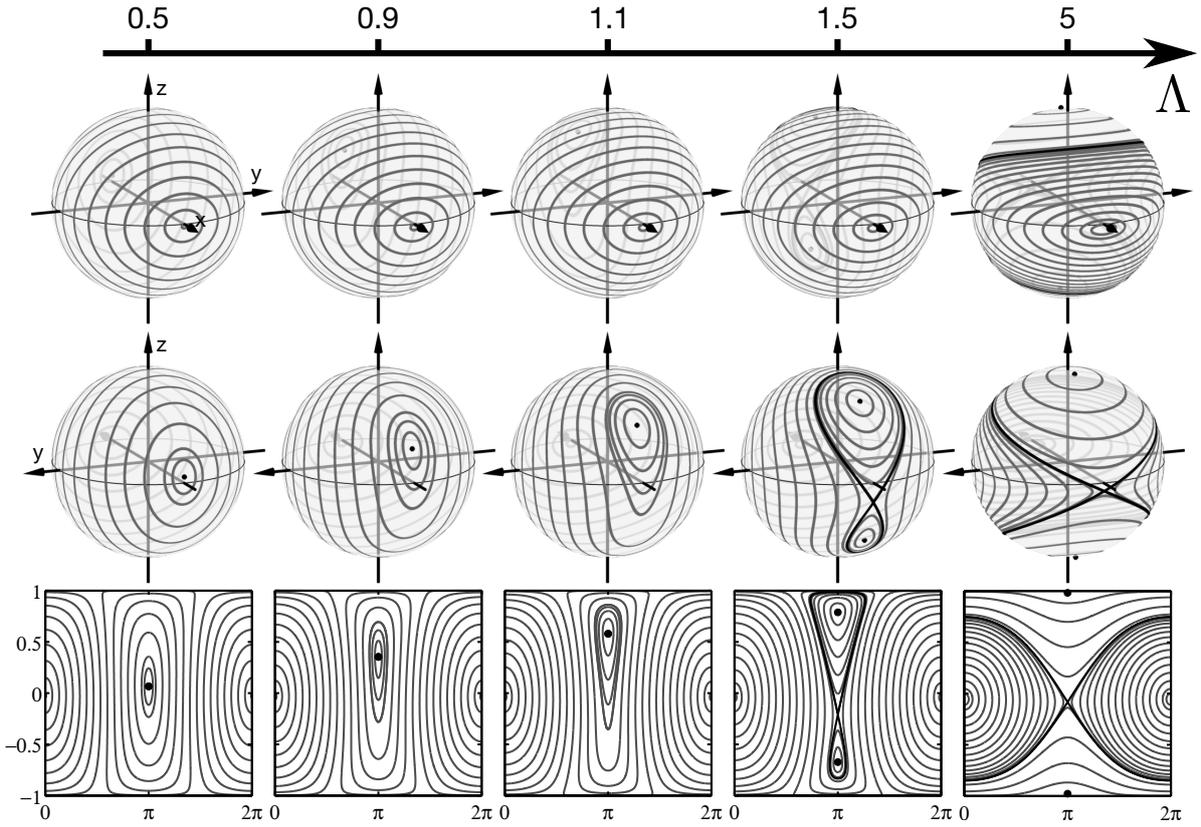


Figure 2.3 The classical phase space for a fixed detuning of $\delta = 2\pi \times 2$ Hz. The interaction energy scale $N\chi = 2\pi \times 30$ Hz is kept constant and Λ adjusted via the Rabi frequency. Decreasing Ω for increasing Λ leads to an upward shift of the fixed point F_π and eventually to the appearance of a second stable and the unstable fixed point (bifurcation "out of blue sky"). For even smaller Ω the system is again dominated by interaction and the detuning merely leads to a shift towards smaller z compared to the resonant case.

is provided by a controllable tunnel barrier, also realized in our group with a precursor of the current experimental setup [63, 64]. The big advantage of the internal system is that the Rabi coupling can easily overwhelm the interaction effects and allows for a very controlled preparation and characterization of the spin states.

For the BJJ all values of $-1 \leq z \leq 1$ are accessible, whereas the SJJ is mostly at $z \approx 0$ due to the external circuit. Furthermore, the BJJ features various nonlinear behavior like macroscopic quantum self trapping at $z \neq 0$ [57, 59, 65], which is a unique effect of dominant interaction. The BJJ has a mechanical analog, where z is the canonical momentum of the displacement angle φ of a pendulum whose length gets shortened by the factor $\sqrt{1 - z^2}$ of its momentum [58]. The striking phenomena are in addition to oscillations with $\langle \varphi \rangle = 0$ and running phase modes (complete revolutions around the pivot) of the standard pendulum two oscillation modes with $\langle \varphi \rangle = \pi$ around the upward standing position. We will show below, that by a coordinate transformation, these complicated behavior can be precisely mapped to a *rigid* pendulum by a coordinate transformation.

The interesting case for nonclassical state generation is $\Lambda > 1$, where the phase space (z, φ) features three stable fixed points ($F_0 = (0, 0)$ and $F_\pm = (\pm\sqrt{1 - 1/\Lambda^2}, \pi)$) and the unstable

hyperbolic fixed point $F_\pi = (0, \pi)$ on the negative x -axis. The energy landscape given by the classical Hamiltonian obtains a saddle point there (quadratically increasing in z -direction and decreasing in φ -direction for positive parameters). The regime $\Lambda \ll 1$ of dominant Rabi coupling is used for preparation and characterization of the quantum state. In the intermediate regime $0 < \Lambda < 1$, the frequency of small amplitude oscillations around the two stable fixed points (plasma- and π -oscillations) are very useful for an independent characterization of the nonlinearity. A linearization ($\sqrt{1 - z^2} \approx 1$, $\sin \varphi \approx \pm \varphi$ and $\cos \varphi \approx \pm 1$) of the equations of motion for $\delta = 0$ around $(0, 0)$ and $(0, \pi)$ gives $\ddot{z} = -\Omega^2(1 + \Lambda)z = -\omega_{\text{pl}}^2 z$ and $\ddot{z} = -\Omega^2(1 - \Lambda)z = -\omega_\pi^2 z$ whose solutions are sinusoidal with the frequencies $\omega_{\text{pl}} = \Omega\sqrt{1 + \Lambda}$ and $\omega_\pi = \Omega\sqrt{1 - \Lambda}$. Observation of both allows to extract the parameters Ω and χ as

$$\begin{aligned}\Omega &= \sqrt{(\omega_{\text{pl}}^2 + \omega_\pi^2)/2} \\ \chi &= (\omega_{\text{pl}}^2 - \omega_\pi^2)/(2N\Omega).\end{aligned}\quad (2.50)$$

One has to keep in mind however that this is only valid for small amplitudes, since larger oscillations are anharmonic. A similar harmonic linearization around the two self-trapping fixed points in the bifurcated regime yields the equations of motion $\dot{\varphi} = \Omega\Lambda(1 - \Lambda^2)(z - z_\pm)$ and $\dot{z} = \Omega\varphi/\Lambda$ and thus $\ddot{z} = -\omega_\pm^2 z$ with the frequency $\omega_\pm = \Omega\sqrt{\Lambda^2 - 1}$, which is bigger than Ω . For $\Omega \rightarrow 0$, i.e. $\Lambda \rightarrow \infty$ but $\Omega\Lambda = N\chi = \text{const.}$, these two stable fixed points coincide with the poles of the sphere, the two fixed points on the equator disappear and the frequency $\omega_\pm \rightarrow N\chi$ corresponds to the phase evolution due to the nonlinearity (see Eq. 2.48), which is maximal at the poles $z = \pm 1$.

2.1.9 Mapping to a rigid pendulum

Here we will show how the meanfield limit of the nonlinear system in the case of zero detuning can be mapped onto the motion of a one-dimensional mathematical pendulum with the help of an elementary coordinate transformation.

We first note that the trajectories are equipotential lines of the classical Hamiltonian and for fixed spin length we have $z^2 = 1 - x^2 - y^2$, thus the classical Hamiltonian yields with $\delta = 0$

$$\frac{N\chi}{2}(1 - x^2 - y^2) - \Omega x = \text{const.}\quad (2.51)$$

By completing the square, we get

$$\left(x + \frac{1}{\Lambda}\right)^2 + y^2 = r^2 = \text{const.}\quad (2.52)$$

This is the equation of a circle, which means that all trajectories including the separatrix upon projection to the x - y -plane fall onto circular lines¹. The center of all these circles is the point $(x, y) = (-1/\Lambda, 0)$, which will turn out to be the pivot of the mathematical pendulum swinging in the x - y -plane. We define the phase of the pendulum as $\phi = \arcsin(y/r)$. The first time derivative yields with the equations of motion (Eq. 2.46)

$$\frac{d\phi}{dt} = \frac{1}{\sqrt{r^2 - y^2}} \frac{dy}{dt} = \frac{(N\chi x + \Omega)z}{\sqrt{r^2 - y^2}} = \frac{(x + 1/\Lambda)N\chi z}{\sqrt{r^2 - y^2}}.\quad (2.53)$$

¹This means that the separatrix and all trajectories are section lines of cylinders with the unit sphere.

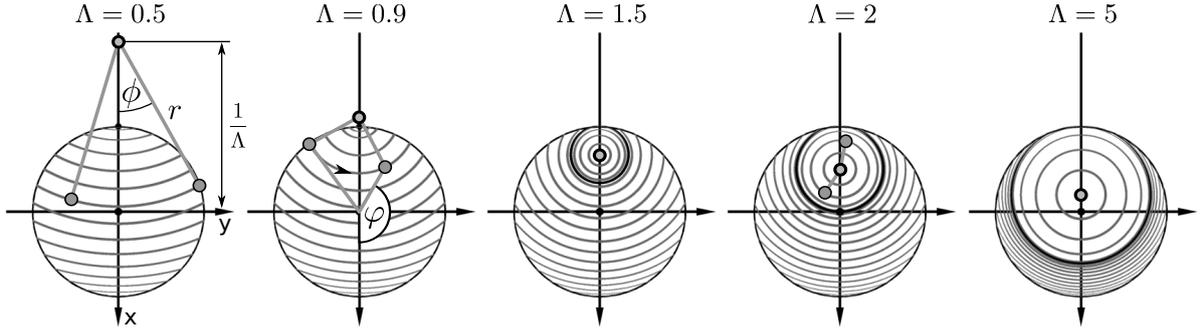


Figure 2.4 Rigid pendulum mapping of the Bose Josephson Junction mean-field limit For the case of zero detuning δ , all trajectories including the separatrix fall onto circular lines with center $(x, y) = (-1/\Lambda, 0)$ upon projection to the x - y -plane. The temporal dynamics can be mapped onto a *rigid* mathematical one-dimensional pendulum with the angle $\phi = \arcsin(y/r)$ as dynamical variable and the atomic imbalance $z = \dot{\phi}/N\chi$ its conjugate momentum. Due to the restriction to the unit circle, only oscillations with $\langle \phi \rangle = 0$ are possible for $\Lambda < 1$. For increasing Λ , the pivot of the pendulum moves towards the origin. If the pivot point lies inside the unit circle, the pendulum is free to perform also complete revolutions, in clockwise direction for $z < 0$ and counterclockwise for $z > 0$. This is the bifurcation scenario [61]. For comparison with the non-rigid 'momentum shortened' pendulum analogy [58] employing the atomic phase φ as dynamical variable, we depict a complicated inverted mode for the case $\Lambda = 0.9$.

With Eq. 2.52, the square root cancels and we have

$$\dot{\phi} = N\chi z. \quad (2.54)$$

The second derivative yields

$$\ddot{\phi} = N\chi \dot{z} = -N\chi\Omega y = -N\chi\Omega r \sin \phi. \quad (2.55)$$

This yields the equation of motion of a one-dimensional mathematical pendulum

$$\ddot{\phi} + \omega_0^2 \sin \phi = 0 \quad (2.56)$$

with the small amplitude angular frequency $\omega_0 = \sqrt{N\chi\Omega r} = \Omega\sqrt{\Lambda r}$. The atomic imbalance z takes the role of the conjugate momentum: $z = \dot{\phi}/N\chi$ and the classical Hamiltonian of this pendulum takes the form

$$\tilde{H} = \frac{N\chi}{2} z^2 - \Omega r \cos \phi \quad (2.57)$$

with the canonical equations of motion

$$\begin{aligned} \frac{dz}{dt} &= -\frac{\partial \tilde{H}}{\partial \phi} = -\Omega r \sin \phi \\ \frac{d\phi}{dt} &= \frac{\partial \tilde{H}}{\partial z} = N\chi z. \end{aligned} \quad (2.58)$$

The pendulum obtains its largest momentum, when $\phi = 0$, where it can be noted most conveniently as

$$\dot{\phi}|_{\phi=0} = N\chi \sqrt{1 - x^2 - y^2}|_{\phi=0} = \Omega \sqrt{\Lambda^2 - (\Lambda r - 1)^2}. \quad (2.59)$$

We note that there is no momentum shortening of the pendulum involved. Since the trajectories are confined to the sphere, this allows to grasp the bifurcation scenario. For $\Lambda < 1$ the pivot lies outside the unit circle and only oscillations with downward mean bob position are possible. The pivot approaches the unit circle for $\Lambda \rightarrow 1$ and crosses it at $\Lambda = 1$. This is where the bifurcation happens. For $\Lambda > 1$ the pivot is inside the unit circle and trajectories with full revolutions around the pivot are possible². In this case, the separatrix is the trajectory with radius $r = (1 - 1/\Lambda)$, that touches the unit circle. The trajectories inside the separatrix are those corresponding to macroscopic quantum self-trapping (MQST) with the bob turning clockwise for $z < 0$ and counter-clockwise for $z > 0$. For $\Lambda > 2$ the separatrix encloses the origin, which enables windings of the atomic phase φ (running phase modes), which in our picture corresponds to the same mode of the pendulum as the π -self-trapped trajectories of [57, 58]. Both feature a running phase ϕ . Equation 2.59 explains the frequency close to the self-trapped stable fixed points, since for $r \rightarrow 0$ (close to the fixed point, which is equal to the pivot in this case), we get the self-trapping frequency $\Omega\sqrt{\Lambda^2 - 1}$ as obtained in 2.1.8 with a linearization of the equations of motion.

For $\Lambda \rightarrow \infty$ ($\Omega \rightarrow 0$), the pivot approaches the origin and the phase becomes equal to the atomic phase ($\phi \rightarrow \varphi$). In this case, the angular momentum stays constant ($\dot{z} = 0$) and the phase evolves with $\dot{\phi} = N\chi z$, corresponding to one-axis twisting.

We note that the 'gravitational acceleration', which points into positive x -direction for our pendulum is $N\chi\Omega r^2$ and thus it has no obvious mechanical analogue. The oscillation frequency increases with the pendulum length contrary to a 'normal' pendulum in gravity. Specifically, for $\Lambda < 1$ we recover the frequencies of plasma-oscillations with $r = 1/\Lambda + 1$ and of π -oscillations with $r = 1/\Lambda - 1$. Note that due to the conservation of the spin length, every ϕ is connected to two possible momenta z only. All realizable turning points of the pendulum lie on the unit circle, where $z = 0$.

Our picture gives a simplified and unifying handle on the periods and temporal shape of the mean-field oscillations, since they are given by those of the mathematical pendulum, which are known very well in terms of elliptic integrals and Jacobi elliptic functions, respectively [66], which was also obtained in [57] for the momentum shortened pendulum. Maybe more importantly, it identifies the correct angle to judge the validity of the small-angle approximation over the whole parameter regime and phase space, which is much less obvious with the atomic phase φ . For example, with $\Lambda < 1$, there are trajectories, where the atomic phase (and thus the momentum-shortened pendulum) makes excursions by almost $\pm\pi/2$, but the oscillation is close to harmonic, which is explained by the rigid pendulum mapping. The limit of strong Rabi coupling is also nicely included with $\Lambda \rightarrow 0$, the pendulum length $r \rightarrow \infty$, but $\Lambda r \rightarrow 1$, which gives $\omega_0 \rightarrow \Omega$. All trajectories are close to harmonic in this case, since the small-angle approximation becomes increasingly better.

Our pendulum analogy does not include an approximation and is different to the standard Josephson mapping, since it does not use the (natural) macroscopic phase as dynamic variable. It is exact, if the rotation axis of the Rabi coupling is perpendicular to the nonlinear twist. In the detuned case, the effective rotation axis obtains a tilt in z -direction and the x - y planar symmetry is broken. However, the pendulum picture still provides a good intuition, if the detuning is small compared to Ω and $N\chi$.

²To simplify the notations, we take $\Lambda > 0$. The case $\Lambda < 0$ corresponds to a rotation of the picture by 180°

In the bifurcated regime, the period of trajectories close to the separatrix diverges. For the creation of non-classical states, we prepare a coherent spin state in the regime $\Lambda \ll 1$ with the pivot on the positive y -axis. We then change the direction of the pivot by $\pi/2$ to the negative x -axis and quench the pivot position nonadiabatically to $1/\Lambda < 1$ by decreasing the Rabi frequency. This prepares the inverted pendulum situation which, due to the coherent state width in z corresponds to a preparation of a 'superposition of trajectories' close to the separatrix, turning clockwise for $z < 0$ and counterclockwise for $z > 0$ during the time evolution. The pendulum falls down to both sides simultaneously.

After writing, we have become aware of [67], where the mapping to the mathematical pendulum was first recognized.

Comparison to the Josephson pendulum We have been describing a system of two coupled largely populated bosonic modes as in the effective description of the Josephson effect of superconductors. However, for the 'classic' Josephson effect, all couplings are *linear*. In our language, the classic Josephson effect would be described by the Hamiltonian $H_J = -\Omega x + \delta z$, i.e. Rabi coupling with a finite detuning. The Josephson equations ($\dot{z} = \Omega \sin \varphi$; $\dot{\varphi} = \delta$) come about by considering $z \approx 0$, which is maintained by the external circuit. If this would not be the case, the Josephson effect would correspond to a (tilted) linear rotation and thus to a rigid rotor. The dc Josephson effect ($\delta = 0$) is explained by the changing projection of the mean spin vector onto the rotation axis with the phase φ . When the mean spin vector is parallel with the rotation axis, the coupling and therefore the current through the junction vanishes ($\dot{z} = 0$), whereas it gets maximal for a relative angle of $\pi/2$. For the ac Josephson effect ($\delta \neq 0$), the phase φ changes with a constant rate (corresponding to the external voltage) and thus the current varies sinusoidally.

An isolated Josephson junction with finite parallel capacitance becomes nonlinear and can be mapped to a pendulum, again for $z \approx 0$ [68]. At first sight, the nonlinearity of our system seems different in nature, since it is based on interactions between the constituent qubits. However the effect is very similar, since the interaction in our case is also shared over the whole ensemble in analogy to the capacitively shunted Josephson junction.

Yet, the approach to employ the nonlinearity is very different in the two systems. For superconducting qubit devices, the nonlinearity is typically used to lift the splitting degeneracy in the energy spectrum for macroscopic (phase) states with $z \approx 0$, such that two of them can be addressed and used as single qubit [69]. In contrast, we deal with nonequilibrium dynamics of the interacting system, which populates a large number of eigenstates and evolves into a nonclassical state. It eventually spreads over a big range of imbalances z (corresponding to a big variation in the number of cooper pairs on the capacitor) and/or phases φ .

2.1.10 Quantum evolution from the classically unstable fixed point

Figure 2.5 shows the ideal quantum evolution obtained by numerical diagonalization of the spin Hamiltonian (Eq. 2.41) starting with a coherent spin state placed on the classically unstable fixed point. The spread of the initial quantum uncertainty during the time evolution follows the 'velocity field' given by the mean-field trajectories close to the separatrix. Short evolution times

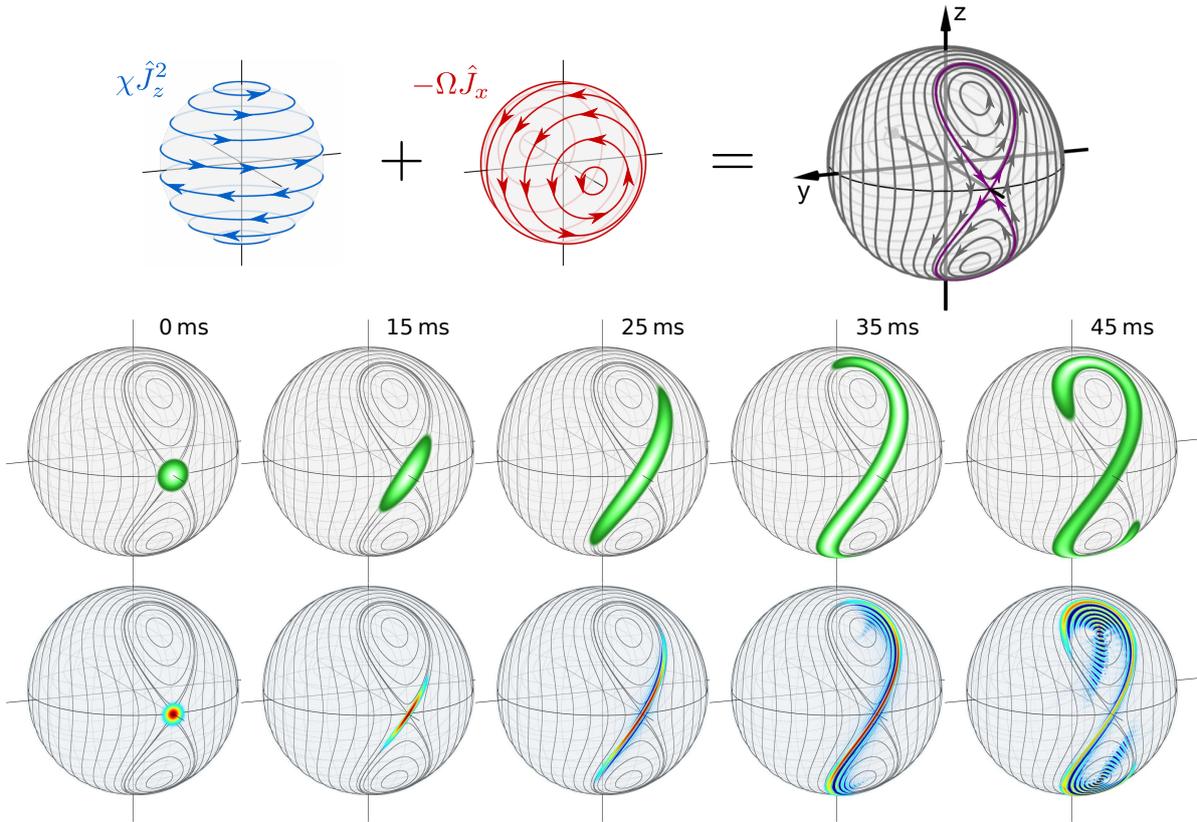


Figure 2.5 Phase space picture of the ideal quantum evolution. **Top row:** Pictorial mean-field representation of the two contributions nonlinearity and linear coupling, which leads in combination (here $\Lambda = N\chi/\Omega = 1.5$ and $\delta = 0$) to the bifurcated phase space. **Middle and bottom row:** Numerical integration of the time evolution starting with a coherent spin state on the classically unstable fixed point (corresponding to the lowest eigenstate of \hat{J}_x) for $N = 380$ atoms, $\Lambda = 1.5$ and $\Omega = 2\pi \times 20$ Hz. The middle row shows the Husimi projection and the bottom row the corresponding Wigner distribution. Short evolution times lead to spin squeezed states, which develop into non-Gaussian states as the quantum uncertainty 'wraps' around the two stable fixed points above and below the equator. This is accompanied by negativities and rapid oscillations of the Wigner function. For better visibility, both distributions are rescaled to their maximum for each evolution time. The classical trajectories, exactly valid for $N \rightarrow \infty$ provide an excellent intuitive guideline to understand the quantum evolution for a finite number of atoms. The Husimi distribution is positive everywhere, but still carries the full information about the quantum state. The Wigner function has the nice property, that the observable spin distribution for some projection axis is obtained by integration in perpendicular direction (marginal distributions). This can serve as basis for quantum state tomography [49, 74].

result in spin squeezed states, which turn into non-Gaussian states for later evolution times as the state begins to bend around the two stable fixed points. The Wigner distribution obtains negativities and rapid oscillations, which can be interpreted as quantum interference in phase space [70, 71]. Compared to pure one-axis twisting ($\chi \hat{J}_z^2$), the squeezing evolution as well as the transition to non-Gaussian states is accelerated by the addition of linear coupling [40, 72, 73]. For the quantitative characteristics of these states see section 2.6.

2.2 Fisher Information and the Cramér-Rao bound

The discussion of statistical parameter estimation dates back to Edgeworth, Fisher, Cramer and Rao [75–79], who laid the grounds for a quantitative and general understanding of the precision on a probability distribution parameter which can be obtained from a finite number of repetitions of an experiment. The role of the Fisher information is to quantify the amount of knowledge about a parameter that can be extracted from observations of a fluctuating observable. In our case, the fluctuating observable will be the imbalance $z = (N_b - N_a)/(N_b + N_a)$ and the parameter a phase (or angle) θ , since every unitary change on a quantum system can be characterized by a phase. The basis of the discussion are the conditional probability distributions $P_z(\theta) = P(z|\theta)$ which characterize the measurement outcomes given the value θ of the parameter. From the knowledge of $P_z(\theta)$, which can be obtained from a calibration measurement, statistical arguments or a complete theoretical description, an unknown future value of θ can be estimated from a single repetition of the experiment with this unknown setting. The Fisher information quantifies the achievable precision of this estimation task via the Cramér-Rao bound. The notation of its derivation in the literature is often somewhat obscured by the aim for the most general expressions in the multivariate case. For our purpose, a single parameter and discrete probability distributions are enough and the derivation can be written down in a very compact form.

We will concentrate here on unbiased estimators and give the biased version in the discussion about the quantum Fisher information. An estimator is called unbiased if its expectation value is the true value of the parameter, that is

$$\langle \hat{\theta}_{\text{est}} \rangle = \theta. \quad (2.60)$$

Specifically, this means that the deviation of the estimator from the true value in a single evaluation is only of statistical nature and averages to zero if the experiment is repeated. This must not be confused with the situation in which more and more statistical samples are used for a single evaluation of the estimator, since then, it should always converge to the true value, if it is a sensible estimator. Consider for example the familiar estimators for mean and variance of a random variable X , which read for m repetitions

$$\bar{X} = \frac{1}{m} \sum_{k=0}^m x_k \quad \text{and} \quad s^2 = \frac{1}{m-1} \sum_{k=0}^m (x_k - \bar{X})^2. \quad (2.61)$$

If X is Gaussian distributed, s^2 is unbiased, whereas the mean squared deviation $(m-1)s^2/m$, which also converges to the variance for $m \rightarrow \infty$ is statistically not centered at the true value for finite m . The bias of s^2 however depends on the form of the probability distribution of X . We use the notation $\Delta X^2 = \text{Var}(X)$ for the variance and $\langle X \rangle$ for the statistical expectation value in the limit of infinite sampling, where X is described by its probability distribution. We will focus on discrete probability distributions, since this is the experimental situation and is more handy to treat. The general continuous case can be found in the extensive literature on the subject, for example [78, 80, 81]. In the discrete case, the only requirements are that $P_z(\theta)$ is differentiable with respect to θ and that the support of $P_z(\theta)$, that is the range for the values of z , where $P_z(\theta)$ is nonzero, does not depend on θ [82].

Cramér-Rao bound For deriving the precision bound of the estimator $\hat{\theta}_{\text{est}}$ we start from the equality

$$0 = \langle \hat{\theta}_{\text{est}} - \theta \rangle = \frac{\partial}{\partial \theta} \sum_z (\hat{\theta}_{\text{est}} - \theta) P_z(\theta) = -1 + \sum_z (\hat{\theta}_{\text{est}} - \theta) \frac{\partial}{\partial \theta} P_z(\theta). \quad (2.62)$$

Under the assumption that the estimator itself is not explicitly dependent on θ , but universal, that is $\partial_\theta \hat{\theta}_{\text{est}} = 0$, we get

$$\sum_z (\hat{\theta}_{\text{est}} - \theta) \frac{\partial}{\partial \theta} P_z(\theta) = \left\langle (\hat{\theta}_{\text{est}} - \theta) \frac{\partial}{\partial \theta} \log P_z(\theta) \right\rangle = 1. \quad (2.63)$$

Taking the square of both sides, the Cauchy-Schwarz inequality for the expectation value $\langle XY \rangle^2 \leq \langle X^2 \rangle \langle Y^2 \rangle$, where $X = (\hat{\theta}_{\text{est}} - \theta)$ and $Y = \partial_\theta \log P_z(\theta)$, yields

$$\langle (\hat{\theta}_{\text{est}} - \theta)^2 \rangle \left\langle \left(\frac{\partial}{\partial \theta} \log P_z(\theta) \right)^2 \right\rangle \geq 1. \quad (2.64)$$

We thus obtain the Cramér-Rao bound

$$(\Delta \hat{\theta}_{\text{est}})^2 \geq \frac{1}{F(\theta)}, \quad (2.65)$$

where $\langle (\hat{\theta}_{\text{est}} - \theta)^2 \rangle$ is the variance $(\Delta \hat{\theta}_{\text{est}})^2$ of the estimator and

$$F(\theta) = \sum_z P_z(\theta) \left(\frac{\partial}{\partial \theta} \log P_z(\theta) \right)^2 \quad (2.66)$$

is the Fisher information. The extension to the case of m independent repetitions of the estimation task leads to the Cramér-Rao bound [80, 81]

$$(\Delta \hat{\theta}_{\text{est}})^2 \geq \frac{1}{mF(\theta)}, \quad (2.67)$$

which is a natural extension obeying the extra scaling with $1/m$ for multiple measurements.

The Cramér-Rao bound expresses the fundamental statistical limit on the precision of unbiased parameter estimation. If it can be attained in a real scenario depends on the details of the implementation and the shape of the probabilities $P_z(\theta)$. Bayesian and maximum likelihood methods (detailed below) attain the limit asymptotically for $m \rightarrow \infty$ [83], which is a relatively weak statement for practical purposes. Therefore it is important for the study of a specific estimation protocol to characterize the required number of repetitions to obtain a certain sensitivity. The proximity to the Cramér-Rao bound reveals the efficiency of the employed estimator.

Characteristics and alternative forms Above, the Fisher information is written in the form of the expectation value of the square (second moment) of the so-called *score* $\partial_\theta \log P_z(\theta)$, which characterizes the sensitivity of $\log P_z(\theta)$ on the parameter. Its expectation value (first moment) yields

$$\sum_z P_z(\theta) \left(\frac{\partial}{\partial \theta} \log P_z(\theta) \right) = \sum_z \frac{\partial}{\partial \theta} P_z(\theta) = \frac{\partial}{\partial \theta} \sum_z P_z(\theta) = \frac{\partial}{\partial \theta} 1 = 0. \quad (2.68)$$

So the Fisher information is identical to the variance of the score. The first derivative of the score reads

$$\frac{\partial^2}{\partial \theta^2} \log P_z(\theta) = \frac{1}{P_z(\theta)} \frac{\partial^2 P_z(\theta)}{\partial \theta^2} - \left(\frac{\partial}{\partial \theta} \log P_z(\theta) \right)^2 \quad (2.69)$$

and its expectation value

$$\sum_z P_z(\theta) \left(\frac{\partial^2}{\partial \theta^2} \log P_z(\theta) \right) = \frac{\partial^2}{\partial \theta^2} \sum_z P_z(\theta) - \sum_z P_z(\theta) \left(\frac{\partial}{\partial \theta} \log P_z(\theta) \right)^2. \quad (2.70)$$

The first term vanishes and the second term is the negative Fisher information, thus we get

$$F(\theta) = - \sum_z P_z(\theta) \left(\frac{\partial^2}{\partial \theta^2} \log P_z(\theta) \right) \quad (2.71)$$

In this representation the Fisher information is interpreted as the negative average curvature of $\log P_z(\theta)$. Since $F(\theta) \geq 0$ this means that the average curvature is negative. The Fisher information can also be written in the convenient forms

$$F(\theta) = \sum_z \frac{1}{P_z(\theta)} \left(\frac{\partial P_z(\theta)}{\partial \theta} \right)^2 = 4 \sum_z \left(\frac{\partial}{\partial \theta} \sqrt{P_z(\theta)} \right)^2. \quad (2.72)$$

Especially the last form is very appealing. We will see later, that the object $\sqrt{P_z(\theta)}$ can be viewed as a coordinate in the Euclidean space of probability amplitudes and the Fisher information is proportional to the square of a statistical speed.

2.3 Maximum likelihood and Bayesian estimation

In this section we will consider two conceptually different but closely related methods of parameter estimation, namely maximum likelihood and Bayesian estimation. Both are based on the likelihood function introduced by Fisher [76], but the historically older Bayesian approach arrives at a different interpretation.

Likelihood function In case we know the probabilities $P_z(\theta)$, and have measured (sampled) m values $\{z_j\}_m = \{z_1, z_2, \dots, z_m\}$ with the fixed setting θ_0 , we can evaluate the likelihood function

$$\mathcal{L}(\theta, \{z_j\}_m) = \prod_{j=1}^m P_{z_j}(\theta) \quad (2.73)$$

It is a function of θ and represents the probability to obtain the outcome sequence $\{z_j\}_m$ if the setting θ_0 had been θ .

Maximum likelihood estimation A natural choice for a statistical estimator of θ_0 is the point at which $\mathcal{L}(\theta)$ obtains its maximum. This is the method of maximum likelihood. Formally stated, the maximum likelihood estimator is

$$(\hat{\theta}_{\text{est}})_{\text{MLE}} = \underset{\theta}{\operatorname{argmax}} \mathcal{L}(\theta, \{z_j\}_m) \quad (2.74)$$

It is often more convenient to work with the logarithm of \mathcal{L} , because probabilities are smaller than one and the likelihood quickly tends to zero. It turns the product into a sum, which reads

$$\log \mathcal{L} = \sum_{j=1}^m \log P_{z_j}(\theta). \quad (2.75)$$

The maximum of $\log \mathcal{L}$ is the same as the maximum of \mathcal{L} , because the logarithm is a monotonically rising function.

Expected curvature of the likelihood function For one observation with the outcome z_1 , the log-likelihood function becomes $\log \mathcal{L}(\theta, z_1) = \log P_{z_1}(\theta)$, which is just the logarithm of the conditional probability evaluated for the fixed value z_1 . It depends on the specific shape of $P_{z_1}(\theta)$ as a function of θ , if it possesses a unique global maximum. If it does, θ_{MLE} is well defined and we know, that $\partial_{\theta} \log \mathcal{L}(\theta, z_1)|_{\theta_{\text{MLE}}} = 0$, which is the necessary condition for a maximum. Furthermore, we know from Eq. 2.71, that

$$\left\langle \frac{\partial^2}{\partial \theta^2} \log \mathcal{L}(\theta, z) \right\rangle = \sum_z P_z(\theta) \frac{\partial^2}{\partial \theta^2} \log P_z(\theta) = -F(\theta). \quad (2.76)$$

This gives a nice intuitive interpretation of the Fisher information: It is the expected negative curvature of the log-likelihood function for a single observation. Evaluated at θ_{MLE} it estimates, how peaked the likelihood function will be on average at the point of the estimation. For m independent observations we get

$$\begin{aligned} \left\langle \frac{\partial^2}{\partial \theta^2} \log \mathcal{L}(\theta, \{z_j\}_m) \right\rangle &= \left\langle \sum_{j=1}^m \frac{\partial^2}{\partial \theta^2} \log P_{z_j}(\theta) \right\rangle = \sum_{j=1}^m \left\langle \frac{\partial^2}{\partial \theta^2} \log P_{z_j}(\theta) \right\rangle \\ &= -mF(\theta). \end{aligned} \quad (2.77)$$

The last two steps reflect, that all z_j are independent and identically distributed. This means, that for m observations, the expected curvature of the likelihood function is m times higher. Our assumption, that the likelihood function has a unique global maximum is not always justified for a limited number of observations. However, the remarkable Bernstein-von Mises theorem of Bayesian statistics (see below) guarantees this with very few assumptions for $m \rightarrow \infty$. Practically speaking, this will be the case for sufficiently many observations.

Bayesian estimation The Bayesian approach differs from the former method mainly in the interpretation of the likelihood function. While the maximum-likelihood estimator only picks the maximum and discards every other information about the shape of the likelihood

function, in Bayesian inference it is regarded as a probability distribution of the parameter θ . This allows to include prior knowledge about the value of θ that is available before the observation. Central to Bayesian estimation is Bayes' theorem, which in our case reads

$$P_{\theta}(\{z_j\}_m) = \frac{\mathcal{L}(\theta, \{z_j\}_m)P(\theta)}{P(\{z_j\}_m)}. \quad (2.78)$$

It assigns a conditional probability $P_{\theta}(\{z_j\}_m) = P(\theta|\{z_j\}_m)$ to the parameter θ based on the observation $\{z_j\}_m$ on which the likelihood function is evaluated, and the prior knowledge $P(\theta)$ which is also expressed as a probability. The denominator merely provides the proper normalization for P_{θ} to be interpretable as probability distribution. If θ is a continuous variable, it can be written as

$$P(\{z_j\}_m) = \int \mathcal{L}(\theta, \{z_j\}_m)P(\theta) d\theta. \quad (2.79)$$

For estimation purposes this value is irrelevant, because it does not depend on θ . If $\log P_{\theta}$ is examined, it only gives a constant offset. The object $P_{\theta}(\{z_j\}_m)$ is called the posterior distribution. In words, Bayes' theorem states: *Posterior probability is proportional to likelihood times prior probability.*

We note that in a puristic view, one would not state Bayes' theorem with the likelihood function, but with exactly the same object named and interpreted differently, which is not very helpful here since we already gained some intuition about the likelihood function.

In contrast to the maximum likelihood approach, the outcome is not just one value but the whole function P_{θ} is interpreted as a probability distribution of θ , which can be evaluated in different ways. For example, the estimator could be the mean θ evaluated with the probability $P_{\theta}(\{z_j\}_m)$, which minimizes the mean squared deviation of the estimator from the true value. Very similarly to the maximum likelihood method, the maximum of the posterior distribution can be used as estimator, which is the maximum-a-posteriori (MAP) estimator.

A big advantage of the Bayesian approach is that confidence regions can be defined (which for example contain 68% of the probability, corresponding to one standard deviation). However, one more or less arbitrary part is the choice of the prior probability $P(\theta)$. It can contain anything from previous observations to personal belief. If no evidence about θ is available before the observation, it should be simply taken constant to keep subjective elements out of the estimation (principle of indifference). After one observation, the obtained posterior can be used as prior for the next observation, which is called Bayesian updating.

One important result in Bayesian statistics is the Bernstein-von Mises theorem [84], which states that independent on the choice of the prior, the posterior distribution converges to a normal distribution, which is centered on the true value θ_0 in the limit $m \rightarrow \infty$. The prior is only required to be nonzero in the neighborhood of θ_0 , which is perfectly reasonable. A prior, which assigns zero probability to the true value is definitely wrong. The only further requirement for our discrete case is, that the space of possible values of z is finite, which can be easily approximated in practical applications and is certainly true in our case, since the number of atoms is finite. The rate of the convergence again depends on the choice of the prior and the shape of the probability distributions $P_z(\theta)$. This theorem is quite remarkable, since it

has the character of a central limit theorem for the Bayesian probability distribution. With this knowledge, we can give the limiting form of the posterior probability as

$$\log P_\theta(\{z_j\}_m) \rightarrow \log \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(\theta - \theta_0)^2}{2\sigma^2} \right] \right) = -\frac{(\theta - \theta_0)^2}{2\sigma^2} + \text{const.} \quad (2.80)$$

independent of the prior. For $P(\theta) = 1$, the Bayesian probability is identical to the (normalized) likelihood function, thus we can identify the Gaussian variance σ^2 with the inverse curvature of the log-likelihood function and get with Eq. 2.77

$$\sigma^2 = \frac{1}{mF(\theta_0)}. \quad (2.81)$$

This means that Bayesian estimation with mean and variance of $P_\theta(\{z_j\}_m)$ is asymptotically unbiased and saturates the Cramér-Rao bound Eq. 2.67. In a practical application, the convergence to a Gaussian for increasing m is an important indicator for the efficiency of the estimator.

2.4 Fisher information extraction

One route for characterization of the experimentally obtained quantum state is full state tomography to obtain an experimental estimate of the density matrix. The problem with this general approach is that it becomes exponentially hard with the number of qubits [85, 86]. Besides this principal difficulty, we do not have access to the individual qubits, which would force to make further a-priori assumptions about the quantum state, for example, that the collective two-mode approximation is valid. Under this assumption, the dimension of the density matrix is $(N + 1)^2$, which still scales quadratically. In our case, this would require the precise estimation of about $(N + 1)^2/2 \sim 10^5$ free parameters, which is not feasible at the moment.

We will pursue a different route here, which employs the observed distributions directly for state characterization. We have seen that the Fisher information quantifies the amount of knowledge about a parameter that is encoded in the probability distributions of the observable and their infinitesimal change with respect to the parameter. In principle, both can be extracted by sampling of the distributions for known settings of the parameter which are closeby. Experimentally, we face the problem that the observed frequencies fluctuate, which is especially problematic for the estimation of the derivative. Therefore we need a statistically more robust way of extracting these quantities. We will discuss the concept of statistical distance here and show how this concept can be used to estimate the Fisher information from experimental probability distributions.

2.4.1 The concept of statistical distance

The notion of statistical distance is about as old as modern statistical sampling theory [76], see also [87]. The main idea is to quantify in statistical terms, how close a given probability distribution is to another, that is how distinguishable they are from each other for a given number

of samples. Quite naturally, this will depend on how structured the compared distributions are, that is how probable it is to observe a sample, which can easily be identified as belonging to one of the two alternatives. The definition is not unique and quite a number of statistical distance measures have been devised, which are closely related [88]. We will concentrate on one specific choice, which is very much related to the classical statistical concepts on which Wootters studied the geometric structure of the quantum mechanical Hilbert space [87].

The basic idea is to count how many in m trials mutually distinguishable probability distributions fit on a smooth curve with minimal length joining the two. In this approach, distinguishability is defined with respect to the size of the uncertainty region for m trials. We shortly reproduce the concepts in a notation consistent with the further treatment.

The probability distributions P_z are thought of as points in a space with as many dimensions as there are possible values of z . We will call the normalized frequencies or experimental probability distributions \mathcal{F}_z with $\sum_z \mathcal{F}_z = 1$. When sampling m times from P_z , the corresponding likelihood (or probability mass function) for the \mathcal{F}_z will be the multinomial distribution

$$\mathcal{L}(\{\mathcal{F}_z\}) \propto \prod_z (P_z)^{m\mathcal{F}_z}, \quad (2.82)$$

where $m\mathcal{F}_z$ is the number of occurrence of the specific value z and the product runs over all possible values of z . We omit the normalization constants for the sake of clarity, whenever possible. Due to the central limit theorem, the likelihood becomes Gaussian for $m \rightarrow \infty$ and reads

$$\mathcal{L}(\{\mathcal{F}_z\}) \rightarrow \propto \exp \left[-\frac{m}{2} \sum_z \frac{(\mathcal{F}_z - P_z)^2}{P_z} \right]. \quad (2.83)$$

The region of uncertainty is now defined as the 1σ radius of this Gaussian, which leads to the condition

$$\frac{\sqrt{m}}{2} \left[\sum_z \frac{(\delta P_z)^2}{P_z} \right]^{\frac{1}{2}} > 1 \quad (2.84)$$

for the distinguishability of the likelihood function of m trials for the distributions P_z and P'_z with $\delta P_z = P_z - P'_z$. Already at this point we can see the relationship with the Fisher information, if we insert $\delta P_z = \partial_\theta P_z(\theta) \delta\theta$ for a parametrized $P_z(\theta)$, which leads to

$$\frac{\sqrt{m}}{2} \left[\sum_z \frac{1}{P_z(\theta)} \left(\frac{\partial P_z(\theta)}{\partial \theta} \right)^2 \right]^{\frac{1}{2}} \delta\theta = \frac{\sqrt{mF(\theta)}}{2} \delta\theta > 1 \quad (2.85)$$

and thus to the condition $\delta\theta > 2/\sqrt{mF}$ for distinguishability, which is two times the Cramér-Rao bound. The definition of the line element proceeds in removing the expected scaling with \sqrt{m} and taking $m \rightarrow \infty$, which ensures the convergence to Gaussian, while keeping the size of the 'distinguishability region' fixed. The statistical length of a curve, which joins $P_z = P_z(0)$ and $P'_z = P_z(\alpha)$ can now be obtained by the line integral (see also Eq. 2.72)

$$l = \frac{1}{2} \int_0^\alpha dt \sqrt{F(t)} = \int_0^\alpha dt \left[\sum_z \left(\frac{\partial}{\partial t} \sqrt{P_z(t)} \right)^2 \right]^{\frac{1}{2}} = \int_0^\alpha dt \left[\sum_z \left(\frac{dx_z}{dt} \right)^2 \right]^{\frac{1}{2}} \quad (2.86)$$

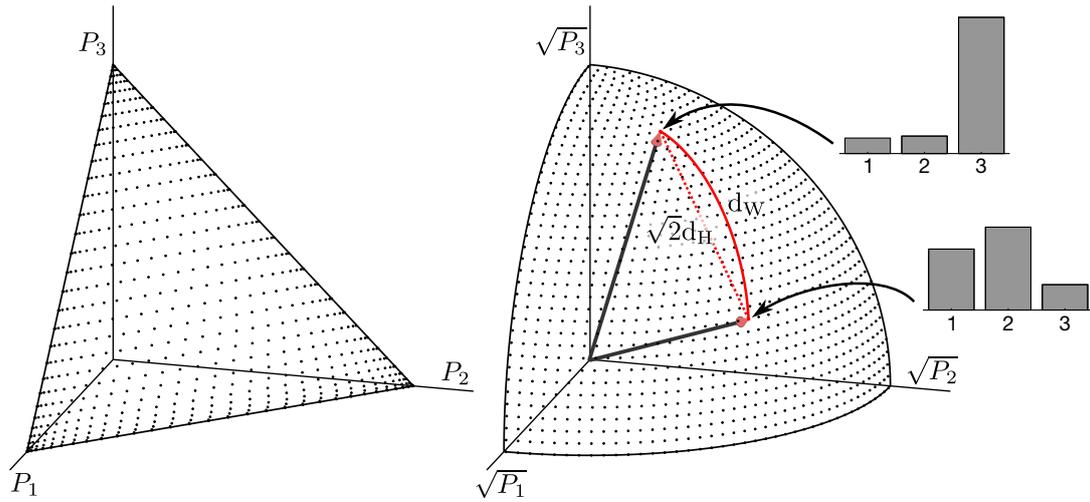


Figure 2.6 Geometrical picture of probability distributions and statistical distance. The probability space for three possible outcomes is a plane in the positive octant, defined by the normalization $P_1 + P_2 + P_3 = 1$. The dots represent equally spaced distributions in the sense of statistical distance. Distributions close to a border can be more easily distinguished, which can be observed as accumulation of the points. A change of coordinates via the square root rectifies this and leads to equidistant points on the unit sphere. The arrows denote two exemplary distributions, which are also shown as histograms. The definition of the Wooters distance d_W is the arc length between them, whereas the Hellinger distance d_H is proportional to the chordal distance, which is the length of the straight connection.

Here, the meaning of $F(t)$ is the Fisher information calculated along the parametrized curve $P_z(t)$. The change of coordinates $x_z = \sqrt{P_z}$ allows to identify the last integral as an Euclidean length. The normalization condition of probability distributions reads $\sum_z x_z^2 = 1$ and identifies the space of possible coordinates x_z as the surface of the unit sphere constrained to the coordinate region, where $x_z > 0$. To define the distance, one can choose the smallest possible length l of a curve on the unit sphere, which is identical to the angle between the probability amplitudes x_z and x'_z and is the definition of the Wooters distance:

$$d_W(P, P') = \arccos(\vec{x} \cdot \vec{x}') = \arccos \sum_z \sqrt{P_z P'_z} \quad (2.87)$$

2.4.2 Hellinger distance

We have chosen to work with another definition of statistical distance, which is up to a constant factor the "normal" Euclidean distance between the probability amplitudes and is called the Hellinger distance. We will mainly work with its square, which reads

$$d_H^2(P, P') = 1 - \vec{x} \cdot \vec{x}' = 1 - \sum_z \sqrt{P_z P'_z} = \frac{1}{2} \sum_z \left(\sqrt{P_z} - \sqrt{P'_z} \right)^2 \quad (2.88)$$

Geometrically, this means choosing the length of the direct connection, which is also called chordal distance, instead of the arc length on the sphere. This is illustrated in Fig. 2.6. It is quite remarkable that the intuitive argument for distinguishability based on the multinomial

distribution ultimately leads to the square root of probabilities as the underlying geometrical object [87] already for classical probabilities³. The Fisher information can be identified as a differential Riemannian metric on probability space [88]. This causes curves which pass regions of high Fisher information to lengthen. From the structure depicted in Fig. 2.6 we directly see that the regions with high Fisher information are located close to the border of the probability space, corresponding to fine structured probability distributions.

We will make this more concrete by investigating the structure of the Hellinger distance for our singly parametrized $P_z(\theta)$. The Hellinger distance is itself a metric, that means that in the probability space it is positive or zero, satisfies the triangle inequality and is symmetric. It is zero only, if the two probability distributions are the same [88] and always smaller or equal to one. It is equal to one, if the probabilities are mutually singular, which is equivalent to orthogonality in probability amplitude space, that is $\vec{x} \cdot \vec{x}' = 0$. In this situation, P_z is zero, where P'_z is not and vice versa.

We will now have a look at the Hellinger distance for neighboring points along our parametrized curve $P_z(\theta)$ with the help of a Taylor expansion at $\theta = 0$. This is no restriction, since we can always reparametrize $\theta' = (\theta - \theta_0)$ for every θ_0 along the curve. For the first local derivatives, we get

$$\begin{aligned}
d_{\text{H}}^2 \Big|_0 &= 0 \\
\frac{\partial d_{\text{H}}^2}{\partial \theta} \Big|_0 &= -\frac{1}{2} \sum_z \frac{\partial P_z(\theta)}{\partial \theta} \Big|_0 = \frac{1}{2} \frac{\partial}{\partial \theta} \sum_z P_z(\theta) \Big|_0 = 0 \\
\frac{\partial^2 d_{\text{H}}^2}{\partial \theta^2} \Big|_0 &= \frac{1}{4} \sum_z \left[\frac{1}{P_z(0)} \frac{\partial P_z(\theta)}{\partial \theta} \Big|_0 - 2 \frac{\partial^2 P_z(\theta)}{\partial \theta^2} \Big|_0 \right] = \frac{F(0)}{4} - \underbrace{\frac{1}{2} \frac{\partial^2}{\partial \theta^2} \sum_z P_z(\theta) \Big|_0}_{=0} \\
\frac{\partial^3 d_{\text{H}}^2}{\partial \theta^3} \Big|_0 &= -\frac{3}{8} \frac{\partial F(\theta)}{\partial \theta} \Big|_0
\end{aligned} \tag{2.89}$$

Thus, we obtain the Taylor expansion of the squared Hellinger distance as

$$d_{\text{H}}^2(P^{(\theta)}, P^{(\theta+\delta\theta)}) = \frac{F(\theta)}{8} \delta\theta^2 - \frac{F'(\theta)}{16} \delta\theta^3 + \mathcal{O}(\delta\theta^4) \tag{2.90}$$

We see that the local curvature of the squared Hellinger distance is proportional to the Fisher information, which is always non-negative. In contrast, $F'(\theta)$ and higher order terms can also be negative.

2.4.3 Extraction procedure

The connection of the Fisher information to the curvature of the squared Hellinger distance provides the basis of our extraction protocol. The idea is to experimentally sample several $P_z(\theta)$ for closely spaced parameters θ , which are characteristic for the produced quantum state and its dependence on θ . In our case θ will be the angle of a linear rotation. We then calculate

³Quantum mechanics is based on probability amplitudes, whose absolute values have the 'dimension' square-root-probability, i.e. $|\Psi(\alpha)| = \sqrt{P(\alpha)}$

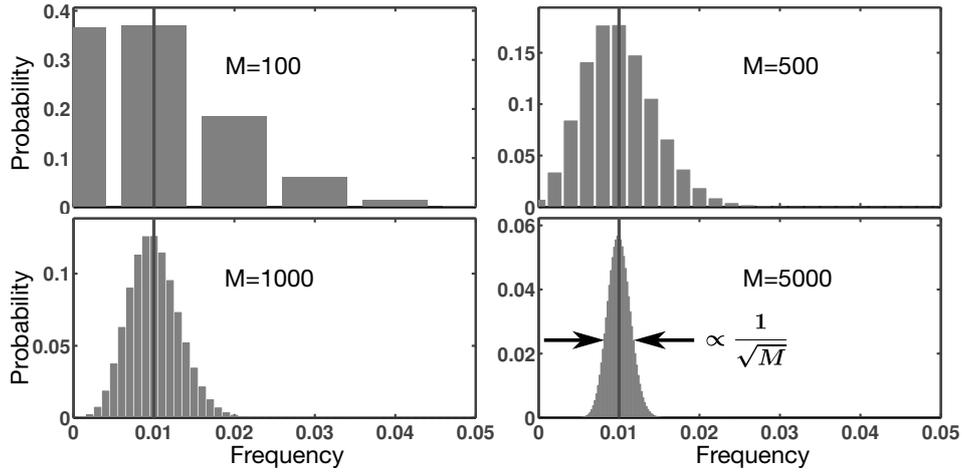


Figure 2.7 Illustration of the probability distribution of \mathcal{F}_z for the specific value $p = \langle \mathcal{F}_z \rangle = 0.01$ at a single value of z for varied number of samples M . The expected distribution of \mathcal{F}_z is multinomial, which results in a binomial distribution $\binom{M}{k} p^k (1-p)^{M-k}$ with $k = M\mathcal{F}_z$ for the single z .

the squared Hellinger distance between the experimental probability distributions and extract the curvature and thus the Fisher information by a quadratic fit. Since we have to work with a finite number of experimental repetitions we consider in the following the statistical properties of this protocol.

2.4.4 Hellinger distance of experimental probability distributions

For a sufficiently high number of experimental observations, $P_z(\theta)$ will be more and more approximated by the observed frequencies $\mathcal{F}_z(\theta)$ ⁴. We therefore consider the extension of the Hellinger distance to frequencies, which reads

$$d_{\text{H}}^2(\mathcal{F}^{(0)}, \mathcal{F}^{(\theta)}) = 1 - \sum_z \sqrt{\mathcal{F}_z(0) \mathcal{F}_z(\theta)}. \quad (2.91)$$

For a finite number of samples, $\mathcal{F}_z(\theta)$ also has a finite uncertainty, which shows up as fluctuation around $P_z(\theta)$, if the sampling is repeated. This is described by the multinomial distribution (Eq. 2.82). If $\mathcal{F}_z(\theta)$ has been sampled very well, the deviations from $P_z(\theta)$ are to a good approximation Gaussian distributed and well characterized by their variance, as illustrated in Fig. 2.7. In this case, standard error propagation can be used for the error estimation of d_{H}^2 , which is not the case any more for poorer sampling conditions, especially with distributions close to the boundary of the probability space, i.e. small probabilities for some z . For small probabilities P_z it takes a very high number of repetitions for \mathcal{F}_z to become Gaussian and statistically centered at P_z because unlikely events naturally happen infrequently. However, the very concept of statistical distance (distinguishability in m trials) ensures, that for each value of the Fisher information there will be a suitable $\delta\theta$ such that the statistical distance for a specific number of repetitions will not be dominated by these effects. In other words: The distance of distributions, which contain a big amount of information about θ will converge fast.

⁴We use the term 'experimental probability' equivalently

Gaussian error propagation For a large number of repetitions, we can separate the signal from the fluctuations and write

$$\mathcal{F}_z(\theta) = P_z(\theta) + \delta f_z(\theta) \quad (2.92)$$

with $\delta f_z(\theta) \rightarrow 0$ for many repetitions. The deviations $\delta f_z(\theta)$ are the multinomial sampling errors characterized by

$$\langle \delta f_z(\theta) \rangle = 0 \quad (2.93)$$

$$\langle \delta f_z(\theta)^2 \rangle = \frac{P_z(\theta)(1 - P_z(\theta))}{M} \quad (2.94)$$

$$\langle \delta f_z(\theta) \delta f_{z'}(\theta) \rangle = -\frac{P_z(\theta)P_{z'}(\theta)}{M} \quad \text{for } z \neq z' \quad (2.95)$$

Here, M is the number of samples and $\langle \cdot \rangle$ denotes the statistical expectation value for repetitions of the whole sampling procedure. The second line is equal to the variance and the third to the covariance, which is nonzero, because observing the value z in one sample excludes the observation of all other possibilities $z' \neq z$. Because of the normalization, the fluctuations sum to zero, that is $\sum_z \delta f_z(\theta) = 0$ for one realization. For small δf , we make a Taylor expansion of Eq 2.91 up to second order, which yields

$$d_{\text{H}}^2 \approx 1 - \sum_z \sqrt{p_0 p_\theta} \left(1 + \frac{\delta f_0}{2p_0} + \frac{\delta f_\theta}{2p_\theta} - \frac{\delta f_0^2}{8p_0^2} - \frac{\delta f_\theta^2}{8p_\theta^2} + \frac{\delta f_0 \delta f_\theta}{4p_0 p_\theta} \right) \quad (2.96)$$

where we have used the abbreviations $p_\theta = P_z(\theta)$ etc. This is justified if $\delta f \ll P_z(\theta)$. Since $\langle \delta f_z(\theta) \rangle = 0$, this leads to the requirement $\sqrt{\langle \delta f_z(\theta)^2 \rangle} \ll P_z(\theta)$ and thus with Eq. 2.94 to $P_z(\theta) \gg 1/M$. The next step is to take the sample average of this expression. We first note, that $\langle \delta f_0 \delta f_\theta \rangle = 0$ because the two are statistically independent and $\langle \delta f_0 \rangle = \langle \delta f_\theta \rangle = 0$ because of Eq. 2.93. By inserting the multinomial variances (Eq. 2.94), we arrive at

$$\langle d_{\text{H}}^2 \rangle \approx 1 - \left(1 + \frac{1}{8M_0} + \frac{1}{8M_\theta} \right) \sum_z \sqrt{p_0 p_\theta} + \frac{1}{8} \sum_{\tilde{z}} \left(\frac{1}{M_0} \sqrt{\frac{p_\theta}{p_0}} + \frac{1}{M_\theta} \sqrt{\frac{p_0}{p_\theta}} \right) \quad (2.97)$$

In general, the number of samples for $\mathcal{F}^{(0)}$ and $\mathcal{F}^{(\theta)}$ can be different, which we have denoted by M_0 and M_θ respectively. The second sum (\tilde{z}) runs over all values of z , where p_0 and p_θ are both nonzero.

Statistical bias Comparing Eq. 2.97 to the definition (Eq. 2.88) we can observe, that the expectation value of the squared Hellinger distance obtains a statistical offset from the true value due to finite sampling even with the assumption of symmetric (Gaussian) fluctuations of each frequency around its mean. The reason for this is the nonlinearity of the square root in the Hellinger distance and that $\langle f(X) \rangle \neq f(\langle X \rangle)$ for nonlinear functions f . Specifically, if f is convex⁵, then $\langle f(X) \rangle \geq f(\langle X \rangle)$ and respectively " \leq " for concave, which is called Jensen's inequality.

⁵For a convex function $f(a_1 x_1 + a_2 x_2) \leq a_1 f(x_1) + a_2 f(x_2)$ for $a_1 + a_2 = 1$ and both positive

To illustrate, how a statistical bias arises for nonlinear functions, we look at the simplified one-dimensional example $f(x) = \sqrt{x}$ (concave) with Gaussian distributed input described by a random variable $X = x_0 + \delta X$ with $\langle X \rangle = x_0$ and variance $\langle \delta X^2 \rangle = \sigma^2$. The Taylor expansion yields

$$f(x + \delta x) = \sqrt{x} \left(1 + \frac{\delta x}{2x} - \frac{\delta x^2}{8x^2} + \mathcal{O}(\delta x^3) \right). \quad (2.98)$$

It contains all relevant terms of the discussion above. With $\langle \delta X \rangle = \langle \delta X^3 \rangle = 0$ ⁶ the expectation value of this expression reads

$$\langle f(X) \rangle = \sqrt{x_0} \left(1 - \frac{\langle \delta X^2 \rangle}{8x_0^2} + \mathcal{O}(\langle \delta X^4 \rangle) \right) \approx \sqrt{x_0} \left(1 - \frac{\sigma^2}{8x_0^2} \right) \quad (2.99)$$

in contrast to the value $f(\langle X \rangle) = \sqrt{x_0}$. Only if $f(x)$ is linear, the quadratic and higher order terms do not arise and $\langle f_{\text{lin}}(X) \rangle = f_{\text{lin}}(\langle X \rangle)$.

To obtain a similar expression as Eq. 2.90, we now expand Eq. 2.97 again up to second order in θ , which yields

$$\langle d_{\text{H}}^2(\mathcal{F}^{(\theta)}, \mathcal{F}^{(\theta+\delta\theta)}) \rangle = c_0 + \left(c_2 + \frac{F(\theta)}{8} \right) \delta\theta^2 + \mathcal{O}(\delta\theta^3, \delta f_z^3), \quad (2.100)$$

with

$$c_0 = \frac{n-1}{4M}$$

$$c_2 = \frac{1}{32M} \left[1 + \sum_z \left(\frac{\partial_\theta P_z(\theta)}{P_z(\theta)} \right)^2 \right] \quad (2.101)$$

where n is the number of z -values with $P_z(\theta)P_z(\theta + \delta\theta) \neq 0$ (both nonzero) which is approximately the same number for the condition $\mathcal{F}_z(\theta)\mathcal{F}_z(\theta + \delta\theta) \neq 0$. In order to obtain consistent results with this expansion, we have set $M = M_\theta = M_0$ here. Otherwise, the evaluation of the derivatives would assume that $\mathcal{F}^{(\theta)} \neq \mathcal{F}^{(\theta+\delta\theta)}|_{\delta\theta=0}$, which is contradictory. A similar treatment for the sample variance of d_{H}^2 gives

$$\text{Var}(d_{\text{H}}^2) = \frac{F(\theta)}{8M} \delta\theta^2 + \mathcal{O}(\delta\theta^3, \delta f_z^3) \approx \frac{d_{\text{H}}^2}{M} \quad (2.102)$$

The above discussion shows that an extraction of the Fisher information using the statistical Hellinger distance for experimentally obtained frequency distributions \mathcal{F}_z is possible in principle and all errors decrease asymptotically with $\sim 1/M$. However, the underlying probabilities $P_z(\theta)$ of an observed experimental state are unknown, which would be needed for the evaluation of Eq. 2.101 and 2.102. The observed \mathcal{F}_z are the best available guesses for P_z . Close to the border of the probability space, that is for small P_z , assigning some errors like $\Delta(\mathcal{F}_z)^2 \sim \mathcal{F}_z(1 - \mathcal{F}_z)/M$ would lead to false conclusions, because \mathcal{F}_z can be far from Gaussian distributed⁷. In other words, the prerequisite $M \gg 1/P_z(\theta)$ for the above treatment cannot be fulfilled for all z . Therefore an independent statistical method for estimation of bias and error bounds is desirable. For this, we will employ a Jackknife resampling technique, detailed in the following.

⁶All odd central moments are zero for a Gaussian distributed random variable

⁷We will be dealing with experimental frequency distributions, which contain as few as one observation for some particular z

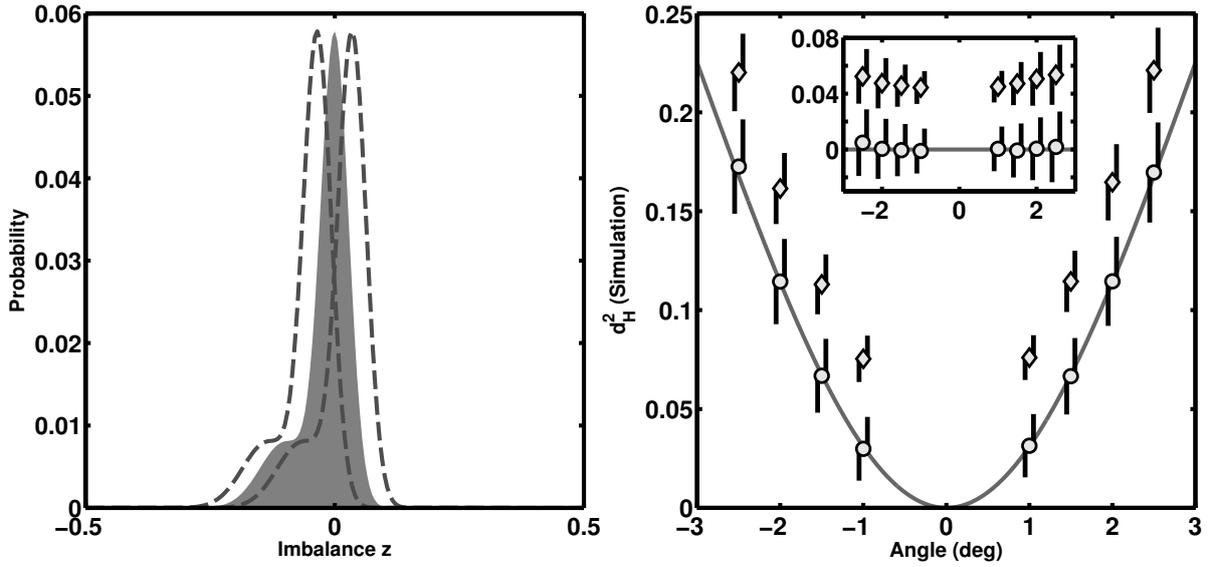


Figure 2.8 Left: Fictitious probability distribution of z used for benchmarking the employed statistical methods. It resembles the admixture of a 5 dB squeezed state with 25% of a coherent state for 400 atoms, shifted by 0.1 with respect to each other. The resulting variance in z is larger than those of a coherent state for this choice. For the Monte-Carlo simulation of the statistical Hellinger distance versus angle θ , the whole distribution was shifted by the amount $\delta z = \delta\theta$, which resembles rotations on the Bloch sphere with maximal spin length. The dashed lines represent shifts of $\delta\theta = \pm 2^\circ$. **Right:** Sampled Hellinger distances (here for the example $M_\theta = 300$ repetitions for $\theta \neq 0$ and $M_0 = 1200$ repetitions for $\theta = 0$). The line is the result of the perfect probability distribution without sampling (true value). The sample mean of d_H^2 (diamonds) obtains a statistical bias, which can be reliably estimated by the Jackknife method and subtracted (circles). Error bars display the statistical fluctuation (1 s.d.) for the uncorrected and the mean of the Jackknife error estimate for the bias corrected case. The inset shows the residuals of the two upon subtraction of the true value.

2.4.5 Jackknife resampling and Monte-Carlo simulation

In this section, we will describe a robust method for estimation of statistical errors and bias for the Hellinger distances of the experimentally observed frequency distributions \mathcal{F}_z . Furthermore we will substantiate the Fisher information extraction method with Monte-Carlo simulations on realistic probability distributions, which allows to benchmark the method for different choices of the sample size M .

With access to an unlimited number of experimental repetitions under reproducible conditions, the whole method could be repeated many times, which would allow to study its statistical fluctuations as well as to predict its statistical bias from the averaging behavior. Of course, the question if the method has a systematic error, i.e. averages to the wrong result for infinite data, cannot be answered by purely statistical means. In our case, we want to first estimate probabilities by observed frequency distributions, which by itself is very demanding for our typical system sizes of $N \sim 400$ atoms, which translates in principle to $N + 1$ possible values for the imbalance z when assuming perfect detection. As discussed above, the further procedure of calculating the Hellinger distances cannot be treated reliably by Gaussian error propagation, which would need the true probabilities and their derivatives as input. The re-

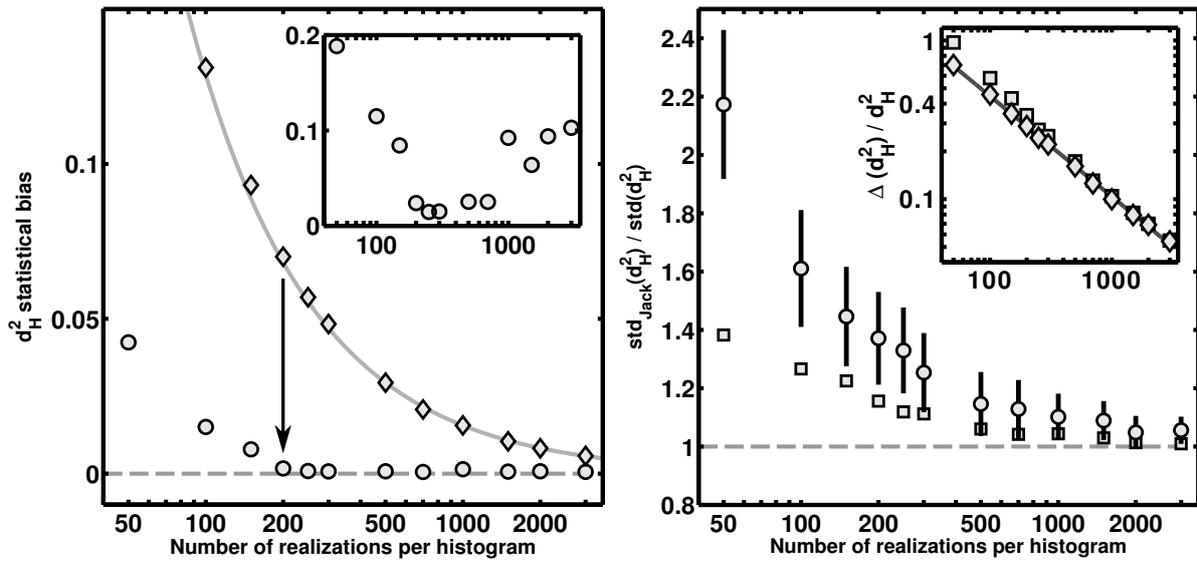


Figure 2.9 Systematic Monte-Carlo simulations for different number of realizations M_θ per histogram of the Hellinger distance estimation, sampled from the probability distribution of Fig. 2.10 with increased statistics $M_0 = 4M_\theta$ for the reference histogram. **Left:** Bias removal by the Jackknife method. The line is a fit of a power law to the mean statistical bias of d_H^2 (diamonds) with the result $\propto 1/M_\theta^{0.92}$. Circles show the average of the corrected values and the inset the division of the two, which demonstrates an efficient and accurate bias reduction of more than 95% in the relevant range (not too few samples but with an appreciable bias). **Right:** The inset shows the scaling of the fluctuations of the biased Hellinger distance estimation (diamonds) and those of the bias corrected values (squares) with the number of realizations. The line is a power law with $1/M_\theta^{0.63}$. The main panel shows the fluctuation increase due to the bias removal (squares), which is remarkably low ($< 10\%$ in the relevant range) and the Jackknife error estimation (circles) with its standard error. The Jackknife method overestimates the true error and lies reliably above the slightly increased fluctuations caused by the bias removal.

maintaining possibility of repeating many times is experimentally difficult to realize due to the enormous amount of necessary experimental repetitions.

Resampling methods provide in some cases an efficient way out of this dilemma. They are well established tools in the field of statistical estimation and decision [89] and allow to infer error and bias margins basically by evaluating the original estimator on generated pseudosamples. These are obtained by random draws with replacement from the original data (nonparametric Bootstrap [90]) or a guess for their distribution (parametric Bootstrap) or generated by taking all but a small subset of data (Jackknife). While Bootstrap methods are more versatile and better suited for the assessment of the whole distribution of an estimator, we chose to work with the Jackknife procedure, which was the precursor of Bootstrap and originally designed for bias reduction. The advantages are that it includes only real experimentally observed data, can be made perfectly reproducible (does not require a random number generator), is easy to use and computationally fast [91]. It is able to remove a statistical bias scaling with $1/M$, which we have seen is the case for the Gaussian approximation (Eq. 2.101) in the limit of big M . We now turn to the exact implementation later used for the experimental data and present Monte-Carlo simulations to test this procedure.

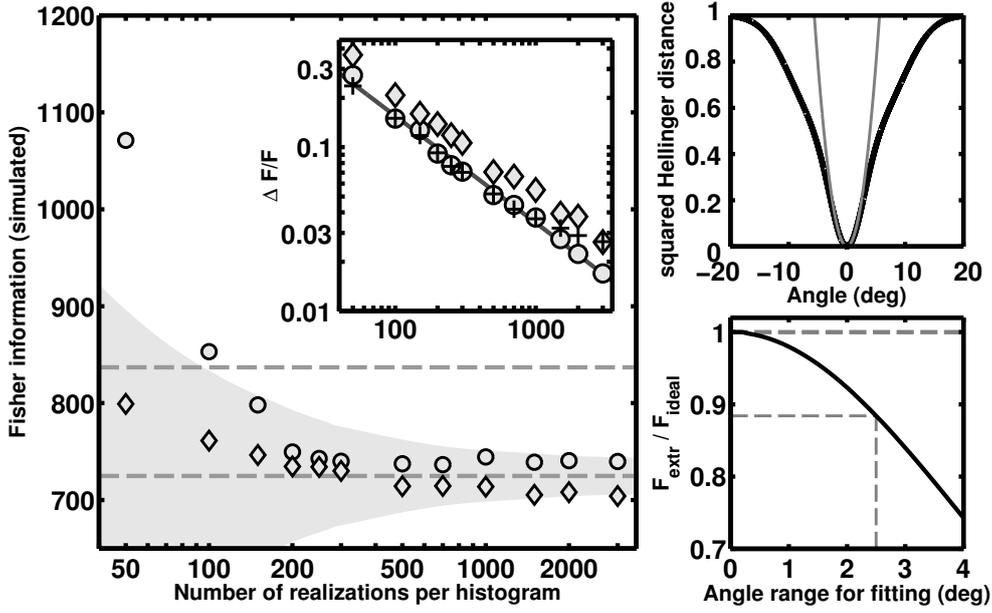


Figure 2.10 Left: Monte-Carlo simulation of the extraction of the Fisher information F from quadratic fits to the estimated Hellinger distances based on the probability distributions of Fig. 2.8. The bias corrected values (circles) were fitted quadratically with the curvature as only fit parameter, whereas for the original biased values (diamonds) a global offset was allowed as second fit parameter. The lower dashed line displays the result of a quadratic fit to the true values of the Hellinger distance in the same angle range ($\pm 2.5^\circ$). The shaded region is the 68% confidence interval of the bias corrected fits, centered on the dashed line as reference. The inset shows the scaling of the relative standard error of the extracted values of F . The statistical fluctuations of the estimate of F for the bias corrected case (circles) closely match the average 68% confidence prediction of the single fits (crosses). The fluctuations of the biased version (diamonds) are bigger due to the additional fit parameter. The deviation from the true infinitesimal Fisher information (upper dashed line) stems from the finite angle range for the extraction, since d_H^2 starts to deviate from parabolic for larger angles. **Right, top:** Hellinger distance values of the probability distributions used for the Monte-Carlo simulation for larger shift angles θ . The curvature of the gray parabola is the ideal Fisher result for infinitesimal changes. The angular shape of the Hellinger distance is specific for the underlying distributions. The shallow shape for larger angles is generic for peaked distributions. **Right, bottom:** Dependence of the extracted Fisher result on the angle range used for the Hellinger fit. For the exemplary case of $\pm 2.5^\circ$ used for the left panel (dashed lines), this leads to an underestimation of $\sim 12\%$.

After data taking, we have M_0 experimental values $\{z_1^{(0)}, z_2^{(0)} \dots z_{M_0}^{(0)}\}$ of the imbalance z for the reference at $\theta = 0$ and M_θ values $\{z_1^{(\theta)}, z_2^{(\theta)} \dots z_{M_\theta}^{(\theta)}\}$ for small $\theta \neq 0$ which we bin and normalize to obtain the discrete frequency distributions $\mathcal{F}_z(0)$ and $\mathcal{F}_z(\theta)$. With these, we can calculate one value for the Hellinger distance at each θ .

For the Jackknife bias and error evaluation, we divide $\{z_1^{(0)}, z_2^{(0)} \dots z_{M_0}^{(0)}; z_1^{(\theta)}, z_2^{(\theta)} \dots z_{M_\theta}^{(\theta)}\}$ in g blocks of h samples ($M_0 + M_\theta = hg$). With this, g pseudovalues $[d_H^2]_{\text{rem}j}$ are obtained by

removing block number j and calculating the Hellinger distance with the remaining samples. The Jackknife estimator

$$\langle d_{\text{H}}^2 \rangle_{\text{Jack}} = g [d_{\text{H}}^2]_{\text{Full}} - (g-1) \frac{1}{g} \sum_{i=1}^g [d_{\text{H}}^2]_{\text{rem}j} \quad (2.103)$$

is the bias corrected value used for the further analysis. Its standard error [91] is estimated as the square root of

$$(\Delta(d_{\text{H}}^2)_{\text{Jack}})^2 = (g-1) \text{JackVar}([d_{\text{H}}^2]_{\text{rem}j}) = \sum_{j=1}^g \left([d_{\text{H}}^2]_{\text{rem}j} - \langle d_{\text{H}}^2 \rangle_{\text{Jack}} \right)^2 \quad (2.104)$$

The factor $(g-1)$ accounts for the fact, that the generated samples are not statistically independent. This block-Jackknife results are averaged up to blocksize $h = 20$ for mean and variance, which provides extra robustness. Before each block-Jackknife evaluation, the samples of both reference and angle measurement are randomly shuffled.

In Figures 2.8 and 2.9 we show a test of this procedure on fictitious probability distributions, which resemble those of a quantum mechanical state consisting of a 5 dB squeezed state with a 25% admixture of a coherent spin state for $N = 400$ atoms. The noise and the relative displacement of the coherent state annihilates all quantum metrological advantage for standard averaging readout of a perfect Ramsey sequence with maximal mean spin length. However, the ideal Fisher information is still ~ 3 dB beyond the shotnoise limit. Fig. 2.8 shows this distribution together with an example of a Hellinger distance estimation for rotations with maximal spin length and slope ($\delta z = \delta\theta$). The knowledge of the probability distributions from which the samples are generated allows to test the statistical estimation. We find the expected positive bias, resulting from the concavity of the Hellinger distance in the probabilities as discussed above. The Jackknife method reduces it very effectively and produces realistic errorbars. Fig. 2.9 shows this systematically for different sample sizes M_θ . For each choice, the reference is sampled with $M_0 = 4M_\theta$, close to our experimental implementation and motivated by the fact that the reference is contained in d_{H}^2 for every angle θ and it is therefore beneficial to prioritize it in situations with a limited total number of samples. The procedure is repeated 200 times for each M_θ to enable a statistical analysis of the Jackknife estimates and a comparison to their true value. The statistical bias of d_{H}^2 scales with $\sim 1/M$ also under very poor sampling conditions and is reliably removed by the Jackknife. In the relevant range of $M > 300$, the bias reduction adds less than 10% noise to the estimated d_{H}^2 , well within the Jackknife error estimate, which is slightly larger than the true statistical fluctuations⁸. Finally, Fig. 2.10 shows the systematic behavior of the curvature extraction via quadratic fits with $f(\theta) = a\theta^2$ or $f(\theta) = a\theta^2 + b$ for the bias corrected and the original values of d_{H}^2 respectively. Both are translated to an estimated Fisher information $F = 8a$. The fits are of least- χ^2 type weighted with the Jackknife error estimates or nonweighted respectively. In the relevant range, the results are very close to the ideal angle-averaged value, which is obtained by a quadratic fit to the true Hellinger distances in the same angle range. The confidence region bound of the fit is consistent with the statistical error of the curvature which is bigger for the uncorrected values due to the additional fit parameter. The d_{H}^2 curvature decreases for larger angles in our case,

⁸This conservative behavior is a generic feature of the Jackknife

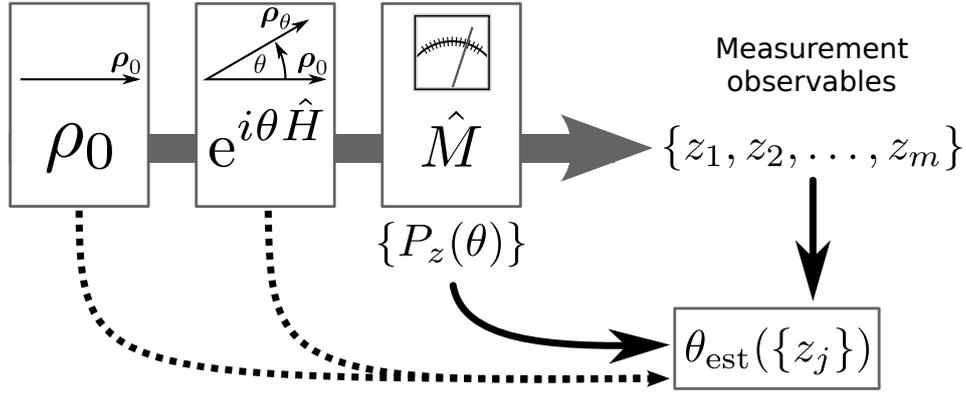


Figure 2.11 Quantum parameter estimation setting. The input state ρ_0 undergoes a unitary transformation (rotation) which depends on the parameter (angle) θ and a measurement, represented by the operator \hat{M} . Either the conditional output probabilities $P_z(\theta)$ or precise knowledge about the input state and the transformation can be used to construct an estimator θ_{est} , which allows to infer the value of θ from the observable outcomes of the measurement. The precision of θ_{est} is ultimately limited by the Quantum-Cramér-Rao bound $\Delta\theta_{\text{est}}^2 \geq 1/mF_Q[\rho_0, \hat{H}]$ with the Quantum Fisher information F_Q , which is the maximum of the Fisher information $F[\rho_0, \hat{H}, \hat{M}]$ over all measurement operations \hat{M} consistent with quantum mechanics and thus $F_Q[\rho_0, \hat{H}] \geq F[\rho_0, \hat{H}, \hat{M}]$.

which is the generic behavior of peaked distributions and leads to a slight underestimation of the true Fisher information. This trend is shown systematically in the lower right panel. A different simulation with Gaussian distributions showed, that the allowable angle range for an underestimation of 10% on the Fisher information is roughly $\pm 90^\circ/\sqrt{F}$ in this case. Non-Gaussian settings are typically more demanding, as in our example. The scaling with $1/\sqrt{F}$ is expected from the Cramér-Rao bound.

2.5 Quantum Fisher Information

Since quantum mechanical states are of statistical nature intrinsically, the concept of Fisher information provides a natural view on parameter estimation settings, where the 'quantum noise' is the fundamentally limiting factor. Optimization over all measurements consistent with quantum mechanics yields the Quantum Fisher information, which separates the question about the sensitivity of a specific state on a parameter from the question of the optimal measurement that has to be performed to take full advantage of it. It is by now well established [8, 9, 17, 92–94], that only unique quantum properties like entanglement at the stage of the parameter dependent transformation allow quantum states to be more sensitive than the best classical state.

The generic unitary parameter estimation setting is depicted in Fig 2.11. The state of a quantum mechanical system described by the statistical operator (density matrix) $\hat{\rho}_0$ undergoes a unitary transformation with \hat{H} being the generator of this transformation (i.e. $U = \exp(i\theta\hat{H})$). The goal is to estimate the value of the parameter θ from the outcomes of m measurements on

$\hat{\rho}_\theta$ with the help of a proper estimator, which is described by a quantum mechanical operator $\hat{\theta}_{\text{est}}$. For an unbiased estimator, the Quantum-Cramér-Rao bound is given by

$$(\Delta\hat{\theta}_{\text{est}})^2 \geq \frac{1}{mF_Q[\hat{\rho}, \hat{H}]} \quad (2.105)$$

with F_Q being the Quantum Fisher information [95], which is defined as the optimal value of the Fisher information over all measurements allowed by quantum mechanics, so called POVM (positive operator valued measures). It characterizes the maximally achievable precision of the estimator given the input state $\hat{\rho}$. The factor $1/m$ is the classical scaling for repeated independent measurements. For pure states ($\hat{\rho} = |\Psi\rangle\langle\Psi|$), the Quantum Fisher information reduces to the very simple form⁹

$$F_Q[\hat{\rho}, \hat{H}] = 4(\Delta\hat{H})^2, \quad (2.106)$$

where $(\Delta\hat{H})^2 = \langle\Psi|\hat{H}^2|\Psi\rangle - \langle\Psi|\hat{H}|\Psi\rangle^2$ denotes the quantum mechanical variance of the operator \hat{H} . This means, that the input state $\hat{\rho}$ is most sensitive on the parameter θ if it features a large spread in \hat{H} .

For example in our qubit setting with linear rotations about the z -axis, that is $\hat{H} = \hat{J}_z$, the pure state with the highest Quantum Fisher information is the one with the largest variance in \hat{J}_z , which is the N -particle GHZ(Greenberger-Horne-Zeilinger)-state (also called NOON-state or Schrödinger cat state)

$$|\text{GHZ}\rangle = |\text{NOON}\rangle = \frac{1}{\sqrt{2}}(|N, 0\rangle + |0, N\rangle). \quad (2.107)$$

It is the prototypical state, that attains the Heisenberg limit [9, 17, 96]

$$\Delta\theta_{\text{HL}}^2 = \frac{1}{mN^2}, \quad (2.108)$$

which is the maximum achievable precision on θ in the two-mode setting.

It is interesting to note, that the Quantum Cramér-Rao bound can be regarded as the precise version of a Mandelstam-Tamm uncertainty relation [97] between the parameter θ and its "conjugate" operator \hat{H} . For the original Mandelstam-Tamm type, the parameter would be time and \hat{H} the operator of the energy [98]. This is fundamentally different from the commutator type of uncertainty relation, since time is not an operator but a parameter in quantum mechanics. For fixed quantum Fisher information, time resolution can be only increased by increasing the spread in energy.

Take, for example, an ensemble of two-level atoms with the energy difference δE between the (single-particle) ground state $|\downarrow\rangle$ and the excited state $|\uparrow\rangle$. The unperturbed state of this system evolves with the Hamiltonian $\hat{H} = \delta E \hat{J}_z$. In this case, the Quantum Cramér-Rao bound for the unbiased estimation of the parameter $\theta = t\delta E/\hbar$ takes the appealing form

$$\Delta t\delta E \geq \frac{\hbar}{\sqrt{F_Q[\hat{\rho}, \hat{J}_z]}} \quad (2.109)$$

⁹A derivation of the Quantum Fisher information for mixed states can be found in Appendix A.1

Thus, if one wants to increase the precision of time estimation in an atomic clock, one aims for a quantum state with high F_Q and makes δE as big as possible, which corresponds to the intuitive notion of "more ticks per second". This is exactly the advantage of modern optical clocks [99], where δE is many THz compared to the traditional time standard involving hyperfine splittings in the GHz regime. For a pure state, using $\Delta E = \delta E \Delta \hat{J}_z$, this yields

$$\Delta t \delta E \geq \frac{\hbar}{2 \Delta \hat{J}_z} \quad \Rightarrow \quad \Delta t \Delta E \geq \frac{\hbar}{2}. \quad (2.110)$$

For a coherent spin state of N atoms with an 'energy spread' $\Delta E = \delta E \sqrt{N}/2$, this yields a limit for time resolution of $\Delta t = \hbar/(\sqrt{N} \delta E)$. In Ramsey spectroscopy, which is the most sensitive and important method for metrological applications, the signal is the accumulated phase $\varphi = T \delta E / \hbar$ during the interrogation time T and the signal-to-noise ratio is bound by $\text{SNR} \leq \varphi / \Delta \varphi = T / \Delta T = T \sqrt{F_Q} \delta E / \hbar = \varphi \sqrt{F_Q}$. For fixed δE , this can be increased by lengthening the interrogation time, which has the drawback of limiting the bandwidth, or by increasing $\sqrt{F_Q}$ beyond the limit \sqrt{N} achievable with coherent spin states of independent atoms.

2.6 Time evolution of Fisher information

As shown in 2.1.10, the quantum evolution starting on the classically unstable fixed point gives access to spin squeezed states for short evolution time as well as twisted non-Gaussian states at later time. In this section, we show quantitative results of numerical calculations for the case without external detuning (see Fig. 2.5). We compare the (Quantum) Fisher information to the optimal spin squeezing in the perfect noiseless case to the obtainable values in the presence of additive Gaussian noise, for example due to finite detection resolution on the state populations N_a and N_b .

Spin squeezing [10, 11] is characterized by the factor $\xi^2 = \xi_N^2 / [\langle \vec{J} \rangle^2 / (N/2)^2]$, where the number squeezing factor ξ_N^2 characterizes the spin variance perpendicular to the mean spin direction compared to an ideal coherent spin state of the same atom number, i.e. $\xi_N^2 = (\Delta J_\perp)^2 / (\Delta J_\perp)_{\text{CSS}}^2$ and $\langle \vec{J} \rangle^2 = \langle J_x \rangle^2 + \langle J_y \rangle^2 + \langle J_z \rangle^2 \leq (N/2)^2$ is the squared mean spin length. As the spin state increases in size, the curvature of the Bloch sphere becomes appreciable and the mean spin length reduces. Besides limitations due to technical noise, this fundamentally limits the obtainable visibility $\mathcal{V} \leq \langle \vec{J} \rangle^2 / (N/2)^2$ of a Ramsey sequence and thus the slope of the Ramsey fringe [100].

In a practical implementation of a spin squeezed Ramsey sequence [12], the state is rotated such that its shortest axis is aligned with the equator of the Bloch sphere, i.e. that its largest spread is in z -direction. After the Ramsey interrogation time, during which a phase θ can be accumulated (corresponding to a rotation around the z -axis), the short axis is aligned with the readout direction z by a final $\pi/2$ rotation. The complete action of this sequence is thus a rotation by the angle θ with the rotation axis $J_{\vec{n}}$ and the readout (projection) axis $J_{\vec{p}}$ perpendicular to each other and both perpendicular to the mean spin direction (see also

Fig. A.1). The inset in Fig. 2.12 shows an analogous situation for a non-Gaussian state. With the relations [9]

$$\Delta\theta_{\text{est}}^2 \geq \frac{1}{F_Q[\rho, J_{\vec{n}}]} \leq \frac{1}{F[\rho, J_{\vec{n}}, J_{\vec{p}}]} \leq \frac{\xi_{\text{min}}^2}{N}, \quad (2.111)$$

the normalized Fisher information $F_{(Q)}/N$ for a single linear rotation is readily comparable to the inverse spin squeezing factor $1/\xi^2$ ¹⁰. Both have a direct operational meaning for the maximal sensitivity of a Ramsey sequence. Since the spin squeezing factor is based on the spin variance, it characterizes the maximally obtainable precision for estimation of θ by the average of the observable¹¹, i.e. the mean spin component $\langle J_{\vec{p}} \rangle$ as projected to the readout direction \vec{p} and employing the phase estimator $\theta_{\text{est}} = \arcsin(2\langle J_{\vec{p}} \rangle/N)$ ¹². To take advantage of a non-Gaussian state with $F/N > 1/\xi^2$ requires methods like maximum likelihood or Bayesian estimation, taking into account substructures of the spin distributions beyond their mean position on the projection axis.

Figure 2.12 summarizes the numerical calculations. In the noiseless case, the Quantum Fisher information closely matches the spin squeezing result for short evolution times. For longer times, the bending of the state leads to a decrease of squeezing since the variance in all directions becomes bigger than that of the coherent spin state and additionally the mean spin length decreases. However, the quantum Fisher information continues to rise and eventually comes close to the Heisenberg limit. Interestingly, it turns out, that the rotation axis for maximal F_Q matches the optimal rotation axis for spin squeezing (perpendicular to the direction of minimal variance) for all evolution times. This is due to the fact that the optimal F_Q is obtained for the axis with maximal spread of the quantum state, which appears to be perpendicular to the axis of smallest variance even for the wrapped states. In the initial time evolution it moves from closely aligned with the tangent of the separatrix at the unstable fixed point towards the z -axis.

The calculation of the Fisher information F additionally requires the optimization of the measurement axis, which for the wrapped states becomes quite tedious since they obtain fine structures on the single particle level. The best values close to F_Q are found for the measurement axis parallel to the mean spin direction. This can be intuitively understood by looking at the Wigner distributions (Fig. 2.5). One relies on perfect visibility of the fine-structures in this case, which are maximal when projecting along the mean spin direction.

This situation is qualitatively changed by the addition of Gaussian noise along the projection axis, which we incorporated by convolution of the spin distributions with a Gaussian. We evaluated two cases: $\Delta J_{\text{meas}} = 0.7$, which is at the threshold to single particle resolution ($\Delta N_{a,b} \sim 1$)¹³ for the two components a and b , and $\Delta J_{\text{meas}} = 3$ ($\Delta N_{a,b} = 4$), which is the level of our detection noise. For both noisy cases, the prediction of the optimal rotation axis from the quantum Fisher information fails (dot-dashed line in Fig. 2.12 for $\Delta J_{\text{meas}} = 3$) and the optimal measurement axis is again found perpendicular to the mean spin direction.

¹⁰We focus here on linear projective measurements along a certain fixed direction \vec{p} for the calculation of F , since it is our experimental situation. Note that this already includes all higher moments of $J_{\vec{p}}$ in the resulting probability distributions.

¹¹This is also true for the non-Gaussian case in the limit of many measurements, where the distribution of the sample mean is well approximated by a Gaussian distribution due to the central limit theorem.

¹²For $J_{\vec{p}} = J_z$ and small angles close to the equator, $\theta_{\text{est}} \approx z = (N_b - N_a)/N$.

¹³This is also equivalent to the loss of a single atom prior to detection.

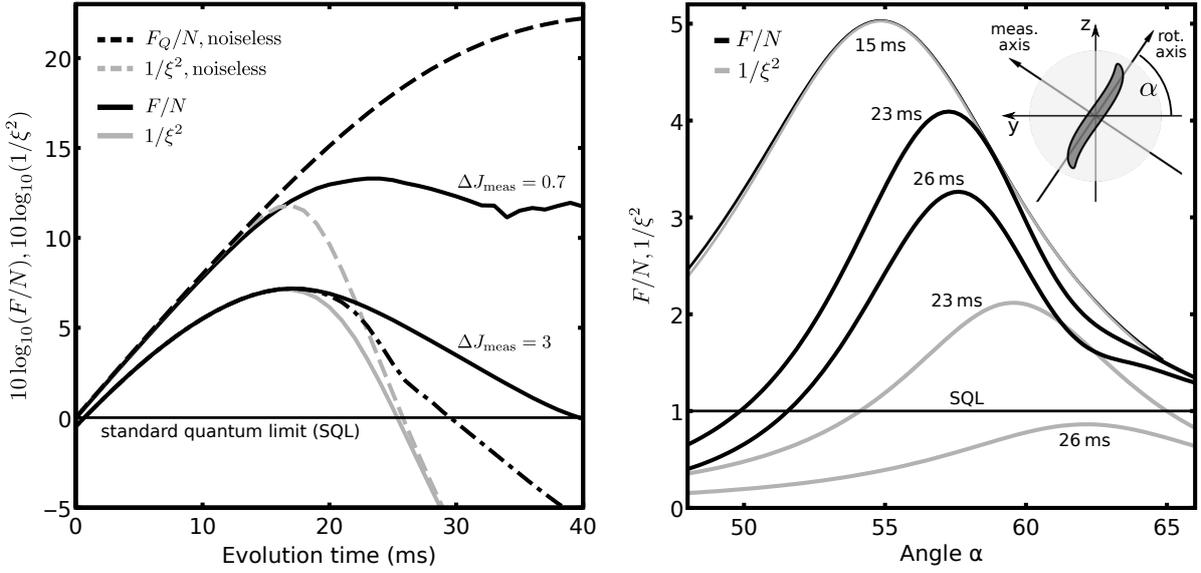


Figure 2.12 Time evolution of Fisher information and spin squeezing. The time evolution without external detuning δ starting with a coherent spin state from the unstable fixed point is calculated numerically for a fixed atom number $N = 300$, interaction parameter $\Lambda = 1.5$ and Rabi frequency $\Omega = 2\pi \times 20$ Hz. The dashed lines correspond to the ideal noiseless case, where the Quantum Fisher information (QFI) rises to a maximum value ~ 22 dB, a factor 2 away from the ultimate Heisenberg limit (~ 25 dB for $N = 300$), whereas the spin squeezing is limited to ~ 12 dB, about a factor 10 lower. The solid lines correspond to the same situation with added Gaussian noise along the measurement axis (threshold to single atom resolution $\Delta J_{\text{meas}} = 0.7$, realistic detection resolution $\Delta J_{\text{meas}} = 3$). The optimization for distinct evolution times is shown in the right panel. Rotation axis, measurement axis and mean spin direction were kept mutually perpendicular and the angle α was varied (see inset). One observes that the angle for optimal Fisher information starts to deviate from the angle of best spin squeezing as the state enters the non-Gaussian regime. The dash-dotted line in the left panel corresponds to the maximum Fisher information for the noisy state with $\Delta J_{\text{meas}} = 3$, but calculated for the optimal rotation axis of the noiseless QFI, which turns out to be not the optimal choice.

Inspired by the spin squeezed case, we therefore consider a scheme with the measurement axis perpendicular to the rotation axis and both perpendicular to the mean spin direction as depicted schematically in the inset. An optimization over the angle α yields a smooth continuation of the spin squeezed states by the noisy wrapped states with the Fisher information beyond the shot-noise limit ($F/N > 1$) where the spin squeezing has already vanished even without added Gaussian noise. The angle α for optimal Fisher information increasingly differs from the angle of minimal variance, i.e. best spin squeezing. This is due to the fact that the wrapped state is still sharply peaked in the center, which is maximally visible along the measurement axis at an angle of increased variance (where the 'tails' of the state are also appreciable). This can be taken into account and exploited by an advanced phase estimation protocol, but not with standard averaging.

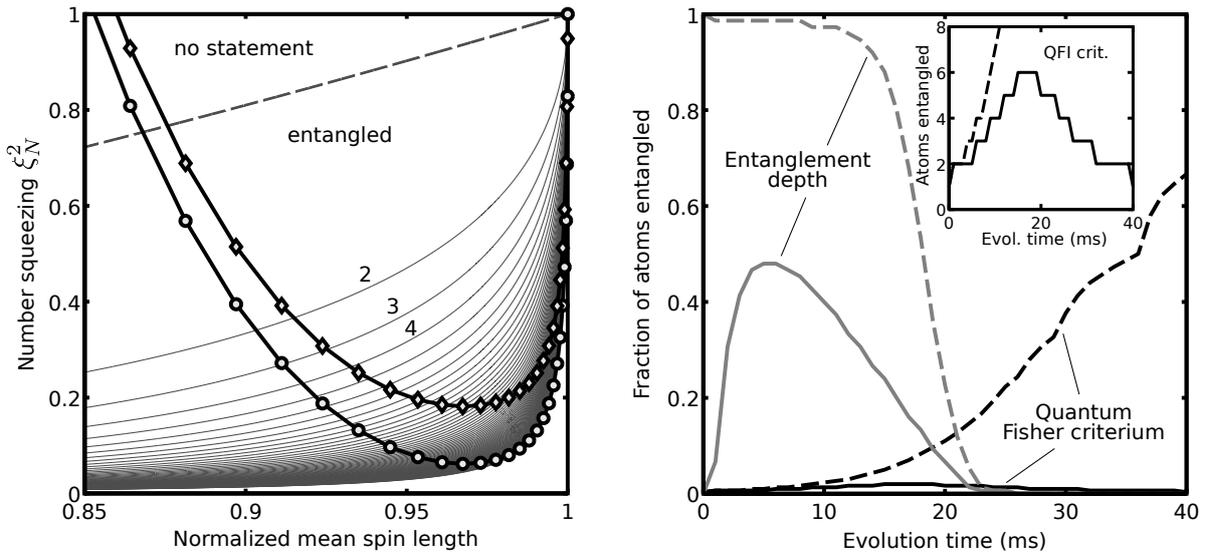


Figure 2.13 Entanglement criteria – Fisher information vs. depth of entanglement. **Left:** The numerical solution of the time evolution starting on the unstable fixed point for $N = 300$ and fixed parameters with $\Lambda = 1.5$ and $\Omega = 2\pi \times 20$ Hz, moving in the parameter space of spin squeezing, i.e. normalized mean spin length (Ramsey visibility) \mathcal{V} and number squeezing parameter ξ_N^2 . The points mark steps of 1 ms, starting in the top right corner. The thin gray lines are the (analytical) bounds given in [101] for entanglement depth. The dashed line represents the condition $\xi_N^2/\mathcal{V}^2 < 1$, sufficient for entanglement. Additional Gaussian noise (typical value $\Delta J_{\text{meas}} = 3$, corresponding to our detection noise but also representative for atom loss, diamonds) leads to a limitation of the reached entanglement depth. **Right:** Comparison of the entanglement criteria for the time evolution in the ideal case (dashed lines) and with Gaussian noise (solid lines). Since the entanglement in a slightly spin squeezed symmetric state is shared between all atoms, the entanglement depth is maximal initially. For the noisy case it obtains a maximum at finite evolution time before the point of maximum number squeezing. The restrictive Fisher information criterium is only sensitive to metrologically relevant entanglement and thus gradually increases with time. The inset shows a zoomed y -axis. The large Fisher information and thus entanglement expected for later evolution times is increasingly contained in small phase space structures and thus very sensitive to added Gaussian noise. Note that 6-particle entanglement according to the Fisher criterium corresponds to a metrological advantage of ~ 8 dB, which shows the large scale of this plot.

2.7 Fisher Information and entanglement

A collection of two-level systems (qubits) is the prototypical setting to discuss quantum entanglement of many-body systems. A system of N distinguishable qubits is called fully separable, if it can be written in the product form

$$\rho = \sum_k p_k \rho_k^{(1)} \otimes \rho_k^{(2)} \otimes \cdots \otimes \rho_k^{(N)} \quad (2.112)$$

with $p_k \geq 0$ and $\sum_k p_k = 1$. $\rho_k^{(j)}$ refers to the density operator of qubit j . Otherwise it is called entangled (non-separable) [102]. If a specific state can be written in this form is highly non-trivial. Inequalities based on (variance) measurements of the collective (pseudo-)spin

have been devised [101, 103, 104], which can detect entanglement in the presence of spin squeezing. It has been pointed out that the observation of a Fisher information $F > N$ for linear collective transformations ($\hat{H}_{\text{lin}} = \hat{J}_{\vec{n}}$) signals the presence of entanglement [9]. Furthermore, it was shown that the value of the Fisher information signals multi-particle entanglement. In particular, a violation of the inequality

$$F_Q[\rho, \hat{H}_{\text{lin}}] \leq sk^2 + r^2 \quad (2.113)$$

signals genuine $(k + 1)$ -particle entanglement [105]. Here, s is the integer smaller or equal to N/k and $r = N - sk$. If N is divisible by k , the right-hand side simplifies to Nk , hence an observation of $F/N > k$ signals $(k + 1)$ -particle entanglement. The Fisher information defines an equivalence class of entangled states, since it recognizes and classifies by definition all states that contain entanglement useful for the improvement of parameter estimation tasks, including spin squeezed states [9].

Comparison with entanglement depth The observation of a spin squeezing factor $\xi^2 < 1$ also implies entanglement [103]. Due to the Heisenberg uncertainty relations for the spin operators, spin squeezing for a given number of qubits is bound from above and thus a depth of entanglement can be defined by comparing with this limit [101, 106, 107]. The spin squeezing depth of entanglement characterizes the smallest possible division into independently squeezed subsystems which is consistent with a given (i.e. measured) mean spin length and number squeezing factor¹⁴.

Along a similar line of thought but much more restrictive, the entanglement criterion based on the Fisher information considers the smallest possible subdivision into *any* entangled states with individually higher $F_{\text{sub}}/N_{\text{sub}}$. This means that the limit is only given by a subdivision into maximally entangled NOON-states (Eg. 2.107), which maximize the quantum Fisher information of the subsystems. For example, the smallest subensembles that can lead to $F = 2N$ are $N/2$ maximally entangled pairs with $F_{\text{sub}} = 2^2 = 4$. Consequently, $F/N > 2$ means that there has to be at least one non-separable subsystem with at least 3 particles.

In Figure 2.13, we compare the two entanglement criteria during the numerical time evolution on the unstable fixed point. In the ideal case, the initial time evolution is very close to the maximally squeezed state for a given mean spin length and thus the spin squeezing entanglement depth is very high, close to the total number of atoms. The addition of Gaussian noise alters this drastically and the entanglement depth is maximal for finite evolution time. In contrast, the Fisher information criterion starts at zero and shows the gradual build-up of metrologically relevant entanglement. Since this is increasingly accompanied by small phase space substructures, it is also very sensitive to noise.

2.7.1 Entanglement and symmetric states

The derivations of entanglement criteria assume the distinguishability of the qubits. If we assume in our case that only two spatial modes $\psi_a(\mathbf{x})$ and $\psi_b(\mathbf{x})$ are accessible, the spin-wavefunction has to be symmetric with respect to particle exchange and partitioning in unam-

¹⁴Depth of entanglement criteria have also been recently devised and employed for the characterization of states close to W-states [28, 108, 109] and symmetric Dicke states [110]

ambiguous subsystems to define entanglement is not dictated by the physical arrangement any more¹⁵.

An important example in this discussion are the Fock or Dicke states $|k, N - k\rangle$ (see 2.1.2), which are the eigenstates of the \hat{J}_z -operator and form a natural basis for indistinguishable qubits. They are separable in the two modes [111] but highly entangled, when distinguishable particles are considered as subsystems. The first Dicke state (also called W-state [28, 112]) for example reads for 3 particles

$$|1, 2\rangle = \frac{1}{\sqrt{3}}(|\uparrow\downarrow\downarrow\rangle + |\downarrow\uparrow\downarrow\rangle + |\downarrow\downarrow\uparrow\rangle), \quad (2.114)$$

which is inseparable with respect to the above definition and features a maximal Quantum Fisher information $\max_{\vec{n}} F_Q[|k, N - k\rangle, \hat{J}_{\vec{n}}] = N(1 + 2k - 2k^2/N) = 7$, which shows 3-particle entanglement. In contrast, any coherent spin state of the 3 particles, for example

$$|\theta = \frac{\pi}{2}, \phi = 0\rangle_3 = \left[\frac{1}{\sqrt{2}}(|\downarrow\rangle + |\uparrow\rangle)\right]^{\otimes 3} = \frac{1}{2\sqrt{2}}(|0, 3\rangle + \sqrt{3}|1, 2\rangle + \sqrt{3}|2, 1\rangle + |3, 0\rangle), \quad (2.115)$$

is separable in the particle basis, features $\max_{\vec{n}} F_Q[|\theta, \phi\rangle_N, \hat{J}_{\vec{n}}] = N = 3$ but is non-separable with respect to the modes ψ_a and ψ_b . However, irrespective of the modes, an individual preparation of N single qubits in $1/\sqrt{2}(|\downarrow\rangle + |\uparrow\rangle)$ would produce the same result in every protocol that does not contain couplings between the qubits, such that no particle entanglement can be established. Yet the same phenomena would be observed when preparing and probing the qubits one by one sequentially. In this respect, the quantum state of a single particle prepared in a superposition of the two modes is (mode-)entangled [113, 114].

Thus, there is no definite answer to the questions 'Does the system contain entanglement?' or 'Is the system separable?', since the choice of the subsystems changes the answer to these questions [115]. Moreover, in view of technological applications, it depends on the task, which kind of inseparability of the quantum state is required to surpass classical performance or to enable operations that are impossible for classical systems [116].

It was argued that since the Dicke states are separable with respect to the modes, entanglement is not required for quantum enhanced metrological tasks [117]. However it was shown recently that the effective particle entanglement contained in symmetric states can be extracted into entanglement of independent modes [118]. The idea of the protocol is to apply (spatial) beamsplitters, which leave the internal state unaltered but can be sequentially applied to achieve any desired partitioning of the system. The crucial observation is that this protocol produces no additional correlations and all the mode-entanglement stems from the structure of the initial state, which is even in one-to-one correspondence with the initial structure. For example, the above Dicke state can be written as

$$\frac{1}{\sqrt{3}} \left(\sqrt{2} \left[\frac{1}{\sqrt{2}} (|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle) \right]_1 \otimes [|\downarrow\rangle]_2 + [|\downarrow\downarrow\rangle]_1 \otimes [|\uparrow\rangle]_2 \right), \quad (2.116)$$

which is identical to what would be observed by fictitious beamsplitting into two spatial modes 1 and 2 combined with projection onto local particle numbers (two particles in mode 1, one

¹⁵If every participating particle is localized, a 'natural' choice is the use of the individual particles as subsystems. However, this is also arbitrary to some degree.

particle in mode 2). In an experimental situation, this could be achieved by counting the total number of particles in each spatial mode for each realization combined with suitable postselection.

In experiments with photons, beamsplitting operations are routinely employed to create entangled modes (beams) [20, 119, 120]. They already proved to be a viable route to various Einstein-Podolsky-Rosen entangled states in the optical continuous variable setting [121–124] and corresponding protocols have been proposed for the atomic system [125, 126].

In the extreme case of splitting into many modes, such that the probability for having two atoms in a single mode tends to zero, the originally contained resource could be transformed into N -mode entanglement. This limiting case is also known as the dilute cloud argument [93, 127], where one imagines a situation in which the original physical cloud of indistinguishable qubits is diluted¹⁶, such that in each bin of the detector (pixel on the CCD camera) there is at most one atom. If the detection is sufficiently fast, local interactions between the atoms are negligible during the probe time. Since the collective internal state is unaffected by the dilution [127] the detector bins can be viewed as individual modes and thus particle entanglement is equivalent to mode entanglement.

When collective observables are sufficient for the estimation task or to characterize the internal state, this splitting is not necessary. Therefore, we can apply criteria derived for collective states of distinguishable qubits for our situation of indistinguishable particles. The state of indistinguishable qubits is not fundamentally different from a symmetric state of distinguishable qubits and thus a fictitious labeling does not lead to false conclusions, as long as no operations are involved which specifically address single qubits. Entanglement criteria based on collective observables are especially useful if the symmetry of the state is not a prerequisite, as it is the case for the Fisher information criterion.

¹⁶This is easily achieved by switching off the external trapping potential and letting the cloud expand in free fall. Usually, interactions that can alter the internal state then vanish very fast.

3 Experimental system

Our experimental setup close to its present form has already been described in numerous theses [41, 107, 128–130] so we will only briefly discuss the BEC production. Since the final optical traps are an essential part of the environment of the mesoscopic BECs, we detail on their characteristics. Furthermore, we describe the interaction with static, radio-frequency and microwave magnetic fields, which we use for spin state manipulation. Furthermore, we discuss the collisional two-body interaction, which is in our system essential for the creation of nonclassical spin states. Finally, we characterize the atomic loss from inelastic collisions.

3.1 The route to BEC

Each experimental run begins with a large volume magneto-optical trap (MOT) which is loaded from a two-dimensional MOT in about 3 s. After a short (~ 3 ms) sub-doppler cooling step in an optical molasses, the atoms are optically pumped to the hyperfine manifold $F = 1$, of which the state with $m_F = -1$ is low-field seeking, and transferred to a magnetic trap. Shortly after loading, the remaining atoms in $F = 2$ are pushed out by shortly pulsing on the MOT laser beams. The magnetic trap consists of a quadrupole field with a vertical gradient of ~ 250 G/cm and a superimposed horizontally rotating bias field with a maximum amplitude of ~ 30 G to avoid spin flips in the center of the trap, where the magnetic field vanishes. This time-orbiting-potential (TOP) configuration is very robust and offers the advantage of easy and reliable implementation of evaporative cooling by ramping down the amplitude of the rotating bias field (circle-of-death evaporation). In this way, the hottest atoms, which are able to reach the edges of the cloud, can perform spin flips when they hit the rotating magnetic zero and are expelled from the trap. The remaining atoms rethermalize and thus cool down. After ~ 25 s in the magnetic trap we are left with $\sim 10^6$ ultracold but still thermal atoms, which can be subsequently loaded into an optical dipole trap formed by two crossed beams with a wavelength of 1030 nm from a multimode solid state Yb:YAG laser (Innolight Corona with a specified longitudinal mode spacing of 15 GHz). The crossing point of the two beams is located ~ 100 μm underneath the position of the magnetic trap at full strength. By ramping down the current of the quadrupole coils, the cloud position is lowered due to gravity and sweeps over the dipole trap beams. This procedure ensures a robust loading of the dipole trap. One of the beams (called 'waveguide') is used throughout the rest of the experiment, while the crossing light sheet ('XDT') is just needed to enhance elastic scattering for efficient rethermalization. After forced evaporation by lowering the laser power of both beams, we are left with a Bose-Einstein condensate of $\sim 10^4$ atoms with indiscernible thermal fraction. The value of the lowest laser power is used to control the atom number in the BEC. Afterwards, the intensity of the waveguide is again increased, while the XDT is slowly ramped to zero, which results in a very elongated cloud in a quasi one-dimensional situation [41]. For the

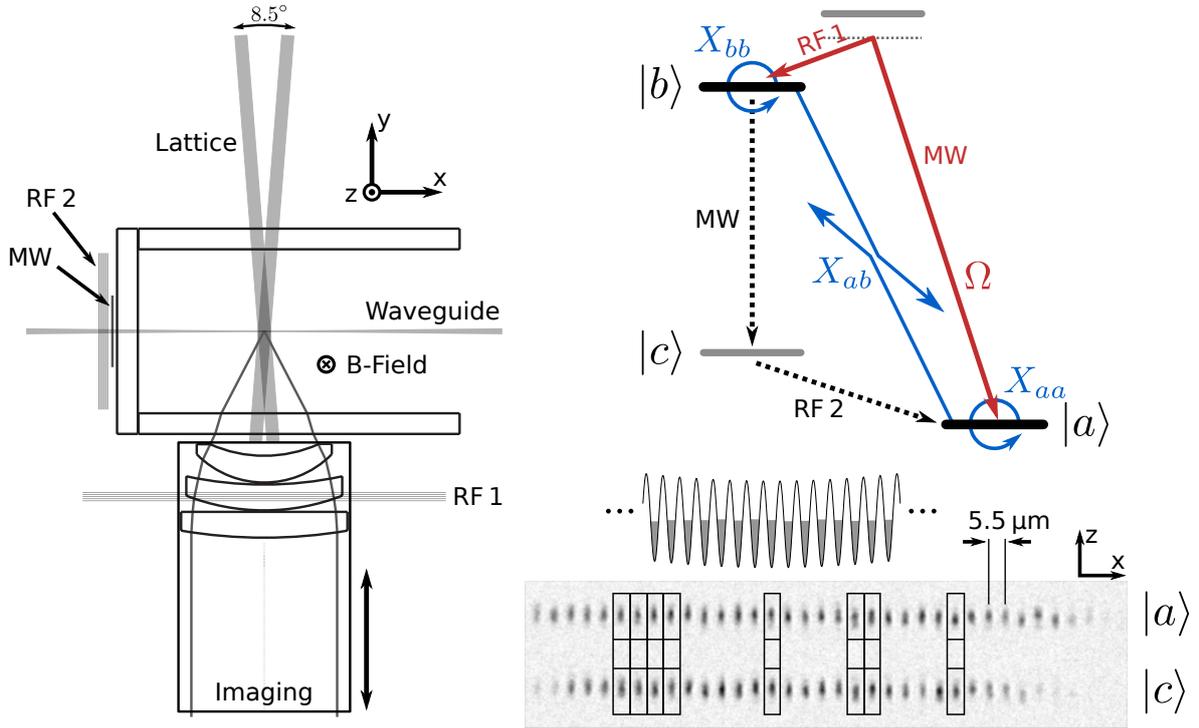


Figure 3.1 Geometry of the experiment and physical system. **Left:** The view from above on the vacuum glass cell with the movable imaging objective shows the arrangement of the optical traps for the simultaneous creation of up to 35 independent BECs with atom numbers in the range $N \sim 100 \dots 600$. The direction of the magnetic offset field of 9.13 G is indicated as well as the arrangement of the radio frequency (RF) and microwave (MW) loop coils to drive magnetic transitions between the hyperfine states. **Right, top:** The sketch schematically shows the involved hyperfine states and their couplings. The levels $|a\rangle = |F = 1, m_F = 1\rangle$ and $|b\rangle = |2, -1\rangle$ form the interacting pseudo-spin two-mode system used to create nonclassical spin states. The atoms initially condense in $|c\rangle = |1, -1\rangle$, from where we transfer them with high fidelity to $|a\rangle$ by an RF sweep. The coupling between $|a\rangle$ and $|b\rangle$ is implemented with MW and phase controlled RF magnetic fields, resulting in a coupling strength (Rabi frequency) Ω . Also the two-body interactions relevant for the experiments are depicted, leading to the nonlinearity $\chi = X_{aa} + X_{bb} - 2X_{ab}$ and a shift of the resonance $\propto (N - 1)(X_{bb} - X_{aa})$. **Right, bottom:** Sketch of the lattice potential and exemplary absorption image after Stern-Gerlach separation. Before the imaging, the population of $|b\rangle$ is transferred to $|c\rangle$ by a microwave π -pulse.

preparation of mesoscopic condensates, an optical lattice potential is formed by crossing two laser beams under a shallow angle of $\alpha \sim 8.5^\circ$ at the position of the atoms which results in a slicing of the cloud and recondensation in each lattice well. Figure 3.1 gives an overview on the arrangement of the final trapping potentials. The trapping frequencies for the spin measurements are $\sim 2\pi \times 260$ Hz transversally and $\sim 2\pi \times 660$ Hz in lattice direction.

3.2 Pseudospin system

In this subsection, we will first briefly discuss the implementation of the spin Hamiltonian

$$\hat{H} = \chi \hat{J}_z^2 + \Omega(\sin \phi \hat{J}_x + \cos \phi \hat{J}_y) + \delta \hat{J}_z \quad (3.1)$$

and later detail the experimental characterization of χ and the various detuning effects, which contribute to δ and have to be considered in the experiment. The magnetic sublevels that constitute the atomic two-level, i.e. qubit system with their couplings and interactions are schematically depicted in Fig. 3.1. Typical values of the parameters in the experiment are $\chi \sim 2\pi \times 0.06$ Hz, $\Omega \lesssim 2\pi \times 350$ Hz and $|\delta| \lesssim 2\pi \times 2$ Hz.

3.2.1 Interaction

We use the pseudo-spin system $|a\rangle = |F = 1, m_F = 1\rangle \leftrightarrow |b\rangle = |2, -1\rangle$, which features a Feshbach resonance at a magnetic field of 9.1 G [131–134]. This can be used to modify the interspecies interaction X_{ab} , which is essential for enhancing the otherwise negligible nonlinearity $\chi = X_{aa} + X_{bb} - 2X_{ab}$. We work at a precisely controlled magnetic offset field of 9.13 G on the high side of the resonance, which results in a reduction of the interspecies scattering length and thus leads to a reduction of X_{ab} , i.e. $\chi > 0$. The miscibility of the two components is ensured by this choice¹. Due to the slightly different intraspecies interactions, the two modes are not expected to be perfectly overlapping in the ground state [41, 129]. Since we start with all atoms in $|a\rangle$, the nonadiabatic preparation of a superposition of $|a\rangle$ and $|b\rangle$ will lead to a slight breathing motion of the modes, which is expected to be negligibly small for our high trapping frequencies [107], such that the single spatial mode approximation can be applied².

3.2.2 Linear Coupling

Due to the required change in m_F by two, Rabi coupling of $|a\rangle$ and $|b\rangle$ can only be realized with a two-photon process. For this purpose, we work with microwave (MW) and radio frequency (RF) magnetic fields both red detuned by $\Delta_{\text{one}} = 2\pi \times 200$ kHz with respect to the 'intermediate level' $|F = 2, m_F = 0\rangle$. The microwave frequency bridges the large hyperfine splitting of 6.835 GHz between $F = 1$ and $F = 2$, whereas the comparably small RF frequency (6.2 MHz) spans the energy difference between $|2, 0\rangle$ and $|b\rangle$ due to the Zeeman splitting. This coupling scheme has the advantage that the RF can be generated and manipulated by an arbitrary waveform generator (AWG). In this way, arbitrary pulse sequences with controlled phase ϕ , detuning δ and Rabi frequency Ω can be programmed with high precision. The phase of the first pulse defines here the frame of reference and can be set to $\phi_1 = 0$ without loss of generality. The two-photon Rabi frequency is connected to the resonant one-photon Rabi frequencies by

$$\Omega = \frac{\Omega_{\text{MW}}\Omega_{\text{RF}}}{2\Delta_{\text{one}}}. \quad (3.2)$$

The maximum values of Ω_{MW} and Ω_{RF} are chosen sufficiently big ($\sim 2\pi \times 12$ kHz) to surpass the nonlinearity during spin rotations and sufficiently small to keep population of $|2, 0\rangle$ at a negligible level ($\Omega_{\text{MW/RF}}^2 / (\Omega_{\text{MW/RF}}^2 + \Delta_{\text{one}}^2) \lesssim 0.4\%$). For further technical details on the MW/RF system see Appendix B.3.

¹The miscibility parameter at this field is $\Delta_{\text{mis}} = g_{aa}g_{bb}/g_{ab}^2 > 1$ [41]

²Note that we do not rely on its validity for most of the conclusions in this thesis.

3.2.3 Detuning

The parameter δ represents the total deviation of the difference of MW and RF frequencies, which defines the frame of reference, from the atomic two-photon resonance. Since the obtainable nonlinearity is relatively weak, we have to work with very small Rabi frequencies typically smaller than $2\pi \times 30$ Hz to reach the interaction-dominated (Josephson) regime with $\Lambda = N\chi/\Omega > 1$. This makes the system susceptible to very small detunings on the order of a few Hertz. Thanks to the linear Zeeman shift of both levels being very similar, the transition has a residual magnetic field shift of only $2\pi \times 10.7$ Hz/mG. This is still demanding, since it requires a magnetic field stability well below 1 mG. With an active magnetic field control, detailed in Appendix B.2, we achieve a short-term stability typically better than 30 μ G. Remaining slow drifts are compensated by automated Ramsey experiments on the two-photon transition. Other important shifts of the resonance are due to AC Zeeman shifts, which are caused by the MW and RF fields driving their respective far detuned one-photon resonance, and interaction shifts (both detailed below).

3.2.4 Detection

After each spin manipulation sequence, we employ a Stern-Gerlach separation, which consists of a 2 ms current pulse on the quadrupole coils of the magnetic trap, during which the waveguide trap is quickly switched off. After a subsequent time of flight of ~ 2 ms, the hyperfine spin states with different magnetic moment separate vertically, while still being confined by the lattice potential. The short time of flight is crucial for the reduction of the atomic density³. A positive side effect is the contraction of the clouds in horizontal direction due to the lattice potential, which facilitates the distinction of the wells.

The individual components are detected by far-above saturation absorption imaging⁴ [135] with high spatial resolution (~ 1.1 μ m). We use the fast kinetics mode of the CCD camera⁵ to acquire five frames with a spacing of 1.2 ms (600 μ s exposure time, 656 μ s shifting time). For this, all but ~ 200 pixel rows of the active region of the CCD are masked by a razor blade directly attached to the camera window. During the first four frames, laser light resonant with the $F = 2$ manifold is pulsed on for 15 μ s. With the pulse for the second frame, repumper light resonant with the $F = 1$ manifold is additionally applied perpendicular to the imaging direction, which optically pumps $F = 1$ population to $F = 2$. The repumper laser is kept on for the subsequent two frames. For the last frame all light is extinguished.

The first frame thus carries the absorption signal of $F = 2$ population, which is quickly accelerated and removed by the strong laser pulse. Absorption signal on the second frame represents the $F = 1$ population, while the third frame is used as reference, ideally containing no absorption signal.

³The subsequent imaging procedure assumes that the atomic column density is approximately constant over the equivalent of one pixel in object space, which is $\sim 420 \times 420$ nm in our case. Strong deviations from this requirement lead to nonlinear behavior, because the connection of absorbed photons to atom number is nonlinear. We observed also diffraction effects when imaging in the trap, which vanished with the time of flight.

⁴typically $\sim 12I_{\text{sat}}$, such that the maximum optical density is smaller than 2

⁵Princeton Instruments PIXIS 1024BR

The forth frame is illuminated to enhance the quality of the reference frame, since each absorption frame contains a small amount of leakage light if the succeeding frame is also illuminated. This results from edge diffraction at the razor blade into the masked region, where the charges of the previous frames are stored before readout. The illumination of the forth frame thus gives the third (reference) frame the right amount and structure of leakage light.

The last frame without illumination is used to determine the ADC offset of the camera, which is subtracted from all frames before any further processing.

In post-processing, the acquired reference frames of typically 700 subsequent repetitions of the experiment are used in linear combination, whose coefficients are found by minimizing the squared per-pixel distance to the absorption frame in specified regions that do not contain atomic absorption signal [136]. The positive effects of this procedure are a significant reduction of the photon shot noise by almost a factor two in variance and a further reduction of the (already small) fringe noise contribution. The shot-noise reduction happens because the linear combination averages uncorrelated noise of the individual pictures. A further advantage is that the reference image is automatically normalized to the intensity of the absorption image.

With the help of this optimized reference frame, the absorption signal is converted into atomic column density per pixel⁶ and integrated in elliptical regions⁷ individually centered on the two components of each well. We systematically varied the size of these regions and chose them sufficiently big without cutting away atomic signal. Further details on the detection scheme and its calibration can be found in [130, 137]. The typical detection noise (one std.) on a single component is ± 4 atoms, which only weakly depends on the imaging intensity.

For each experimental picture, a Gaussian error propagation is performed, which takes into account the camera gain calibration and the calibration factors of the conversion from camera counts to atoms/pixel. This 'noise image' is integrated in the same region as the atomic signal, which yields a very reliable online value for the photon shot noise part of the detection noise.

3.3 Optical traps

Here, we detail the physical characteristics of the optical dipole traps, which determine the atomic density and thus the interaction strength of the two modes. A more technical characterization can be found in the Appendix B.1.

3.3.1 Lattice

The lattice beams are derived from a Ti:Sa ring laser (Coherent 899), have a wavelength of $\lambda \sim 820$ nm and are thus red detuned to both the D2 (780 nm) as well as the D1 (795 nm) line of ⁸⁷Rb. They form a one-dimensional lattice along the waveguide, which has a spacing of $d = \lambda/[2 \sin(\alpha/2)] \approx 5.5 \mu\text{m}$. The trapping frequency along the lattice was determined by modulating the phase of one of the two beams with an electro-optic modulator. If the resulting shaking of the lattice is resonant with the on-site trapping frequency, this leads to a transfer to excited trap states accompanied by a detectable growth in transversal cloud size. Due to the

⁶This is achieved by an integration of the Lambert-Beer law for a saturable absorber, whose coefficients are calibrated with a measurement of the atomic shot-noise of a coherent state

⁷A typical size of their principal axes is 14 and 36 pixels

sinusoidal form of the trapping potential, the obtained resonance frequency depends on the magnitude of the modulation. This can be taken into account by successively decreasing the driving amplitude and extrapolating to zero. The inhomogeneity along the lattice is found to be less than 2% (see Appendix B.1). From the local trapping frequency, we get an estimate for the barrier height of the optical lattice, if we assume the potential to take the form [138]

$$V_{\text{lat}}(x) = V_0 \cos^2(kx) \quad (3.3)$$

with $k = \pi/d$. Taylor expansion at the minima $x_0 = \pi/2 + n\pi$ yields the harmonic approximation

$$V_{\text{lat}}(x) \approx V_0 k^2 (x - x_0)^2 = \frac{M}{2} \omega_{\text{ho}}^2 (x - x_0)^2 \quad (3.4)$$

and thus the barrier height

$$V_0 = \frac{M \omega_{\text{ho}}^2}{2k^2} = \frac{M \omega_{\text{ho}}^2 d^2}{2\pi^2}. \quad (3.5)$$

With the mass $M \approx 1.443 \times 10^{-25}$ kg this leads to $V_0 \approx 2\pi\hbar \times 5.7$ kHz for the highest trapping frequency of $2\pi \times 660$ Hz. Compared to the recoil energy $E_r = \hbar^2 k^2 / (2M) \approx 2\pi\hbar \times 19$ Hz, this is a factor $s = V_0/E_r \approx 300$ higher. This is deep in the tight binding limit and the energy scale of coherent tunneling in the lowest band can be estimated to [139]

$$\frac{4J}{E_r} = \frac{16}{\sqrt{\pi}} s^{3/4} \exp(-2\sqrt{s}) \approx 5 \times 10^{-13} \quad (3.6)$$

The corresponding time scale of $\sim 10^{11}$ s is way below relevance. Even the full photon recoil energy of $\hbar^2 k_L^2 / (2M) = 2\pi\hbar^2 / (2\lambda^2 M) \approx 2\pi\hbar \times 3.4$ kHz is below the barrier height. This means that only multiple photon scattering from the lattice beams allows to acquire sufficient energy for classical hopping. Since the atoms are trapped close to the maxima of the lattice, the photon scattering rate can be estimated to $\hbar\Gamma_{\text{sc}} = \Gamma V_{\text{lat}}/\Delta$ and in our case of the two relevant transitions to [140]

$$\hbar\Gamma_{\text{sc}} \approx \Gamma V_{\text{lat}} \frac{2\Delta_1^2 + \Delta_2^2}{\Delta_1 \Delta_2 (2\Delta_1 + \Delta_2)}. \quad (3.7)$$

Here, Δ_1 and Δ_2 are the detunings from the D_1 and D_2 lines respectively. For 820 nm and $\Gamma \approx 2\pi \times 6$ MHz this yields $\Gamma_{\text{sc}} \lesssim 15$ mHz and thus a worst case heating rate [140] of $(2/3)E_r\Gamma_{\text{sc}}/k_B = (2/3)(\hbar^2 k_L^2 / 2M)\Gamma_{\text{sc}}/k_B \lesssim 2$ nK/s, if all atoms are retrapped after scattering a photon. Besides heating effects before the actual spin experiments, the expected (classical) hopping rate is much too small to cause any significant effect.

3.3.2 Transversal confinement

The confinement in transversal direction is provided by a focused laser beam (waveguide) with a wavelength of 1030 nm. From trap frequency measurements (see Appendix B.1), we can estimate its Gaussian waist (2σ -radius) in the focus to $2\sigma_{\text{WG}} \approx 25.7$ μm . This leads to an aspect ratio of $\omega_{\perp}/\omega_{\parallel} \approx 110$ and trap depths of $(V_{\text{hor}}, V_{\text{vert}}) \approx 2\pi\hbar \times (106, 40)$ kHz⁸ at our working point. The high aspect ratio leads to a very elongated cloud of the expanded BEC

⁸ $V_{\text{vert}} < V_{\text{hor}}$ due to the gravitational potential, which weakens the trap in vertical direction

and thus to the population of many lattice wells with similar mean atom number in the central part. The clear disadvantage of the shallow focus is the high trap depth and thus high laser power needed to obtain a reasonably strong confinement. This is accompanied by an increased ability to trap thermal atoms. In terms of temperature, the trap depth of 40 kHz corresponds to 2 μ K, significantly higher than the recoil temperature $\hbar^2 k_L^2 / (2Mk_B) \approx 165$ nK of lattice beam photons.

In comparison, the previously employed 'charger' trap, which consisted of a more tightly focused beam with a Gaussian waist of $2\sigma_C \sim 5 \mu\text{m}$, allowed to reach a typical transversal trapping frequency of $2\pi \times 460$ Hz with trap depths of only $2\pi\hbar \times (11, 1)$ kHz, corresponding to ~ 50 nK in the weak direction, such that scattering from the lattice beams mainly caused an increased single particle loss rate, because retrapping was unlikely. The required trap depths of the waveguide for the same transversal confinement would be $2\pi \times (300, 220)$ kHz, corresponding to 11 μ K in the weak direction. The required optical power for the charger is a factor $(\sigma_{\text{WG}}^2 / \sigma_C^2)(300/11) \approx 700$ times lower, which is a big advantage. The comparably strong longitudinal confinement (aspect ratio of ~ 21) made it possible to robustly condense in a single beam. However the resulting big variation of the mean atom number per well in the lattice configuration made it difficult to obtain sufficient statistics for the unstable fixed point dynamics, which sensitively depends on the atom number.

In addition, in previous experiments, a different waveguide beam configuration with an aspect ratio of ~ 260 (waist of $\sim 60 \mu\text{m}$) was employed, which is in principle favorable for homogeneity, but it was difficult to reproducibly expand the initial BEC in it for loading of the lattice potential. Furthermore, the required trap depth for tight trapping is even higher. We therefore decided to slightly increase the beam size before the final focusing lens, leading to the compromise described above.

The calculation of the scattering rate for the waveguide beam at maximal power yields $\Gamma_{\text{sc}} \lesssim 45$ mHz and a heating rate of $\lesssim 3$ nK/s, calculated with the photon recoil energy $2\pi\hbar \times 2.2$ kHz for 1030 nm photons. Together with the heating from the lattice potential, this is in the same ballpark as previously observed heating rates in a similar configuration [64]. In the degenerate regime, the heat capacity depends on the temperature, so this is only a very rough estimate. Nevertheless, it shows, that it is imperative to minimize the holding time at high trapping frequencies. We observe trap lifetimes of ~ 15 s for the BECs in the hyperfine ground state, which compares well with $1/\Gamma_{\text{sc}}^{\text{tot}} \sim 17$ s. Thus, the heating rate is presumably lower than this worst case scenario.

3.4 Resonance Shifts

In this section, we provide details of the shifts of the two-photon transition, which directly influence the detuning δ of the spin system, namely shifts due to magnetic fields, the employed dipole traps and interaction shifts.

3.4.1 Magnetic field

Figure 3.2 summarizes the hyperfine structure in a DC magnetic field including the employed Rabi coupling and the resulting AC shifts.

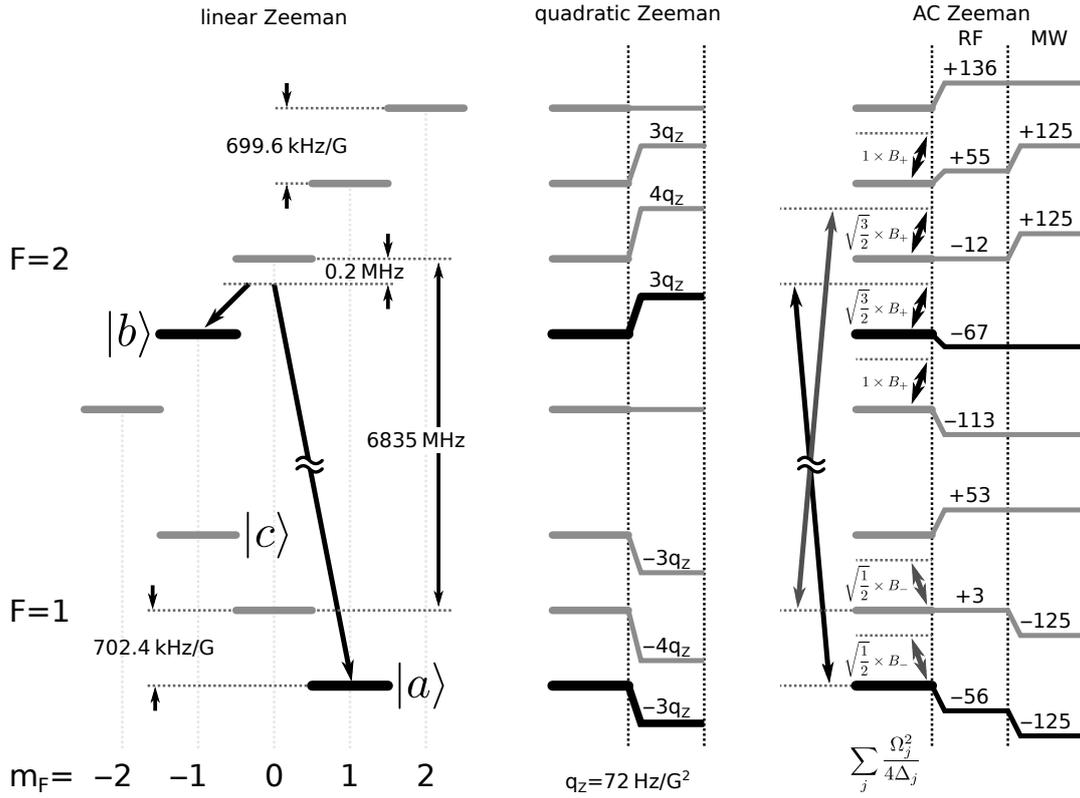


Figure 3.2 Hyperfine structure and magnetic field shifts. Here, the contributions to the hyperfine energy splittings of the $^{87}\text{Rb } 5S_{1/2}$ groundstate are schematically depicted. The employed qubit levels for the pseudospin are highlighted in dark black. **Left:** The linear Zeeman shift, which is the dominant contribution in absolute terms, contributes a small differential sensitivity of $\approx 2.8 \text{ kHz/G}$ of the two-photon resonance. **Middle:** The quadratic contribution of the Zeeman shift adds to this a differential sensitivity of $6 \times 72 \text{ Hz/G}^2 \times 9.13 \text{ G} \approx 7.9 \text{ kHz/G}$. The sum of linear and quadratic contributions gives a total sensitivity of 10.7 kHz/G at the working field. **Right:** The AC Zeeman shift of the RF field yields approximately an additional linear Zeeman splitting. Depicted is the case of equal contributions of the two circular components B_- and B_+ (see text) for a realistic RF Rabi frequency of 12 kHz on $|2, 0\rangle \leftrightarrow |2, -1\rangle$. The MW Rabi frequency was chosen as 10 kHz . One observes, that the MW AC shift is always increasing the transition frequency, whereas size and direction of the RF AC shift depends on the balance between B_+ and B_- .

DC magnetic fields The DC Zeeman shifts can be for small magnetic fields well approximated by a quadratic function

$$\nu_Z = \pm |p_Z| m_F B \pm |q_Z| (4 - m_F^2) B^2, \quad (3.8)$$

where the lower (upper) sign is for $F = 1$ ($F = 2$). The absolute value of p_Z is slightly different for $F = 1$ and $F = 2$ due to the small but nonzero nuclear g-factor and given in Fig. 3.2. More accurate values can be obtained with the Breit-Rabi formula [141] (see also Appendix A.4).

AC magnetic fields The AC Zeeman shifts are caused by the detuned coupling of all possible one-photon transitions to the states of the pseudospin system. Far from all resonances, a good approximation for these shifts is the incoherent sum of all contributions

$$\delta_{\text{AC}} = \sum_j \frac{\Omega_j^2}{4\Delta_j}. \quad (3.9)$$

Because the RF couples all Zeeman sublevels, a deeper understanding of these shifts requires a detailed look at the individual coupling strengths. In terms of spin-1 ($F = 1$) and spin-2 ($F = 2$) operators, the RF driving represents a spin rotation of the form [142, 143]

$$\hat{H}_{\text{cpl}} = \Delta_{\text{spin}} \hat{S}_z + |\Omega_{\pm}| \hat{S}_x, \quad (3.10)$$

where the coupling strengths are given by $\Omega_{\pm} = \sqrt{2}|g_F|\mu_B B_{\pm}$ and B_+ (B_-) is the amplitude of right-handed (left-handed) polarisation amplitude of the RF magnetic field with respect to the direction of the static offset field. μ_B is the Bohr magneton and g_F the magnetic g-factor. For example $B_+ = B_- = B_y/(2\sqrt{2})$ for linear polarisation along y . In our case, the coupling strength in $F = 1$ ($F = 2$) is proportional to B_- (B_+) and $|g_F|\mu_B$ in terms of frequencies is approximately $2\pi \times 0.7 \text{ MHz/G}$. The relation of the individual coupling strengths between the levels is therefore determined by the entries of the \hat{S}_x spin operators, which have exactly the same form as the collective spin operators \hat{J}_x for $N = 2$ (spin-1) and $N = 4$ (spin-2). These are given in Fig. 3.2 together with calculated shifts with the assumption of $B_- = B_+$ for a realistic RF Rabi frequency of $2\pi \times 12 \text{ kHz}$ on the $|2, 0\rangle \leftrightarrow |2, -1\rangle$ transition, relevant for the two-photon coupling. The action of the RF dressing is approximately analogous to an additional linear Zeeman shift. However the detailed balance of B_+ vs. B_- determines the resulting shift of the RF on the two-photon transition. The AC Zeeman shift of the MW magnetic field is more direct, since there is only one relevant transition reasonably close to resonance.

We directly characterized both contributions by Ramsey spectroscopy at 9.2 G keeping either the RF or the MW or none of them on during the interrogation time with the result $\delta_{\text{MW}} \approx 2\pi \times 120 \text{ Hz}$ and $\delta_{\text{RF}} \approx 2\pi \times 70 \text{ Hz}$ with the same sign, adding up to $\delta_{\text{AC}} \approx 2\pi \times 190 \text{ Hz}$ for the fastest pulses with a two-photon Rabi frequency of $\Omega \approx 2\pi \times 320 \text{ Hz}$. Apparently, the B_- component is about 60% stronger than B_+ for the employed loop coil antenna, presumably due to effects of induction in the surrounding coils and metal parts.

A calibration of the AC shifts was performed for each employed pulse power on a regular basis. It turned out, that the most sensitive measure for this purpose is the offset of small amplitude π and plasma oscillations (also used for the direct characterization of the nonlinearity, detailed below). We equalize both the offsets as well as the amplitudes of the atomic imbalance oscillations by adjusting the RF frequency. This procedure is reliable also in the presence of nonlinearity, where the frequency of the plasma(π)-oscillations is enhanced (reduced). In the Rabi regime ($\Omega > N\chi$), the offset gives the tilt of the rotation axis due to the detuning ($\propto \arctan(\delta/\Omega)$), which is the relevant parameter for a particular Ω .

3.4.2 AC electric fields of the dipole traps

The scalar AC Stark shifts from the red-detuned trapping laser beams are slightly larger for $F = 2$ than for $F = 1$. The resulting differential shift is approximately $-(\omega_{\text{hfs}}/\Delta)V_0$, where

ω_{hfs} is the hyperfine splitting, Δ the detuning from the optical resonance and V_0 the absolute value of the shift, which is close to the trap depth. We roughly estimate a shift of $\sim 2\pi \times 3$ Hz coming from the lattice and $\sim 2\pi \times 8$ Hz from the waveguide. At first sight, it seems surprising that the largely detuned waveguide has a bigger influence, but this is due to the total shift being about 18 times as large. Due to the power stabilization of the beams, which is better than 10^{-3} , these shifts have a negligible influence on the stability of the two-photon resonance.

The vectorial contribution to the AC Stark shift scales with the magnetic sensitivity of the states and the circular polarization component of the light and can thus be seen as an additional effective magnetic field [144]. The differential shift for the two-photon resonance is comparable in magnitude to the discussed scalar shift in the case of pure circular polarization [145]. Since we work with linearly polarized light for the trapping beams, this effect can be neglected.

3.4.3 Interaction shifts

One of the intrinsic effects of the BEC spin system is the interaction dependent resonance shift $\delta_{\text{int}} = (N - 1)(X_{bb} - X_{aa})$, which depends on the difference of the intraspecies interaction parameters. In our range of atom numbers it scales approximately with $\delta_{\text{int}} \propto \sqrt{N}$ (see Fig. 3.5 and discussion below). We determined its value by Ramsey experiments at a magnetic field of 9.2 G relatively far away from the Feshbach resonance, since a dependence on X_{ab} is not expected. The extracted value is $\delta_{\text{int}} = 2\pi \times -0.79(2)\sqrt{N}$ [130], which we verified with an independent characterization in the presence of significant nonlinearity (see below). The interaction shift is quite substantial; at $N_0 = 500$, a change of 60 atoms corresponds to a shift of about 1 Hz. The loss of atoms during a time evolution thus leads to considerable detuning drift. Furthermore, a postselection based on the sum of both components is mandatory to obtain consistent results.

3.5 Characterization of the interaction parameters

The most versatile measurement for the characterization of the interaction effects in the spatial single mode approximation are plasma- and π -oscillations⁹ (see 2.1.8) in the Rabi regime of dominant coupling ($\Lambda = N\chi/\Omega \ll 1$). These are initiated by a preparation pulse of angle $\pi/2 - \epsilon$ with pulse phase $\phi = 0$, followed by a sudden change of the RF coupling phase by $\pi/2$ for plasma or $3\pi/2$ for π -oscillations. Figure 3.3 depicts this procedure schematically. The finite preparation imbalance $z < 0$ results in small amplitude oscillations which reflect remarkably many features of the interacting two-mode system. In the limit of small amplitudes, we obtain the following approximations for their frequencies:

$$\begin{aligned}\omega_{\text{pl}} &= \Omega\sqrt{1 + \Lambda} = \Omega \left(1 + \frac{\Lambda}{2} - \frac{\Lambda^2}{8} + \mathcal{O}(\Lambda^3) \right) \\ \omega_{\pi} &= \Omega\sqrt{1 - \Lambda} = \Omega \left(1 - \frac{\Lambda}{2} - \frac{\Lambda^2}{8} + \mathcal{O}(\Lambda^3) \right)\end{aligned}\quad (3.11)$$

⁹The naming refers to the atomic phase φ between the two components being zero or π .

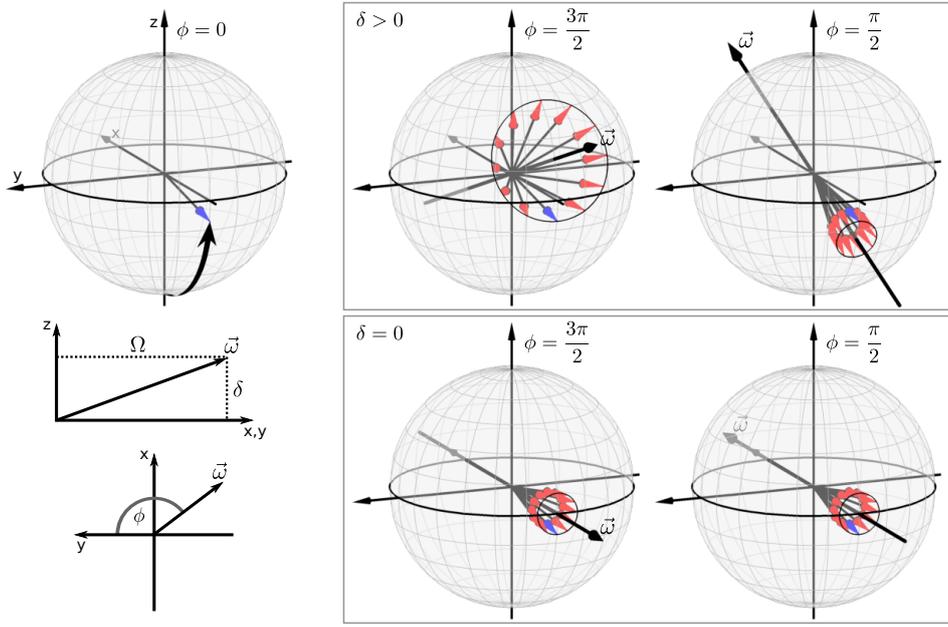


Figure 3.3 Scheme of plasma- and π -oscillations in the Rabi regime Here, small amplitude oscillations are depicted for dominating Rabi coupling ($\Omega > N\chi$). They are initiated by a preparation with a $\frac{\pi}{2} - \epsilon$ pulse (with reference phase $\phi = 0$) followed by a nonadiabatic change of the coupling phase by $3\pi/2$ for π -oscillations (mean atomic phase $\langle \varphi \rangle = \pi$) or $\pi/2$ for plasma-oscillations (mean atomic phase $\langle \varphi \rangle = 0$). The top row shows the preparation with a mean spin vector below the equator of the Bloch sphere, and the subsequent rotation of the mean spin in the two cases in the presence of a detuning (here exaggerated $\delta/\Omega \sim 0.6$). The effective rotation axis $\vec{\omega}$ is tilted with respect to the equatorial plane by an angle $\sim \arctan(\delta/\Omega)$. The resulting oscillations have differing amplitudes and offsets $\langle z(t) \rangle$ along the projection axis z . For $\delta = 0$ (lower panel), the oscillations share the same offset and amplitude, however their frequencies differ in the presence of nonlinearity, which is used for its characterization. The detuned behavior is a sensitive means to verify the compensation of AC Zeeman shifts in the presence of the full coupling power, which cannot be accomplished with a Ramsey sequence.

and thus for their mean and difference frequency

$$\begin{aligned} (\omega_{\text{pl}} + \omega_{\pi})/2 &\approx \Omega \\ \omega_{\text{pl}} - \omega_{\pi} &\approx \Omega\Lambda = N\chi. \end{aligned} \quad (3.12)$$

Note, that for the nonlinearity χ , the approximation is to second order. A further interesting quantity is the offset $\langle z(t) \rangle$ of the oscillations, which reflects the position of the fixed point in the dynamics of the mean and can be obtained by the requirement $(\dot{z}, \dot{\varphi}) = 0$ from a linearization of the equations of motion (Eq. 2.48) which yields

$$\begin{aligned} z_{\text{off,pl}} &\approx -\frac{\delta}{\Omega(1+\Lambda)} \approx -\frac{\delta}{\Omega}(1-\Lambda) \\ z_{\text{off,\pi}} &\approx \frac{\delta}{\Omega(1-\Lambda)} \approx \frac{\delta}{\Omega}(1+\Lambda) \end{aligned} \quad (3.13)$$

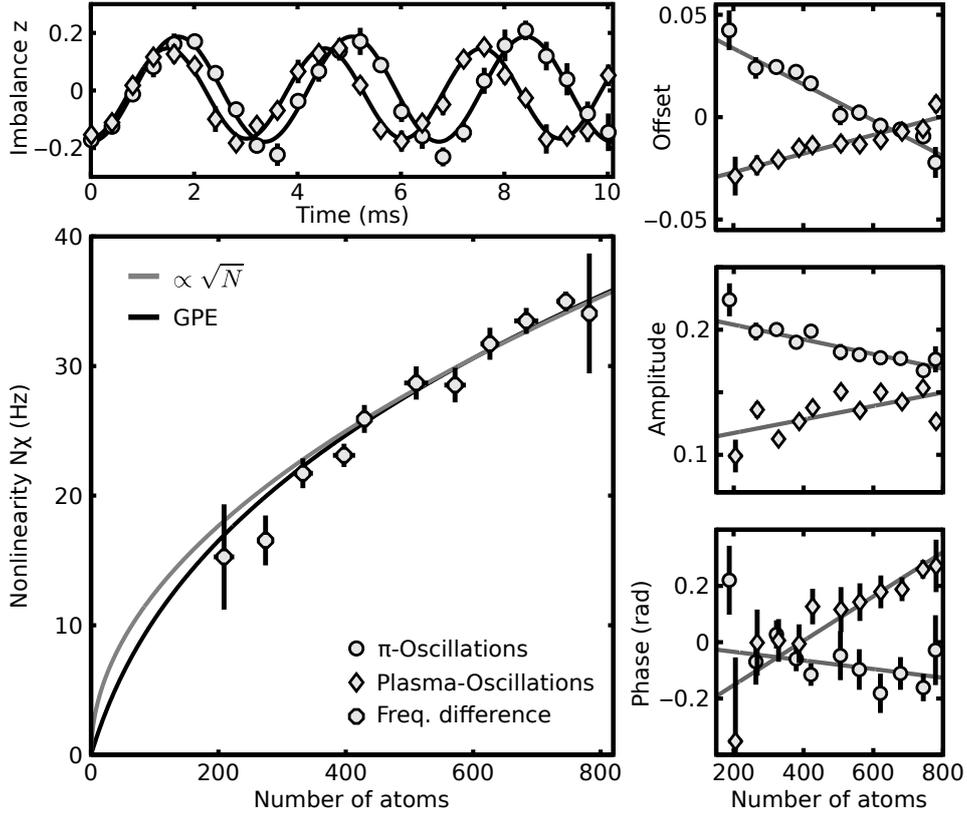


Figure 3.4 Characterization of the nonlinearity Here, we show the result of small amplitude plasma- and π -oscillations in the Rabi regime with $\Omega \approx 2\pi \times 320$ Hz and an evolution time up to 10 ms at our working magnetic field of 9.13 G. **Left top:** Exemplary measurement in a single lattice well with a mean atom number of 510, averaged over 6 realizations. The deviations of the oscillation frequencies are a direct measure of the nonlinearity due to atomic interactions. **Left bottom:** Difference frequency as extracted from single-well measurements and binned by atom number. Additionally, we show the expected scaling with the atom number from a numerical solution of the stationary Gross-Pitaevskii equation (GPE) for our trapping parameters. The empirical fit of a simple square root describes very well the behavior in the relevant range of atom numbers with the result $N\chi = 2\pi \times 1.25(2) \sqrt{N}$ Hz. From the GPE result, we can estimate the change in scattering length $\Delta a_{ab} \approx 6.4a_B$ due to the Feshbach resonance. **Right:** The three parameters offset, amplitude and phase of the oscillations. The detuning changes with the atom number as can be seen in the behavior of offset and amplitude. For a precise extraction of the frequencies, offset and phase are fixed to the value of the linear fit (see text).

and thus $(z_{\text{off},\pi} - z_{\text{off},\text{pl}})/2 \approx \delta/\Omega$. At fixed external detuning, the scaling of δ with the atom number gives the interaction shift.

Figure 3.4 shows a corresponding measurement at a magnetic field of 9.13 G. The measurement of the time evolutions were repeated six times for two different evaporation depths of the initial BEC to obtain a large range of atom numbers and good statistics.

The extraction of the nonlinearity was done by first fitting the oscillations sinusoidally for each well and measurement with the free parameters amplitude, offset, phase and frequency. With the help of the systematic behavior of phase and offset versus atom number, the fit was repeated with those values fixed to the linear approximations shown in Fig. 3.4 in order to

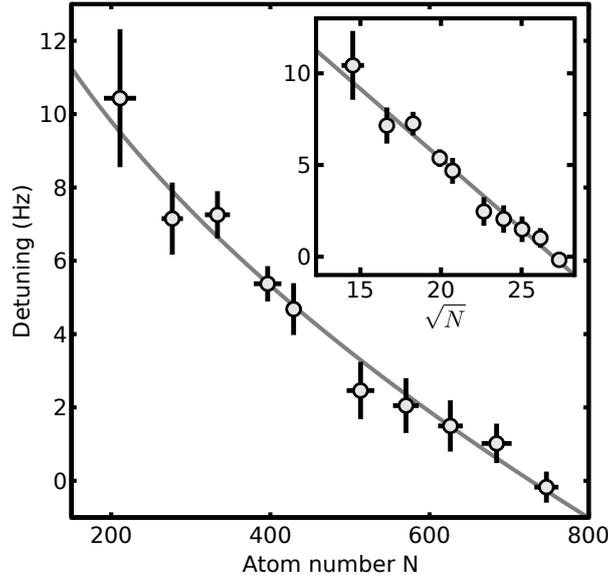


Figure 3.5 Interaction shift of the two-photon resonance From the same measurement as Fig. 3.4, the interaction shift of the resonance is extracted from the offset of the oscillations. It shows nice agreement to the expected scaling with the fit result $\delta_{\text{int}} = 2\pi \times -0.77(6)\sqrt{N}$ Hz. Note that the characterization is done in the presence of nonlinearity directly at 9.13 G and agrees very nicely with the result $2\pi \times -0.79(2)\sqrt{N}$ Hz obtained from Ramsey experiments at 9.2 G, farther away from the Feshbach resonance.

obtain a better estimate of the frequencies. The obtained frequencies were subtracted and binned by mean atom number. In each bin, a weighted mean was calculated, taking into account the RMS fit errors $\Delta\omega$ by weighting with $1/(\Delta\omega_{\text{pl}}^2 + \Delta\omega_{\pi}^2)$.

Both nonlinearity χ and interaction detuning δ_{int} depend on the shape of the BEC mean field (wavefunction) $\psi(\mathbf{x})$ of the two modes on the single lattice site as well as on the relevant interaction parameters g_{ij} . The expected behavior in the single spatial mode approximation is

$$\begin{pmatrix} N\chi \\ \delta_{\text{int}} \end{pmatrix} \approx \frac{N}{2} \begin{pmatrix} g_{aa} + g_{bb} - 2g_{ab} \\ g_{bb} - g_{aa} \end{pmatrix} \int |\psi|^4 d^3x, \quad (3.14)$$

where g_{ab} is altered close to the Feshbach resonance.

In Fig. 3.4 we compare the experimental values for $N\chi$ with the expected scaling from the Gross-Pitaevskii self-consistent groundstate solution for ψ obtained with a single component of scattering length $100a_B$. The prefactor fitting best to the observations is $g_{aa} + g_{bb} - 2g_{ab} \approx 2\pi \times 0.98 \text{ Hz } \mu\text{m}^3$, from which we estimate a change of the interspecies scattering length of $\Delta a_{ab} \approx 6.4a_B$ due to the Feshbach resonance. An empirical square-root fit gives a very good approximation for the relevant range of atom numbers with the result $N\chi = 2\pi \times 1.25(2)\sqrt{N}$ Hz.

Figure 3.5 shows the corresponding extraction of the interaction shift using the differences in offset, multiplied with the local Rabi frequency for each well. The result again fits very nicely to the \sqrt{N} -behavior with the result $\delta_{\text{int}} = 2\pi \times -0.77(6)\sqrt{N}$ Hz in very good agreement with the result of the Ramsey experiments at 9.2 G.

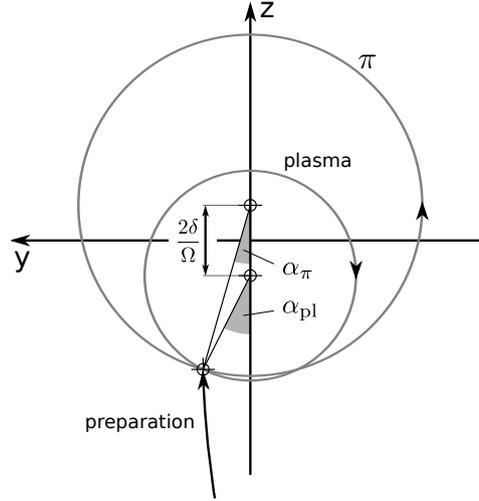


Figure 3.6 Influence of preparation error and detuning Here, a projection of the π - and plasma-oscillations to the y - z -plane (viewed from the negative x -axis) is schematically depicted for $\delta > 0$, as is the case for the calibration measurement. The initial y -position is influenced by both nonlinearity as well as detuning during the preparation pulse, which leads to a phase offset of the subsequent oscillations, which is different in magnitude and sign for the two types. The detuning shifts their offsets in opposite directions, which also makes their amplitudes differ. The effects are exaggerated here for clarity.

The behavior of the phase of the oscillations can be understood as combined effect of nonlinearity and detuning during the rotation about the y -axis for preparation. The nonlinearity results in a phase evolution which linearly depends on z , whereas the detuning contributes proportional to the mean spin length in x -direction, which builds up due to the rotation. The linearized equation of motion for y in this case reads $\dot{y} \approx N\chi z(t)x(t) - \delta x(t)$, where $z(t) \approx -\cos(\Omega t)$ and $x(t) \approx -\sin(\Omega t)$. This can be easily integrated, which yields

$$y(t) \approx \frac{N\chi}{2\Omega} \sin^2(\Omega t) + \frac{\delta}{\Omega} [\cos(\Omega t) - 1]. \quad (3.15)$$

With $\Omega t \approx \pi/2$ of the preparation pulse, one gets $y \approx (N\chi/2 - \delta)/\Omega = \Lambda/2 - \tilde{\delta}$ and thus an atom number dependence of y after the preparation. The situation is schematically depicted in Fig. 3.6. Since $\chi > 0$ and $\delta_{\text{int}} < 0$ in our case, linear and nonlinear interaction contributions add up and a compensation can be only achieved for one specific atom number by an external detuning or a phase adjustment of the second pulse. For the calibration measurement, the preparation pulse was ideal (i.e. $y \approx 0$ after the pulse) at atom numbers around 220, although the atom number 'most resonant' was ≈ 730 (see Fig. 3.5). This reflects in the phase of the oscillations (Fig. 3.4) becoming equal for low atom numbers and increasingly different for higher numbers, although the coupling comes closer to resonance.

It is interesting to note that due to the closely matching atom number dependencies of $N\chi$ and δ , the phase error of the preparation pulse can in principle be minimized by a suitable choice of the interspecies interaction parameter¹⁰, such that the linear interaction part compensates the nonlinear one for a broad range of atom numbers.

¹⁰This could be achieved with $g_{ab} = (3g_{aa} - g_{bb})/2$. Since $g_{aa} > g_{bb}$, this would correspond to a required increase of g_{ab} and thus to a position on the lower-field immiscible side of the Feshbach resonance. However,

3.6 Gradients of the parameters

The advantage of our BEC array besides the tight trapping in lattice direction is that each experimental run yields several BECs with similar atom numbers. In order to interpret them as independent realizations of the same experiment, it is important, that the relevant physical parameters are as homogeneous as possible along the lattice (x -direction). The possible deleterious effects can be divided into influence of the trapping, which potentially alters nonlinearity and detuning, the gradient of the offset magnetic field, which gives a spatial variation of the detuning, and the inhomogeneity of the coupling fields, which can contribute rotation angle errors and detuning through the AC Zeeman shift.

3.6.1 Trapping

The influence of the trapping on interaction effects can be estimated by noting that the relevant interaction parameter of a harmonically trapped BEC is given by Na/a_{ho} [146], where $a_{\text{ho}} = \sqrt{\hbar/m\omega_{\text{ho}}}$ is the harmonic oscillator length and $\omega_{\text{ho}} = (\omega_x\omega_y\omega_z)^{1/3}$ is the geometric mean of the trapping frequencies. We have seen that the interaction dependent parameters scale with the interaction energy per particle $E_{\text{int}}/N \propto Na \int |\psi|^4 d^3x \simeq \alpha\sqrt{N}$. For nonlinearity and interaction detuning, a has to be substituted with the relevant scattering length combinations. Due to the universality of the interaction parameter, the scaling in units of the mean harmonic oscillator can be obtained from the scaling with the atom number [146]. This yields

$$\frac{E_{\text{int}}}{N} \propto \frac{\omega_{\text{ho}}}{2} \sqrt{N\sqrt{\omega_{\text{ho}}}} \propto \sqrt{N}\omega_{\text{ho}}^{5/4} = \sqrt{N}\omega_{\perp}^{5/6}\omega_{\text{lat}}^{5/12} \approx \sqrt{N}\omega_{\perp}\sqrt{\omega_{\text{lat}}}. \quad (3.16)$$

In the Thomas-Fermi limit, which neglects the kinetic energy over the interaction and trapping contribution, the corresponding quantity would be $E_{\text{int}}/N = (2/7)\mu = \hbar\omega_{\text{ho}}/7(15Na/a_{\text{ho}})^{2/5} \propto N^{2/5}\omega_{\text{ho}}^{6/5}$, which yields almost the same scaling with the trap frequencies¹¹. From Eq. 3.16, we see that the waveguide trapping frequency has an approximately linear influence, whereas the lattice potential is less critical. The inhomogeneity of the trapping frequencies is $\Delta\omega_{\text{lat}}/\omega_{\text{lat}} < 1.5\%$ and $\Delta\omega_{\perp}/\omega_{\perp} < 1\%$ (see Appendix B.1), which leaves an influence of $< 2\%$ on the interaction effects. While negligible for the nonlinearity χ , the influence on the detuning is noteworthy, since the total interaction shift is $\sim 2\pi \times 20$ Hz and the 2% turn into $2\pi \times 0.4$ Hz total shift.

Due to the hyperfine splitting, there is also a direct effect of the trapping on detuning, which is proportional to the optical power (see section 3.4.2). From the above consideration of the differential AC Stark shifts, we estimate an influence smaller than $2\pi \times (2\Delta\omega_{\text{lat}}/\omega_{\text{lat}} \times 3 \text{ Hz} + 2\Delta\omega_{\perp}/\omega_{\perp} \times 8 \text{ Hz}) \approx 2\pi \times 0.25$ Hz.

We note that these are rather pessimistic estimates, but showcase the delicacy of the effects (we aim for a homogeneity and stability of better than 1 Hz) and the advantage of direct characterizations and calibrations via Ramsey spectroscopy and π /plasma-oscillations.

we know from test experiments on this side aiming for best spin squeezing, that it is not favorable for the generation of nonclassical states to increase g_{ab} with our trapping parameters, presumably because the mode overlap decreases. For tighter traps, this could be an option.

¹¹The scaling with the atom number in Thomas-Fermi approximation is experimentally indistinguishable from a \sqrt{N} -behavior. However, since $a \approx 5.3$ nm and $a_{\text{ho}} \approx 570$ nm in our case, the requirement $Na/a_{\text{ho}} \gg 1$ for Thomas-Fermi is only valid for $N \gg 100$.

3.6.2 Magnetic offset field

The magnetic sensitivity 10.7 Hz/mG of the two-photon resonance implies that the gradient of the magnetic field B_0 has to be smaller than about $0.3 \mu\text{G}/\mu\text{m}$ for a detuning homogeneity better than 0.5 Hz over 30 sites of the BEC array. Due to the relatively large value of B_0 , we are only sensitive in its direction and perpendicular components are suppressed by a factor $2B_0$, since the different field directions sum quadratically to the absolute value. We optimized the gradient with the help of Ramsey experiments with long interrogation time on the maximally Zeeman sensitive transition $|1, 1\rangle \leftrightarrow |2, 2\rangle$. Due to the magnetic field gradient, the different lattice sites eventually dephase and the atomic imbalance after the second $\pi/2$ -pulse shows a periodic pattern along the array. Due to magnetic field fluctuations, the absolute position of this pattern is not reproducible for long interrogation times, whereas the periodicity is stable, reflecting the magnetic field gradient. This proved very handy for optimization, since any experimental change of the gradient also comes with a change in B_0 . We optimized by placing small permanent refrigerator magnets close to the experimental chamber towards increasing period of the pattern. Finally, we were limited by the interaction shift on this transition, which changes the behavior of the pattern from spatially periodic to atom number dependent. With a careful analysis of the scaling of the differential Ramsey signal along the lattice for small interrogation times¹², we measured the final gradient to $0.196(6) \mu\text{G}/\mu\text{m}$ [12, 130]. This means that the inhomogeneity over 30 wells ($\sim 30 \mu\text{G}$) is on the order of the shot-to-shot reproducibility of B_0 .

3.6.3 Coupling fields

The influence of gradients of the Rabi coupling fields is twofold. On the one hand, they lead to a variation of the rotation angles for fast rotations of fixed pulse-length or of the parameter $\Lambda = N\chi/\Omega$ via a change in Ω . More importantly, they also cause an inhomogeneity of the detuning via the AC Zeeman shift. To characterize these effects, we performed Rabi flopping with many cycles either on the one-photon microwave transition or on the two-photon transition. In each case, the gradient is revealed by sinusoidally fitting the imbalance results of each lattice well and extracting the spatial dependence of the Rabi frequency by a linear fit.

The gradient of the microwave coupling field is quite sensitive to the position of the antenna loop. For optimization, we chose the Rabi pulse time on the one-photon transition sufficiently long, such that a dephasing along the lattice was apparent in a single experimental run and maximized the period along the lattice by small displacements of the antenna. The disadvantage is that this changes also the Rabi frequency which makes the period maximization for fixed Rabi pulse time ambiguous and necessitates several recalibrations of the Rabi frequency during optimization. The final gradient was $\Delta\Omega_{\text{MW}}/\Omega_{\text{MW}} \approx 1.4\%/165 \mu\text{m}$ (30 wells) at $\Omega_{\text{MW}} \approx 2\pi \times 10.0 \text{ kHz}$.

For the radio frequency we tried different configurations and sizes as well as distances of loop coils¹³ and ended up with the present configuration (see Fig. 3.1 RF 1 and Appendix B.3) with the loop around the imaging objective as big as possible to not obscure the MOT

¹²The employed sequence contained a swapping of a spin-squeezed state to the magnetically sensitive transition $|1, 1\rangle \leftrightarrow |1, -1\rangle$ with negligible interaction shift.

¹³This includes an unsuccessful try of two loops simultaneously on opposite sides of the glass cell

beams. Combined with the microwave, this yielded a gradient of the two-photon coupling of $\Delta\Omega/\Omega \approx 0.53\%/165\mu\text{m}^{14}$ at $\Omega \approx 2\pi \times 344\text{ Hz}$, that is at an RF Rabi frequency of $\Omega_{\text{RF}} \approx 2\pi \times 13.8\text{ kHz}$. This means that the RF gradient partly compensates the MW gradient. Since $\Delta\Omega/\Omega \approx \Delta\Omega_{\text{MW}}/\Omega_{\text{MW}} + \Delta\Omega_{\text{RF}}/\Omega_{\text{RF}}$, we get $\Delta\Omega_{\text{RF}}/\Omega_{\text{RF}} \approx -0.38\%/165\mu\text{m}$.

We note that the relative gradient of the two-photon coupling does not depend on the relation $\Omega_{\text{MW}}/\Omega_{\text{RF}}$, since Ω is directly proportional to both contributions.

The relative gradient value of the two-photon coupling directly translates into angle inhomogeneities, for example a worst case of $\approx 0.5^\circ$ for a 90° rotation ($\pi/2$ -pulse)¹⁵.

Since the AC Zeeman shifts are proportional to the square of the one-photon Rabi frequencies, they are not coupled to the two-photon Rabi frequency, but depend on the strength of the individual coupling fields. In light of the Rabi gradients, this is most important for the regime $\Omega < N\chi$ during state evolution, which requires $\Omega \lesssim 2\pi \times 20\text{ Hz}$. This can be either reached by attenuating the RF or the MW or both, the total attenuation required from $\Omega_0 \approx 2\pi \times 320\text{ Hz}$ being $\gtrsim 24\text{ dB}$. If attenuating only one component, the AC Zeeman shift gradient of the other component remains, which gets significant, since the important parameter is δ/Ω . For example, if only the RF is attenuated, which is easy to implement with the arbitrary waveform generator, a sizable gradient of $\approx 2\Delta\Omega_{\text{MW}}/\Omega_{\text{MW}}\delta_{\text{MW}} \approx 3\%/165\mu\text{m} \times 2\pi \times 120\text{ Hz} \approx 2\pi \times 4\text{ Hz}/165\mu\text{m}$ would remain, while the shift due to the RF would become negligibly small. We therefore take the additional complication¹⁶ of lowering the microwave coupling by 10.8 dB with the help of a fixed attenuator on a fast RF switch for this purpose (for details see Appendix B.3). The attenuation leads to a drastic reduction of δ_{MW} to $\approx 2\pi \times 10\text{ Hz}$ and of its gradient to $\approx 2\pi \times 0.3\text{ Hz}/165\mu\text{m}$. The additionally employed attenuation of the RF by 14 dB makes its AC Zeeman shift contribution to the detuning gradient negligible ($\delta_{\text{RF}} \approx 2\pi \times 3\text{ Hz}$).

3.7 Atom loss

The loss of atoms from the trap is a severe limitation for the creation of highly nonclassical states in our pseudospin implementation. The most relevant loss channels are two-body loss [147] in $F = 2$ and losses due to the Feshbach resonance [148]. Here, we discuss and characterize these two contributions. All other loss channels, for example due to photon scattering from the dipole traps or collisions with background atoms in the vacuum chamber happen on a timescale of $\sim 15\text{ s}$ and are thus negligible for the spin evolution times $\lesssim 30\text{ ms}$. For the loss characterization, we have prepared different initial imbalances with a two-photon Rabi coupling pulse followed by a variable hold time of up to 70 ms at our working field 9.13 G close to the Feshbach resonance.

¹⁴This corresponds to a reduction by an order of magnitude from the gradient we had with the older antenna RF 2 (Fig. 3.1)

¹⁵On average, these effects are smaller, since similar atom numbers are usually spaced by much less than 20 wells. However, the effect systematically increases for a post-selection of smaller atom numbers, since they are located at the edge of the BEC array.

¹⁶The attenuated and non-attenuated paths have a different phase shift, which we directly calibrated with π /plasma-oscillations and compensated with a corresponding RF phase shift.

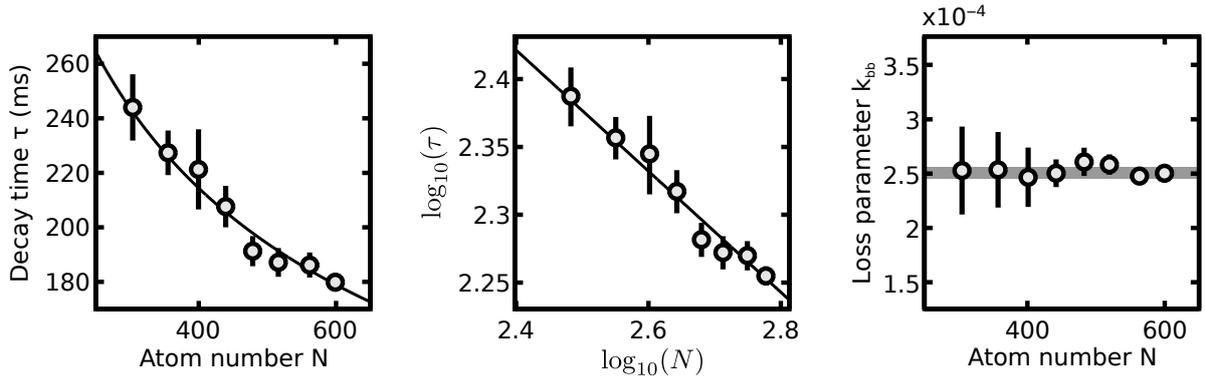


Figure 3.7 Two-body spin relaxation loss Here, all atoms are prepared in $|b\rangle = |F = 2, m_F = -1\rangle$ and kept for different hold times up to 70 ms in the trap. The resulting decay is fitted for each lattice well and the results binned and averaged for the initial atom number. The result of an exponential fit (left) shows a characteristic scaling with the initial atom number. A linear fit in a double-logarithmic scale (middle) gives the result $\tau \propto N^{-0.45(4)}$ (also line in left plot), consistent with the expectation $1/\sqrt{N}$. A fit to Eq. 3.19 yields the loss parameter $k_{bb} = 2.51(5) \cdot 10^{-4} \text{ms}^{-1}$ (gray error band), which is to good approximation independent of the initial atom number.

3.7.1 Two-body spin relaxation

For the collisional interaction of two atoms of $|b\rangle = |F = 2, m_F = -1\rangle$, the total spin projection $M_F = m_F^{(1)} + m_F^{(2)} = -2$ is a conserved quantity in a small magnetic field, while this is not the case for the total F quantum number. During the collision one of the two or both atoms can relax to $F = 1$, which releases the hyperfine energy splitting of 6.8 GHz and is in either case sufficient for both atoms to escape the trap. In absence of component $|a\rangle$, we model this loss process [147, 149] with

$$\frac{dn_b(t)}{dt} = -K_{bb}n_b(t)^2, \quad (3.17)$$

where $n_b(t)$ is the density of component $|b\rangle$. In the single spatial mode approximation, we can integrate out the spatial degrees of freedom and obtain

$$\frac{dN_b(t)}{dt} = -K_{bb}N_b(t)^2 \int |\psi|^4 d^3x \approx -k_{bb}N_b(t)^{3/2}, \quad (3.18)$$

where we have used our approximation $\int |\psi|^4 \propto 1/\sqrt{N}$. The Thomas-Fermi scaling would be $\propto N^{-3/5}$. By separation of variables, this differential equation can be integrated with the result

$$N_b(t) = \left(\frac{k_{bb}}{2}t + N_b(0)^{-1/2} \right)^{-2}. \quad (3.19)$$

Experimentally, the initial decay cannot be distinguished from an exponential law, but we expect the decay time to have a scaling of $\tau \propto 1/k_{bb}\sqrt{N_b(0)}$. We therefore fit the data for the imbalance $z \approx 1$ with both alternatives. Figure 3.7 shows the results. The exponential fit yields the scaling $\tau \propto N_b(0)^{-0.45(4)}$, which is consistent with our expectation. The fit to Eq. 3.19 yields the parameter $k_{bb} = 0.251(5) \text{s}^{-1}$, which turns out to be a good description for all initial atom numbers. Including the GPE result $\sqrt{N} \int |\psi|^4 d^3x = 2.60(4) \mu\text{m}^{-3}$ we obtain

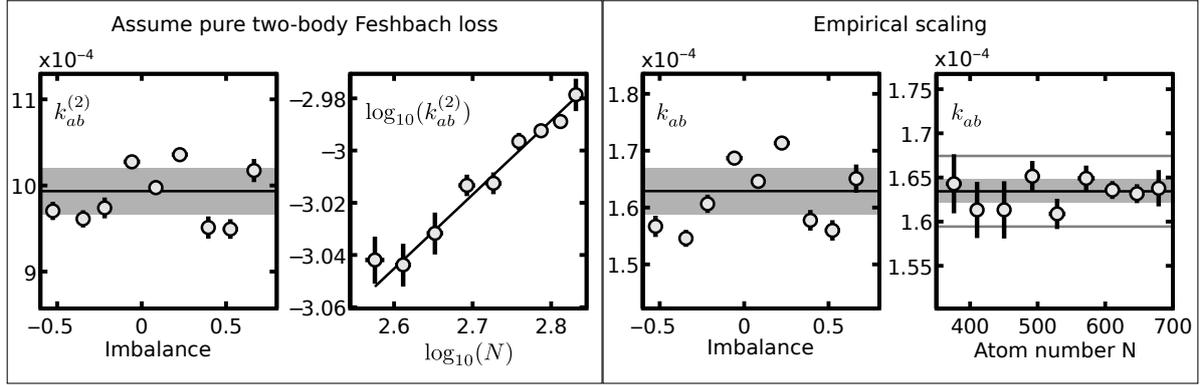


Figure 3.8 Losses due to the proximity of the Feshbach resonance The preparation of different initial imbalances between $|a\rangle = |F = 1, m_F = 1\rangle$ and $|b\rangle = |F = 2, m_F = -1\rangle$ systematically varies the strength of the additional loss channel due to the Feshbach resonance. Including the result of the two-body spin relaxation loss of $F = 2$, the numerical solution of the coupled differential equations for N_a and N_b are fitted to the experimental results. **Left:** The assumption of additional two-body loss already captures quite well the imbalance dependence. However, the resulting loss parameter $k_{ab}^{(2)}$ shows a clear scaling with the initial total atom number ($\propto N(0)^{0.28(2)}$, linear fit in the log-plot), which is a strong indication of other loss channels and dismisses a pure two-body Feshbach loss mechanism. **Right:** Including an additional scaling of $\propto N^{-0.3}$ in the differential equations yields a nice description with smaller relative deviations vs. imbalance and no significant dependence on the initial atom number with the result $k_{ab} = 1.63(4) \cdot 10^{-4} \text{ ms}^{-1}$ (given with one standard deviation of its mean vs. imbalance, which is larger as vs. initial atom number, depicted as gray band and gray lines in the rightmost panel). The empirical scaling lies in between the expected scaling of pure two-body and pure three-body loss (see text).

$K_{bb} = 9.7(3) \cdot 10^{-14} \text{ cm}^3 \text{ s}^{-1}$ in agreement with the result $(10.4 \pm 1) \cdot 10^{-14} \text{ cm}^3 \text{ s}^{-1}$ of [147], which was obtained at a magnetic field of 3 G.

3.7.2 Feshbach loss

The loss mechanism due to the Feshbach resonance has been considered in [148] as additional two-body loss channel taking the form

$$\begin{aligned} \frac{dn_a(t)}{dt} &= -K_{ab}^{(2)} n_a(t) n_b(t) \\ \frac{dn_b(t)}{dt} &= -K_{ab}^{(2)} n_a(t) n_b(t) - K_{bb} n_b(t)^2 \end{aligned} \quad (3.20)$$

From the above consideration of two-body loss, we can write

$$\begin{aligned} \frac{dN_a}{dt} &\approx [-k_{ab}^{(2)} N_a N_b] N^{-1/2} \\ \frac{dN_b}{dt} &\approx [-k_{ab}^{(2)} N_a N_b - k_{bb} N_b^2] N^{-1/2} \end{aligned} \quad (3.21)$$

Here, we have assumed, that the single spatial mode does not depend on the relative populations, but only on the total atom number $N = N_a + N_b$ and we have omitted the notation of the time

dependence for clarity. This coupled system of equations cannot be easily solved analytically, but a numerical solution can be obtained quite efficiently. With an optimization algorithm, we fitted its solution to the time dependence of N_a and N_b independently for each lattice well and measurement by minimizing the squared deviations to the experimental data. The value for k_{bb} was fixed to the value obtained above. The results for $k_{ab}^{(2)}$ were binned for averaging either by initial imbalance or by initial atom number. The range of the other parameter was chosen in the range of statistical relevance ($N(0) = 350 \dots 700$ and $z(0) = -0.6 \dots 0.7$ respectively). The results are shown in Fig. 3.8 left. While the imbalance dependence is captured quite reasonably, we observe a clear scaling of $k_{ab}^{(2)}$ with $N(0)^{0.28(2)}$, which points at an additional three-body loss channel.

The consideration of three-body loss is motivated by the notion of molecule formation due to the coupling to the molecular states involved in the Feshbach resonance [150]. For molecule formation, three atoms have to come together such that the momentum conservation can be fulfilled and the binding energy can go into the relative motion of molecule and third particle. For the loss rates, we would write down

$$\begin{aligned}\frac{dn_a}{dt} &= -K_{ab}^{(3)}(n_a^2 n_b + n_a n_b^2) = -K_{ab}^{(3)} n_a n_b (n_a + n_b) \\ \frac{dn_b}{dt} &= \frac{dn_a}{dt} - K_{bb} n_b^2\end{aligned}\quad (3.22)$$

or in terms of the atom numbers

$$\begin{aligned}\frac{dN_a}{dt} &\approx -K_{ab}^{(3)} N_a N_b N \int |\psi|^6 d^3x \approx -k_{ab}^{(3)} N_a N_b \\ \frac{dN_b}{dt} &\approx \frac{dN_a}{dt} - k_{bb} N_b^2 N^{-1/2}\end{aligned}\quad (3.23)$$

Here, we have used $\int |\psi|^6 d^3x \propto N^{-1}$, which is a good approximation to the GPE result¹⁷ in our range of atom numbers. The corresponding Thomas-Fermi scaling is $N^{-6/5} = N^{-1.2}$. However, the best description of the experimental data is found with

$$\begin{aligned}\frac{dN_a}{dt} &= -k_{ab} N_a N_b N^{-0.2} \\ \frac{dN_b}{dt} &= \frac{dN_a}{dt} - k_{bb} N_b^2 N^{-1/2},\end{aligned}\quad (3.24)$$

which was anticipated from the observation $k_{ab}^{(2)} \propto N^{0.3}$ from the assumption of two-body loss. The results are shown in Fig. 3.8 right. The relative deviations from the mean parameter value vs. imbalance are reduced compared to the two-body loss assumption, which is caused by the good description of the dependence on initial atom number. We find a parameter $k_{ab} = 0.163(4) \text{ s}^{-1}$, where the given error is the standard deviation from the mean vs. imbalance. If the Thomas-Fermi approximation would be justified, the observed scaling would be perfectly consistent with pure three-body loss, since the relevant term is $\propto N_a N_b N^{-0.2} = N_a N_b N N^{-1.2}$ (see Eq. 3.23). However, the scaling of the GPE solution is weaker ($\propto N^{-1}$ instead of $\propto N^{-1.2}$). As we have seen, two-body loss is expected to scale with $\propto N_a N_b N^{-0.5}$, which has the same

¹⁷The numerical value is $N \int |\psi|^6 d^3x \approx 9.1(1) \mu\text{m}^{-6}$

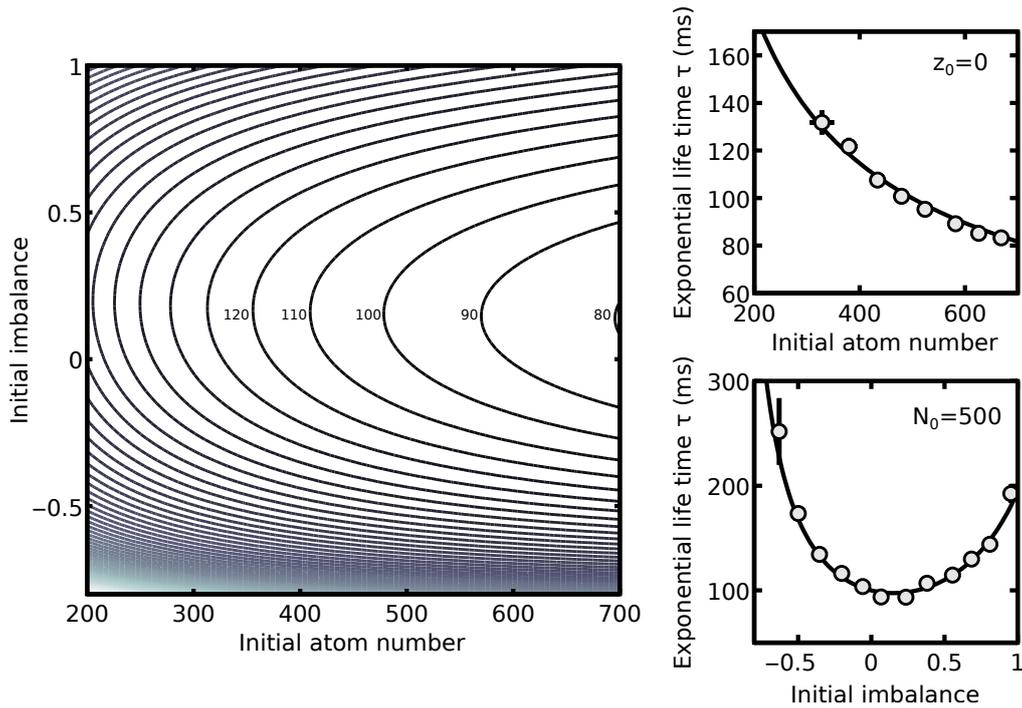


Figure 3.9 Exponential lifetime for different initial conditions Here, we show the dependence of the exponential lifetime of the total number of atoms on the initial imbalance and on the initial atom number. The lines in the left panel are contours of equal lifetime, spaced by 10 ms, beginning with 80 ms on the right edge. The right column shows two cuts, one for an initial $z \approx 0$ vs. atom number and one for an initial atom number of $N \approx 500$ vs. initial imbalance. Due to the spin relaxation loss in $F = 2$, the behavior vs. imbalance is asymmetric with respect to $z = 0$. The shown experimental data points are the results of exponential fits for evolution times up to 70 ms.

form and only differs by the scaling with the total atom number. One could suspect, that the difference is caused by the fast timescale of the loss [132], such that the picture, that the shape of the wave function can adiabatically follow during the time evolution is not justified and therefore a smaller scaling of two-body loss is plausible. However, the time scale of atom loss is ~ 100 ms, whereas the mean harmonic oscillator time scale is ~ 3 ms with tight trapping in all directions, such that non-adiabaticity can be neglected. We checked the validity of this rough argument with a time dependent GPE and found the deviations from the given scalings to be on the order of 10^{-3} during the time evolution.

In this respect we can conclude, that we observe a clear tendency towards three-body Feshbach loss, while the admixture of additional two-body loss cannot be discriminated in our range of atom numbers. Although we made rather simplified assumptions in view of the complex nature of the mixed-spin-channel Feshbach resonance [131, 134], we have obtained a reasonable description.

In Figure 3.9, we show the exponential lifetime of the total number of atoms as fitted to the numerical solution of Eq. 3.24 with the obtained parameters for an evolution time of up to 70 ms. Due to the Feshbach loss, the lifetime is imbalance dependent and diverges for $z \rightarrow -1$ (all atoms in $|a\rangle$). For $z \sim 0.15$ the lifetime is minimal (~ 100 ms for 500 atoms) while for $z \rightarrow 1$ it is limited by the spin-relaxation timescale (~ 200 ms for 500 atoms). The lifetimes

closely match the experimental data fitted exponentially for the same time scale, which gives a nice consistency check of the extracted parameters. We note, that the exponential lifetime depends on the fitting timescale, since for small evolution times the non-exponential decay is steeper than the exponential approximation.

4 State preparation and readout

Here, we detail the experimental steps for the diabatic production and characterization of non-classical states by time evolution close to the classically unstable fixed point. We give intuitive insights with the help of the phase-space Husimi distributions.

4.1 Spin manipulation sequence

After the production of the BEC array in the internal state $|F, m_F\rangle = |1, -1\rangle = |c\rangle$, the magnetic offset field is ramped to its final value of 9.13 G and actively stabilized. The spin state manipulation sequence begins with a radio-frequency adiabatic passage¹ in $F = 1$ to the state $|a\rangle = |1, 1\rangle$. Since it sweeps over the two Zeeman resonances, this transfer is robust and obtains a very high fidelity. The state $|a\rangle^{\otimes N}$ is the coherent spin state with $\vartheta = \pi$, located at the south pole of the sphere in our two-mode description. Subsequently, we apply an $\Omega_0\tau = \pi/2$ -pulse with the Rabi frequency $\Omega_0 \sim 2\pi \times 310$ Hz to rotate this state onto the equator. This sets the frame of reference between the atomic phase and the local oscillator, which in our case corresponds to the subtraction of the RF frequency from the MW frequency. We define the axis of the first pulse to be the positive y -axis. During a pause of $2\ \mu\text{s}$, we attenuate the microwave by 10.8 dB and the radio frequency by 14 dB, which reduces the Rabi frequency to $\Omega_{\text{evo}} \sim 2\pi \times 18$ Hz, corresponding to $N\chi/\Omega_{\text{evo}} \sim 1.5$, i.e. in the bifurcated regime. In combination with the phase shift of the MW attenuator ($\sim 83^\circ$), the phase of the RF is adjusted such that the rotation axis is approximately parallel to the mean spin direction, which corresponds to a preparation close to the unstable fixed point. The subsequent evolution time of variable length is interrupted by a 'spin echo' pulse with length $\Omega_0\tau = \pi$, which partly suppresses the effect of detuning fluctuations from realization to realization. We discuss below that in our nonlinear situation it doesn't only cause a phase inversion. After the second half of the attenuated evolution, we apply two types of rotations for state characterization: One with the rotation axis approximately antiparallel to the mean spin direction, which we call tomography rotation, characterized by the angle $\alpha = \Omega_0\tau$, and one with the rotation axis approximately at 90° to the mean spin direction, which we call interferometry rotation by the angle $\theta = \Omega_I\tau$. Since we want to apply also very small rotations in the order of 1° with the last pulse, we attenuate the RF by 6 dB, such that $\Omega_I \sim 2\pi \times 160$ Hz. This results in about 100 cycles of the RF per degree of rotation, such that switch-on effects, which typically distort the first ~ 2 cycles of the pulse become negligible. The timing sequence is illustrated in Fig. 4.1 with the corresponding rotations and a theoretical time evolution of the two-mode model.

¹The RAP consists of a linear ramp of the RF frequency from 7 MHz to 5.8 MHz in 20 ms with a Rabi frequency of ~ 10 kHz

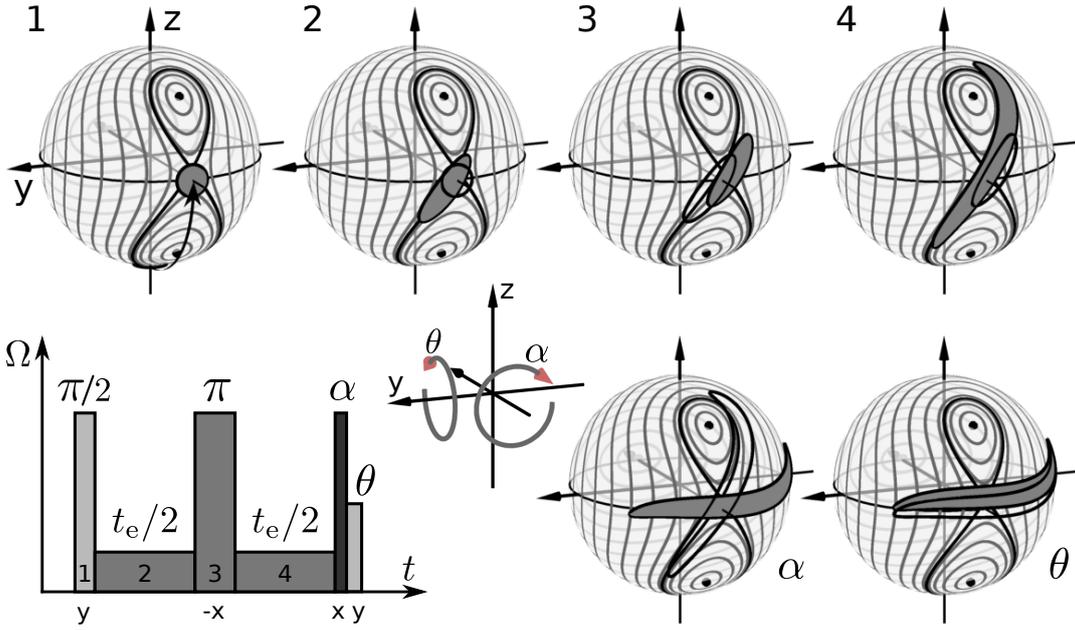


Figure 4.1 Experimental pulse sequence Here, we show the experimental timing sequence of the microwave and radio frequency pulses. The top row depicts the nonadiabatic state preparation sequence with 1) $\pi/2$ -pulse for preparation close to the unstable fixed point, 2) first half of the evolution with attenuated Rabi frequency, 3) spin-echo π -pulse, 4) second half of the evolution. The bar graph schematically shows the different Rabi frequencies of the pulses and the different pulse phases in grayscale. The corresponding rotation axis direction is noted underneath. The two spheres at the bottom show the action of the last two pulses used for state characterization. The rotation about the x -axis with angle α is used for state tomography, which is followed by a rotation about the y axis with (small) angle θ for the Fisher information extraction. The depicted classical phase space is the situation during the time evolution, which is shown with a detuning of $2\pi \times 1$ Hz, and the corresponding initial and final states are depicted by empty and filled contours, respectively. Without detuning, the resulting state would be always point symmetric to its intersection with the negative x -axis. Due to the asymmetry of the phase space in z -direction, the spin-echo pulse can only partly compensate the effect of detuning.

4.1.1 Towards the right parameters

The detuning for the fast pulses was coarsely minimized with the characterization of the AC Zeeman shifts by Ramsey spectroscopy at 9.2 G, farther away from the Feshbach resonance but still at a comparable offset field. A confirmation and fine adjustment was performed at 9.13 G with the symmetrization of plasma- and π -oscillations as described in Section 3.5, which turned out to be an invaluable tool.

The adjustment of the detuning during the evolution time is much more subtle, since the atom number dependence of the interaction shift has a much larger effect. For example, for 500 atoms the total interaction shift is ~ 17 Hz, which changes by $\sim \pm 1$ Hz with ± 60 atoms (this means also during the time evolution in the presence of atom loss). To come close to a reasonable setting, we measured time evolutions without spin echo pulse and decided for the detuning, which showed the most symmetric statistical spread in imbalance. Here as for all characterizations, the BEC array with several realizations at similar atom numbers in one realization is indispensable. Coarse adjustments can be typically done by hand while observing

the behavior along the lattice on the absorption pictures of single shots and statistical spreads can be coarsely estimated by on-the-fly analysis, which is close to impossible in reasonable time in the case of single realizations per experimental run, which currently requires ~ 38 s to complete.

The Rabi frequency of the fast pulses was separately calibrated by Rabi flopping for the two employed pulse powers. Typical drifts are smaller than 0.5%, which we correct by slight adjustments of the RF power before each long-term measurement of $\sim 3 - 5$ days duration.

The relative phase shift (modulo 2π) between the two paths of our microwave attenuator assembly was also determined by plasma/ π -oscillations. Here, we prepared the coherent state directly on the equator, switched the attenuator but kept the RF power at its initial level. The phase was found by adjustment of the RF phase of the second pulse such that the oscillation amplitude was minimized. The successful preparation on the ' π -side' can be checked by a small positive phase displacement from this setting and a rotation angle $\leq 180^\circ$ of the second pulse. The resulting sign of the imbalance is positive on the π -side and negative on the plasma-side². For unattenuated microwave, the same result is obtained for an RF phase shift of $3\pi/2 + \phi_{\text{prep}}$, where $\phi_{\text{prep}} \sim 3^\circ$ is the phase acquired during the preparation pulse due to finite detuning and the nonlinearity (see Fig. 3.6). The difference in RF phase between the two settings is the attenuator phase shift. Since the phase shifts are critical ingredients of the unstable fixed point evolution, we checked the stability of the assembly on a network analyzer while rapidly toggling its setting between attenuated and unattenuated. It turned out to be remarkably stable under physical stress like knocking with a screwdriver handle³ and temperature changes. Even heating it with a heat gun well beyond the human touch tolerance resulted in a change in phase shift of only about one degree. Since the lab temperature stability is better than 1K, this is more than sufficient.

In the following analyses, we concentrate on the time regime where the bending dynamics leads to non-Gaussian features and the spin squeezing vanishes. Experimentally, we found this time by tomographic characterization of the squeezing via the angle α for different evolution times [13, 130]. The day before a detailed characterization measurement with the acquisition of experimental probability distributions, we typically performed such a squeezing tomography, for which reasonable statistics can be acquired in one night of measurement time. Especially the angle of best squeezing can be robustly extracted. Since we expected the best performance in the neighborhood of this angle with minimal variance, we concentrated the detailed characterization there. In the presence of systematic drifts during the measurement time, it is important to shuffle and interleave the different measurement settings. For example, we changed the angle θ in every realization and the angle α for every row of realizations between two Ramsey corrections. For the measurements without θ -pulse, the setting of α was randomly shuffled.

²This is specific for positive nonlinearity χ . For $\chi < 0$ the picture changes by a phase of π .

³To obtain mechanical stability it is thoroughly screwed to an aluminum base-plate and interconnected with semi-rigid coaxial lines

4.1.2 Keeping track of the detuning

To ensure the resonance during a long-term measurement, we perform automated Ramsey spectroscopy on the two-photon transition after 30 – 40 experimental realizations, corresponding to a time interval of 20 – 25 minutes. For this, we acquire 10 shots of a $\pi/2 - T - \pi/2$ sequence with $T = 30$ ms free interrogation time with a phase offset of $\pi/2$ for the last pulse to convert acquired phase into detectable atomic imbalance. Due to the presence of the nonlinearity and thus one-axis twisting, the Ramsey contrast is reduced, which is taken into account by the automatic analysis. Due to the interaction shift, a quite narrow postselection on total atom number is required. We obtained a robust improvement of this issue by a linear fit to the measured imbalances versus total atom number and a correction of the setpoint of the magnetic field stabilization based on the evaluation of the fit at a particular target atom number. The correction values are used in the final data analysis: If the correction step between two consecutive Ramsey measurements is bigger than the equivalent of 1.5 Hz, which in the absence of big drifts occurs in the form of spiky outliers, we discard the preceding and following row of realizations, since this means that either a detuning change likely happened before the Ramsey shots or the evaluation was invalid, which erroneously altered the detuning of the following experiments. Note that the two-photon resonance is not only affected by magnetic fields (see section 3.4), which is why we chose to live with the disadvantage of the reduced sensitivity by one-axis twisting and the comparably weak magnetic field sensitivity in order to maintain the two-photon resonance under all influences. To take advantage of the spin squeezing is not possible here, since the required upright orientation of the long axis would only worsen the phase spread during the interrogation time.

4.1.3 Spin-echo pulse

The basic idea of a π spin-echo pulse in a linear situation (under the influence of the Hamiltonian $\delta\hat{J}_z$ with static δ only) is to invert the z -direction in the middle of the evolution time t_e such that the unitary $\hat{U}_1 = \exp[-i\delta\hat{J}_z t_e/2]$ of the first half cancels with $\hat{U}_2 = \exp[+i\delta\hat{J}_z t_e/2]$ of the second half and the sequence becomes robust against detuning fluctuations from realization to realization or detuning inhomogeneity in a single realization. In NMR spectroscopy, this leads to a rephasing of the spins and a revival of the precession signal [151] which is why it is called (Hahn) spin-echo. With the addition of the quadratic Hamiltonian $\chi\hat{J}_z^2$ this is still possible, since it is invariant under inversion of the z -axis. The situation changes fundamentally with the addition of the weak Rabi coupling $-\Omega\hat{J}_x$ during the time evolution. In this case, an additional detuning breaks the z -symmetry, which means that its effect cannot be perfectly canceled any more by z -inversion. In this sense, 'spin-echo' is a bit of a misnomer in our situation, it rather has to be understood as an integral part of our preparation scheme, partly compensating phase evolution but also symmetrizing to some degree the resulting quantum states in z -direction. This is illustrated in Fig. 4.2 with the help of the classical phase space and the contours of theoretical Husimi distributions. The situation further complicates since due to the interaction shift the detuning changes with the evolution time in the presence of atom loss. However, a reduction of the criticality and a concentration of the final states around the tomography axis can still be obtained. Note, that the big variation of the detuning depicted in Fig. 4.2 is much larger than the experimental reproducibility (~ 0.4 Hz for long-term measurements), but

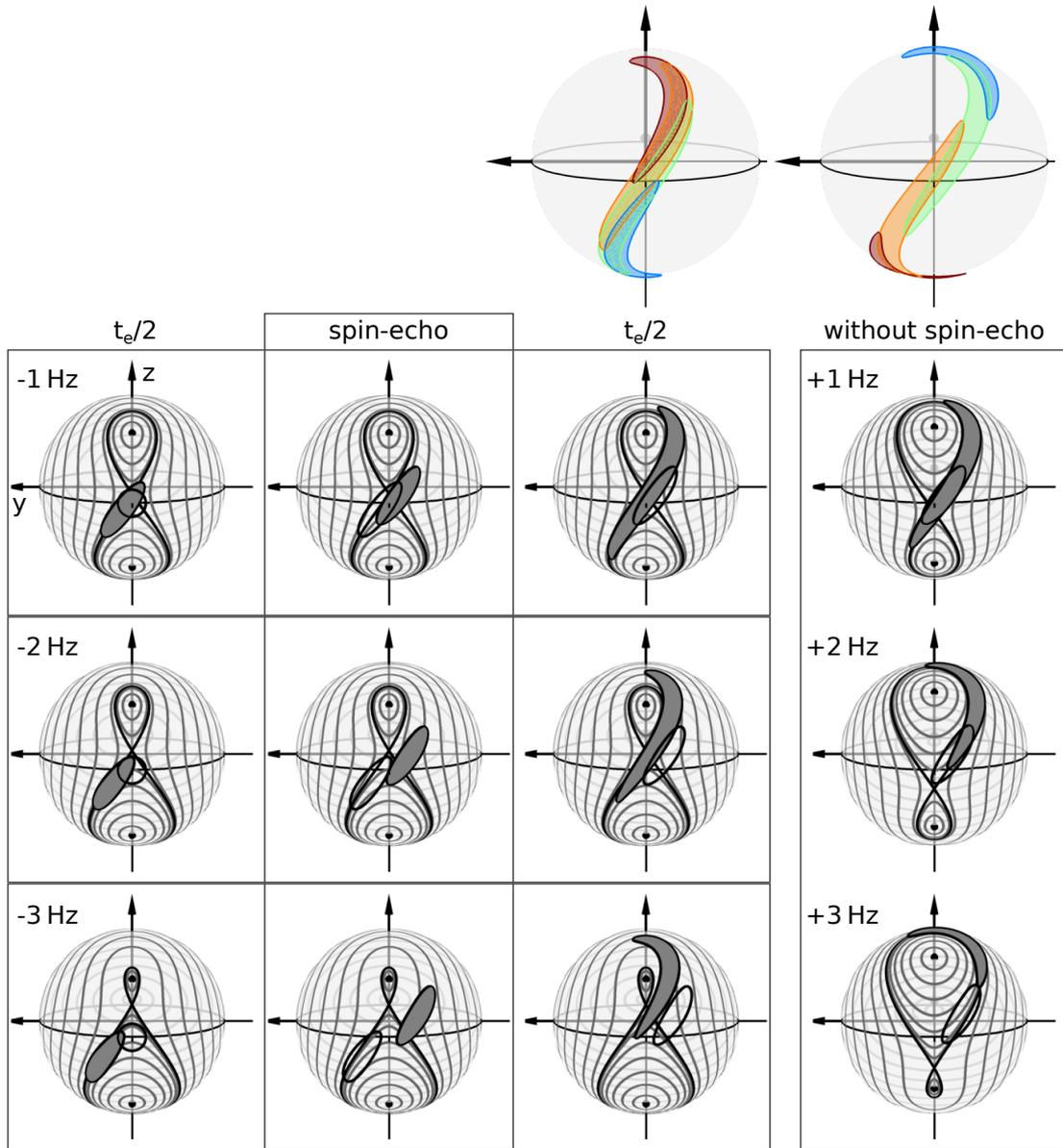


Figure 4.2 Spin echo pulse Here, we illustrate the action of the spin-echo π -pulse (rotation around the negative x -axis) in the middle of the time evolution for different detunings $\delta/2\pi$ and otherwise identical parameters ($\Lambda = 1.5, \Omega = 2\pi \times 20$ Hz). Due to the asymmetry of the phase space in z -direction, the spin-echo pulse can only partly compensate the effects of finite detuning. Nevertheless, it leads to a much reduced influence on the shape of the final state, as can be seen from the comparison to the case without spin-echo pulse. Since the π -pulse can also be seen as an z -inversion of the phase space, we compare it to the case of positive detuning during the whole evolution (right panel). It also leads to a concentration of the mean spin direction around the negative x -axis, which is beneficial for the readout rotations after the state preparation. The upper spheres show a summary of the position and orientation of the final states in the two cases for detunings of ± 1 Hz and ± 3 Hz. The time dependence of the parameters due to atom loss is neglected here to show the basic features.

provides an intuitive insight into the dependence on the total atom number (the range of 4 Hz corresponds approximately to the range from 300 to 500 atoms).

4.2 Readout and data analysis

After state preparation and rotation, the population of $|b\rangle = |2, -1\rangle$ is quickly transferred to $|c\rangle = |1, -1\rangle$ by a microwave π -pulse to stop further spin relaxation and Feshbach loss. This is followed by a controlled ramp-down of the magnetic offset field to a value of ~ 1 G in 300 ms such that the Zeeman splitting is below the linewidth of the resonant detection light. After Stern-Gerlach separation, the two components are destructively detected by absorption imaging (details see section 3.2.4).

Before post-processing, each row of 30 – 40 realizations is visually inspected and filtered for obviously bad realizations, for example rare cases in which the final π -pulse has not worked properly, the image has a big background noise level, the total atom number was abnormally low or the Stern-Gerlach splitting is too small, which happens if the imaging sequence is not triggered correctly. After this step, the optimal reference pictures, the atomic column densities and finally the atom numbers per component and lattice well are calculated.

Since the interaction parameters largely depend on the total atom number, we group the results in bins chosen as small as possible to still provide sufficient statistics. For the characterizations discussed in the following sections, they were fixed to a width of 30 atoms.

The analysis proceeds in calculating the atomic imbalances $z = (N_b - N_a)/N$ for every setting. These values are either directly analyzed in terms of their mean and variance or binned in histograms with a bin width of $4/N$. This width is chosen slightly smaller than the typical photon shot-noise contribution to our detection noise of $(\Delta z)_{\text{PSN}} \approx 6/N$. Smaller widths only lead to worse statistics per bin, while bigger ones tend to hide essential features of the states. To obtain the experimental probabilities (frequencies), the count numbers are normalized to the total number of counts in the histogram.

For comparability of our methods, we do not subtract detection noise in the variance analyses unless otherwise stated.

5 State tomography

5.1 Quantum state reconstruction

After the evolution time of 25 ms, the unrotated state features a very broad distribution of the imbalance with a variance about 40 – 60 times larger than that of a coherent spin state as can be seen in Fig. 5.1 in the top row for three different mean atom numbers. Essential further insight is gained by the application of the tomography rotation, which reveals that the shape of the state is anisotropic, non-Gaussian and strongly depends on the total atom number. We acquired a set of such distributions for tomography angles in steps of 10 degrees and finer sampling in 5 degree steps in the range of smallest variance (examples of obtained distributions can be seen in Fig. 5.1). With this data, we performed a maximum-likelihood reconstruction of the density matrix in the symmetric subspace of the final atom number.

Apart from the projection of the finite window in total atom number to its mean value, we assume in this approach that the qubits are indistinguishable and thus still confined to the two BEC modes after the evolution time, also in the presence of atom loss. With this assumption, a quantum state reconstruction seems to become less impossible, since this subspace scales linearly with the number of particles instead of exponentially for full state tomography. However, there are still $N(N + 1)/2$ complex and $N + 1$ real entries to be estimated.

For the reconstruction, we decreased the bin size in z to $2/N$, according to the granularity of the spin projections with $\{m_\alpha\} = -N/2, \dots, N/2$. We then apply the iterative method presented in [152–154] starting from the normalized identity matrix in \hat{J}_z -basis. The algorithm iteratively maximizes the likelihood for the given measurements to be the result of projections of the reconstructed state along the measured directions. Since we can exclude from full Rabi flopping with orthogonal rotation axis that the state has contributions on the positive x -axis, we include one projection in this direction and assign zero probability to the positive side in the calculation of the iteration step.

The method introduced in [74] on the basis of the Wigner function akin to the planar filtered backprojection used in optics has the intrinsic problem, that the obtained density matrix is not physical, i.e. it usually obtains negative eigenvalues. Furthermore, nonuniform sampling leads to artifacts as in any tomographic backprojection approach, which is also not as problematic for the maximum likelihood method, since a natural weighting occurs in the calculation of the iterative step and all measured data can be included. Furthermore, the reconstruction stays in the space of physical density matrices.

The maximum likelihood approach is very tempting, since it was argued that it should be also possible to deal with informationally incomplete data in this way and obtain the 'most probable' density matrix given the measured data, which contains the least assumptions about things that have not been measured (maximum likelihood maximum entropy approach [155]).

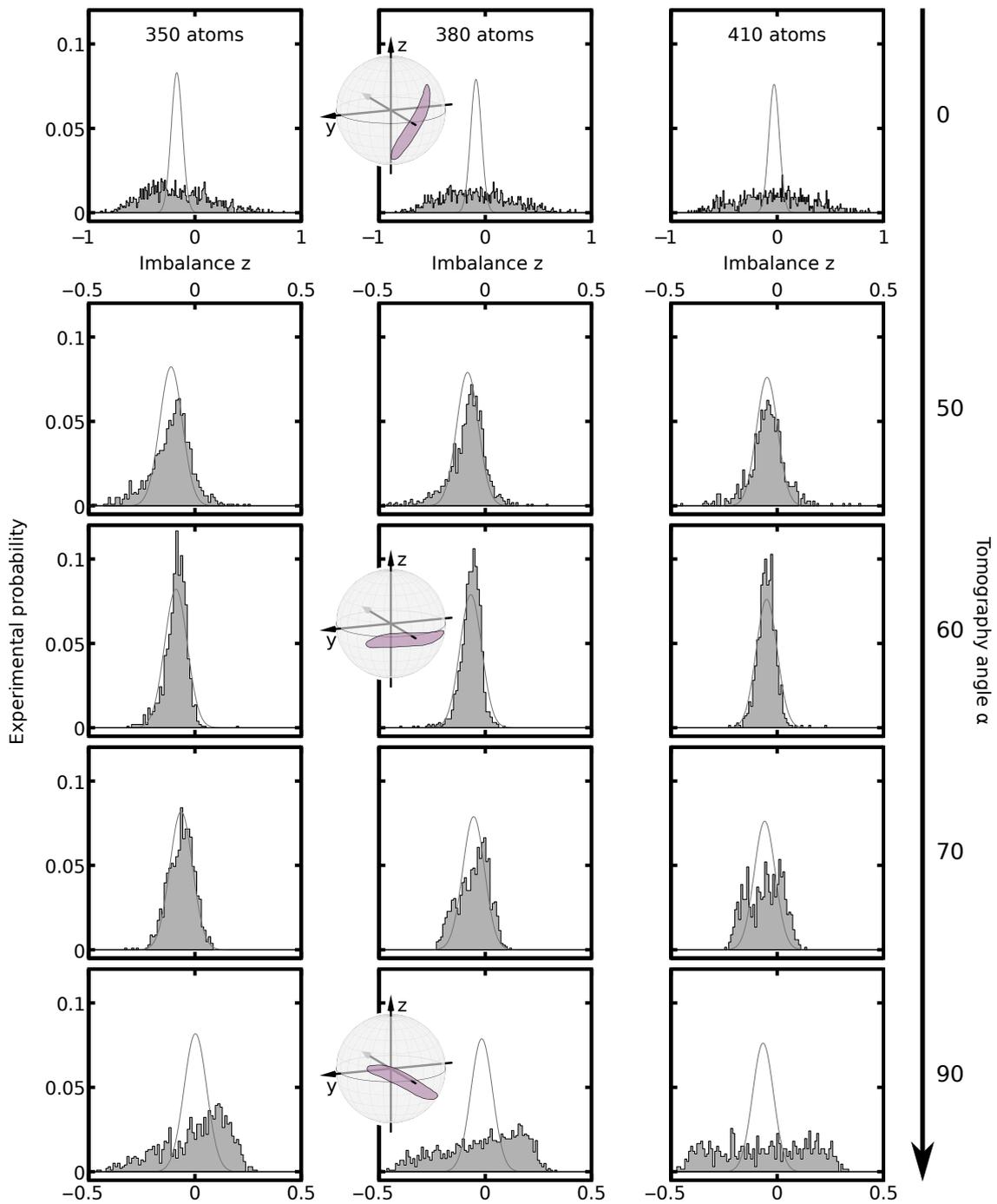


Figure 5.1 Experimental probability distributions of the tomographic characterization Here, we show exemplary histograms of the state tomography after 25 ms of evolution time for three total atom numbers. The unrotated distributions are shown in the topmost panel. Note that for the rotated states the axis is zoomed to half its size. In all panels, the line represents the Gaussian distribution of a coherent spin state with the same atom number, positioned at the mean imbalance of the experimental histogram. The distributions are clearly non-Gaussian and show a big variation versus total atom number.

One could hope to obtain some coarse-grained version of the density matrix, which allows to extract information for example about low spin moments ($\langle \hat{J}_\alpha^k \rangle$ for small k) or the Quantum Fisher information. Usually, a parametric bootstrap method is employed to give error margins on such parameters R calculated from the reconstructed state. One approach is to use the reconstructed density matrix to generate new datasets by random sampling. With the same statistics as in the first estimation, the reconstruction can then be repeated many times. While the scatter of R based on the simulated reconstructions becomes a reasonable estimate of the statistical error ΔR_{stat} , the bias away from the true value cannot be easily obtained in this way. We tested this approach on our reconstructed states for the Quantum Fisher information and obtained typical values of $\max(F_Q/N) \sim 12 \pm 1$ for 380 atoms, which is even larger than we would expect from the perfect evolution in the presence of detection noise (see Section 2.6). We found that even the calculation of the second spin moments \hat{J}_α^2 on operator level, which can be efficiently cross-checked with the experimental data along the measured directions, is not reliable. This particular feature can be easily understood by inspection of the projections of the density matrix. While the shape of the experimental distributions and the reconstructed projections match, there is finite background noise on the latter, which increases the variance in all directions. Higher spin moments share of course the same problem. It was recently shown that statistical bias is an unavoidable feature of maximum-likelihood reconstructions [86]. However, the proposed method using linear inversion of the projections to circumvent this is intractable in our case.

If any proper method would be feasible and sufficient statistics could be acquired, the restriction on the symmetric subspace or another truncation would still be required. A full quantum state reconstruction would not only require local operations on the qubits, but would also scale exponentially with the atom number in both required statistics and required computational power or even memory. Since the dimension of the whole Hilbert space for 300 qubits is already larger than the estimated number of atoms in the universe, reconstruction procedures for mesoscopic systems will always contain assumptions.

While quantitative statements based on reconstructions have to be enjoyed with caution, in our case it is still very useful to obtain intuitive insights on the shape and position of the quantum states with the help of quasiprobability distributions thanks to their inherent filtering properties (see section 2.1.5). We chose the Husimi distribution (Eq. 2.33), i.e. the overlap with coherent spin states of the same atom number for this purpose, since it is intuitive and the Wigner function does not provide further insights.

Figure 5.2 shows the results displayed on the spherical phase space. For these plots the coherent state overlap was calculated on a grid of size 250×250 for ϑ and φ and normalized to the respective maximum of each distribution. The brightness scale was chosen such that the values around $1/e^2$ from the maximum are perceived as dark (green) rim. The rotation axis for the reconstruction was chosen as the positive x -axis and thus the phase offsets of the states from this axis have to be understood as offsets from the rotation axis and are not referenced to the initial preparation axis. Since the axis of the additional θ -rotation (omitted here) for the Hellinger distance evaluation was performed quickly after the tomography rotation and at a precise phase offset of $\pi/2$ from the tomography axis, this directly constrains the mean phase error of this additional transformation. The Husimi distributions show the expected bending of the state, their comparably steep orientation and the systematic change with the atom number, which is caused by the different detuning during the time evolution due to the interaction shift

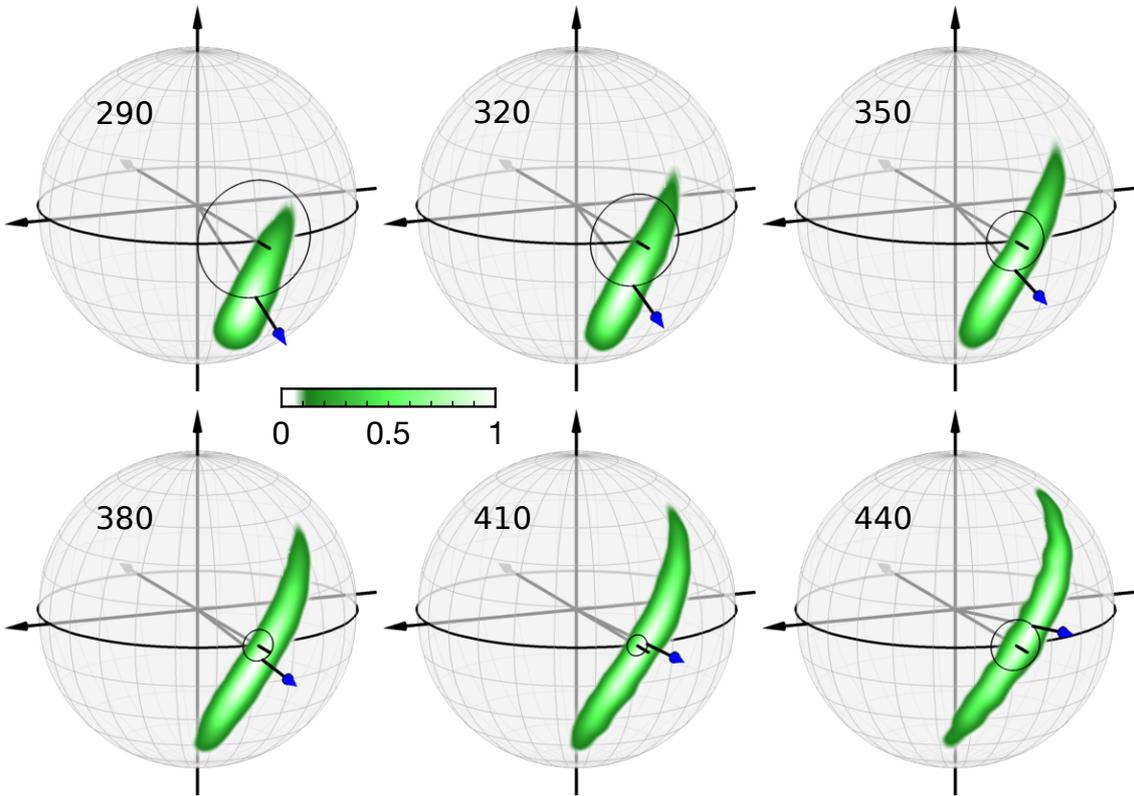


Figure 5.2 Gallery of reconstructed quantum states for 25 ms of evolution time Here, we show the normalized Husimi distributions calculated from the reconstructed density matrices. The tomography rotation axis is the axis pointing to the backside of the sphere. The states are turned clockwise for increasing tomography angles. The path of the mean spin vector (arrow) during the tomography is indicated as line. The middle of the post-selection window for the total atom number is indicated on each sphere. The Husimi distributions provide an intuitive picture of the systematic behavior versus total atom number. The shift of the mean spin vector and the comparably strong behavior of the shape is a combination of the interaction detuning in combination with the spin-echo pulse (see Fig. 4.2).

and the nonlinearity $N\chi \propto \sqrt{N}$, which steepens the meanfield separatrix around the unstable point for decreasing atom number. The mean spin vector and its path for the tomography rotation is indicated as arrow and line respectively. The ripple-artifacts for larger atom numbers are the result of comparably lower statistics of this reconstruction. The total number of experimental realizations in this measurement was $\{13408, 16852, 20766, 20007, 13296, 6093\}$ for $\{290, 320, 350, 380, 410, 440\}$ atoms, respectively.

Figure 5.3 shows the result of a reconstruction measurement for the squeezed Gaussian states with 15 ms of evolution time. Here, a lower number of realizations (6–13 thousand) is sufficient for all atom numbers, since the distributions of these states are much more concentrated.

In view of the conceptual difficulties and experimental demand of quantum state estimation, independent and scalable statistical methods that directly employ the obtained experimental probability distributions are required for the characterization of non-Gaussian states. Our method based on the Hellinger distances constitutes such a method, which in addition does not require the symmetric subspace assumption.

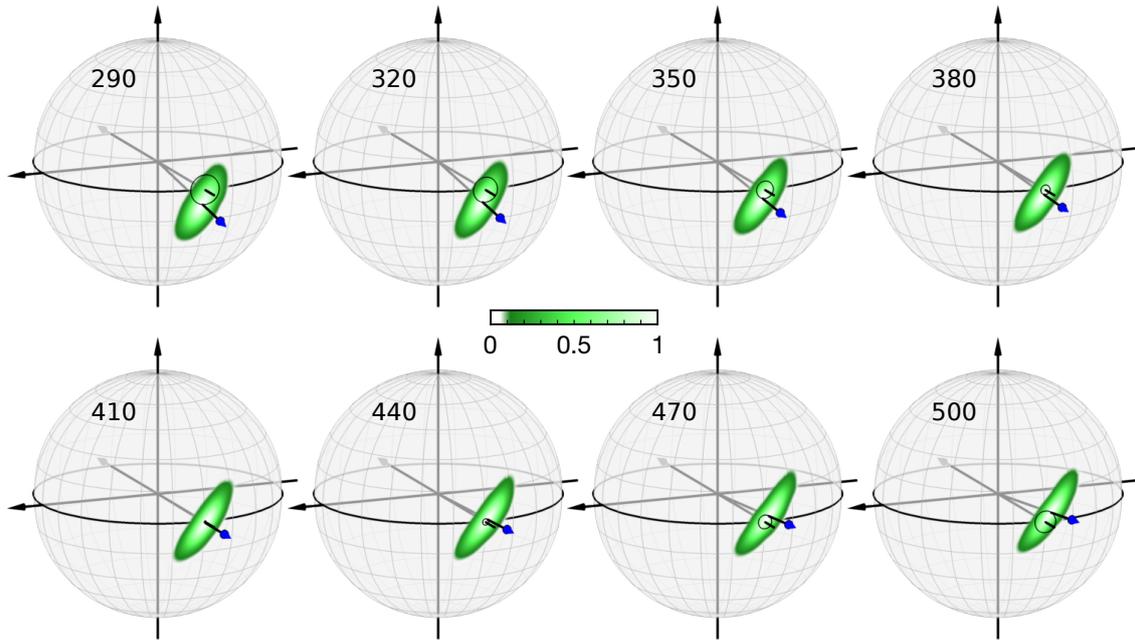


Figure 5.3 Gallery of squeezed states For intermediate evolution times (here 15 ms), squeezed states are generated by the unstable fixed point evolution. The maximum likelihood reconstruction nicely shows the systematic behavior for different final atom numbers. The shift of the mean spin direction vs. atom number is the remaining effect of the interaction detuning while the stronger squeezing and decreasing angle of the long axis to the equator for larger atom numbers is caused by the increasing nonlinearity $N\chi \propto \sqrt{N}$ and $\Lambda = N\chi/\Omega$ respectively. The arrow indicates the mean spin direction and the line its path during the tomography rotation.

5.2 Squeezing tomography and two-mode model

Here, we present quantitative results of the tomography measurements in terms of the variance parameters, i.e. we compare the variance of the generated states with those of the corresponding coherent spin state of the same atom number N . We work with imbalances $z = (N_b - N_a)/N$, so the coherent states have the binomial variance $(\Delta z^2)_{\text{CSS}} = 4p(1-p)/N$, where p is the fractional mean population of state $|b\rangle$, that is $p = \langle N_b/N \rangle = (\langle z \rangle + 1)/2 = \cos^2(\langle \vartheta \rangle/2)$ (see Eq. 2.21). In particular, we compute the number squeezing factor

$$\xi_N^2 = \frac{\Delta z^2}{(\Delta z^2)_{\text{CSS}}} = \frac{N\Delta z^2}{4p(1-p)}. \quad (5.1)$$

When we work with a logarithmic scale, quantities in decibel (dB) are calculated as $\xi_{(N)}^2[\text{dB}] = 10 \log_{10}(\xi_{(N)}^2)$. Easy to remember orientation points are $\{0, 3, 5, 10, 20\} \text{ dB} \simeq \{1, 2, \pi, 10, 100\}$. We show two example plots of this quantity as a function of the tomography angle in Fig. 5.4 (left). They show the expected sinusoidal behavior and a robust squeezing below the coherent state level of 0 dB. The results presented here are directly calculated for the whole datasets spanning ~ 3 days of measurement time without subtraction of detection noise, whose contribution is $(\Delta z^2)_{\text{det}} \approx (2\sigma_{\text{PSN}}^2 + 2\sigma_{\text{fr}}^2)/N^2$ with the photon shot noise $\sigma_{\text{PSN}} \approx 4$ atoms and fringe noise $\sigma_{\text{fr}} \lesssim 2$ atoms per component [137]. We extract the minimal value (short axis), the

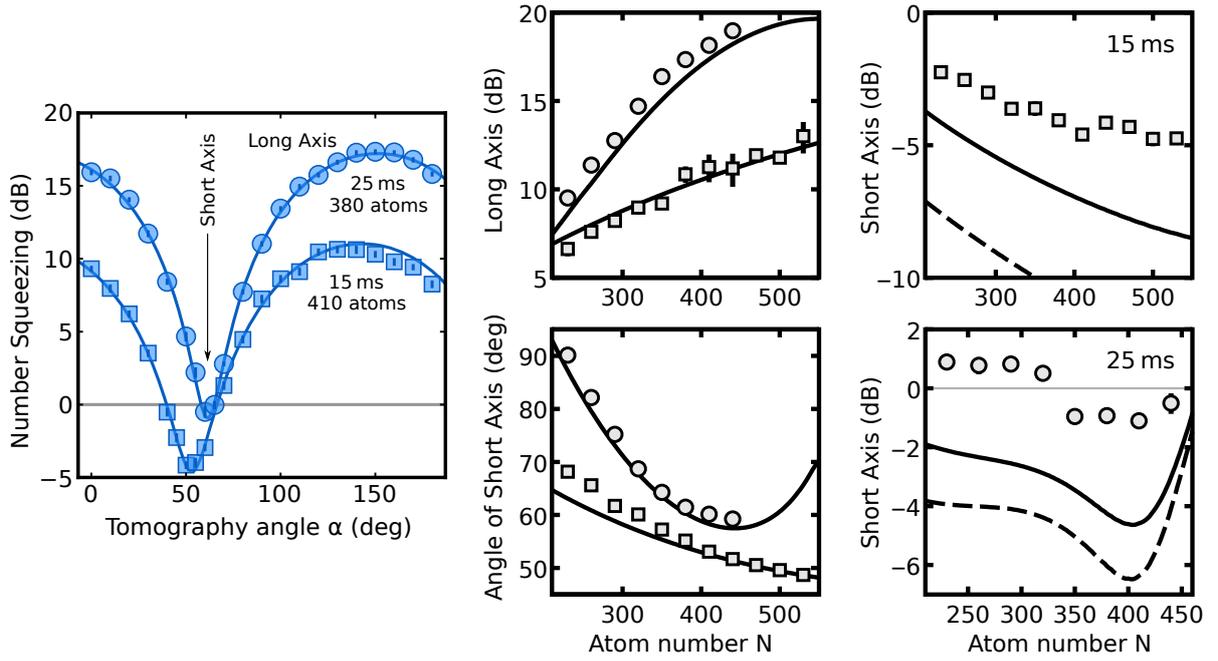


Figure 5.4 Squeezing parameters of the quantum state tomography. **Left:** The number squeezing parameter shows a sinusoidal behavior versus tomography angle, from which the sizes of short and long axes and from their angle the orientation of the quantum state can be obtained. **Right:** The extracted parameters show a systematic scaling with the total atom number, resulting from the dependence of the nonlinearity and the interaction detuning on the atom number. The behavior of the long axis and the angle of the short axis can be reproduced with a two-mode model, taking into account the independently characterized time scales of atom loss and the resulting time dependence of the parameters of the Hamiltonian (solid lines). The experimental size of the short axis is larger than that of the model (dashed line), since it does not include technical noise contributions and noise effects of atom loss. For the solid line the photon shot-noise contribution of the measurement is taken into account.

maximum value (long axis), and the tomography angle of the minimum from sinusoidal fits. The squeezing and the size of the long axis increases with the atom number, while the angle systematically shifts to shallower angles as the interaction parameter $\Lambda = N\chi/\Omega$ increases with the atom number, which decreases the steepness of the meanfield separatrix at the unstable point. The loss of atoms during the evolution time leads to a decrease in Λ and thus to an increasingly steeper separatrix. Therefore larger tomography rotation angles are required to reach the minimal projection in z -direction. The bending of the non-Gaussian state leads to a further increase, since for these states the smallest variance is not reached when the state is most sharply peaked in readout direction, which would correspond approximately to the angle of the separatrix, but at slightly larger rotation angles when the smallest z -projection of the 'tails' is realized. As the state increases in size and bends, the squeezing vanishes.

With the help of the characterized time scales of atom loss, the scaling of the interaction shift and the nonlinearity with the atom number, the variance results of the tomography for

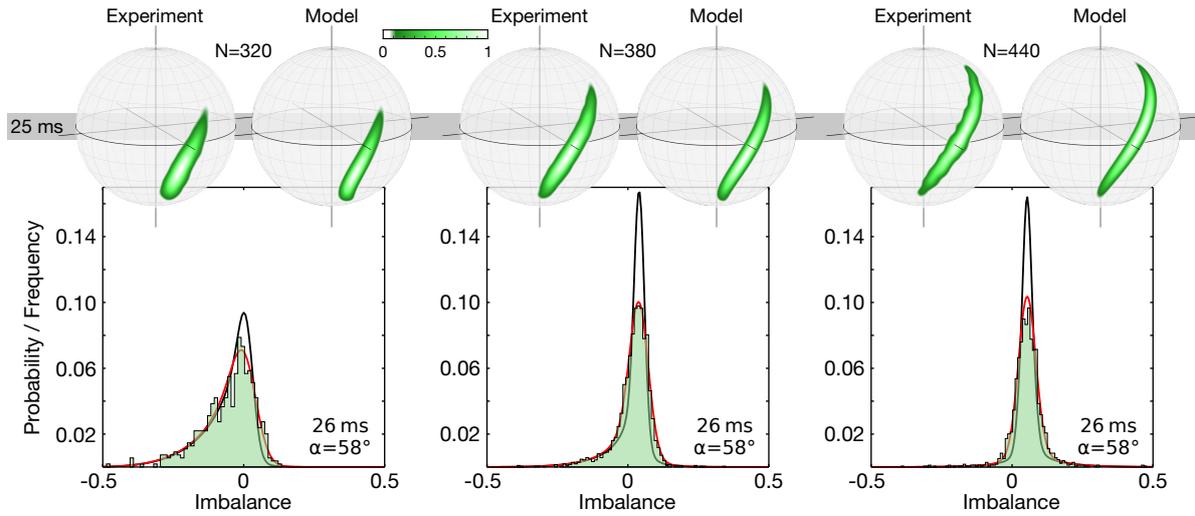


Figure 5.5 Comparison of the two-mode model with experimental distributions. The top row shows a comparison of selected Husimi distributions from the tomographic reconstruction to the corresponding results of the two-mode model. The systematic coarse structure of the Husimi distributions is reproduced very well, while the experimental states are broader than those of the idealistic model. The bottom row shows a comparison of the model with experimental non-Gaussian probability distributions obtained for an evolution time of 26 ms. The black line shows the probability distribution of the model convolved with the detection noise of the measurement. With an additional Gaussian convolution with $\Delta z_{\text{add}} = 10/N$, the model and the observed distributions closely match. The additional broadening can be motivated by technical phase noise of the tomography rotation axis (see text).

the two evolution times can be compared with the two-mode model. For this, we numerically integrate the Schrödinger equation $i\partial_t |\psi\rangle = \hat{H} |\psi\rangle$ with the Hamiltonian

$$\hat{H} = \frac{\Omega}{N_f} \Lambda(t) \hat{J}_z^2 + \Omega (\sin \phi \hat{J}_x + \cos \phi \hat{J}_y) + \delta(N(t)) \hat{J}_z \quad (5.2)$$

in the symmetric subspace of the final atom number N_f with the help of a forth-order Runge-Kutta method. The interaction parameter $\Lambda(t)$ is time dependent and calculated as $\Lambda(t) = N(t)\chi(N(t))/\Omega$. This step is required to obtain consistent dynamics, since we simulate in a Hilbert space which is smaller than those of the actual atom number $N(t)$.

The individual steps of the experimental sequence are replicated with piecewise adjustment of Ω and ϕ . All pulses are modeled in the presence of nonlinearity and we include a small phase offset ϕ_{off} for all periods with high Rabi frequency Ω (preparation, spin-echo and tomography). This is motivated by an experimental script error we discovered after the measurements in conjunction with the switching of the microwave attenuator. We estimated this error to be $\sim -6^\circ$. Comparing the mean imbalance and the angle of minimal variance during the

tomography rotation of both measurements, we found a good agreement to the observed atom number systematics for the parameters

$$\begin{aligned}\Omega_{\text{rot}} &= 2\pi \times 310 \text{ Hz} \\ \Omega_{\text{evo}} &= 2\pi \times 18.4 \text{ Hz} \\ \delta(N(t)) &= 2\pi \times (18 - 0.77\sqrt{N(t)}) \text{ Hz} \\ N(t)\chi(N(t)) &= 2\pi \times 1.25\sqrt{N(t)} \text{ Hz} \\ \phi_{\text{off}} &= -8.5^\circ.\end{aligned}$$

The results are shown in the right panels of Fig. 5.4 as lines. The scaling of χ and δ as well as the loss timescales were set according to the results of the independent characterizations (see sections 3.5 and 3.7). The Rabi frequency during the evolution Ω_{evo} is consistent within 0.3 dB with the characterized total attenuation of 24.8 dB. Slightly better agreement can be obtained with an additional phase shift $\sim 1^\circ$ of the preparation pulse, which would be caused by an additional detuning of only ~ 5 Hz during this pulse, corresponding to $\sim 3\%$ of the total AC Zeeman shift of ~ 190 Hz and well within our calibration uncertainty. Of course the same uncertainty applies to the other two pulses with high Rabi frequency. In this respect, our simplified model can capture the systematic behavior surprisingly well. Not surprising on the other hand is that the exact values of the minimal variances are smaller than experimentally observed, since noise contributions of technical nature and from atom loss are not included. The effect of detection noise is the difference between the dashed and solid lines shown in Fig. 5.4. Apart from mean detuning and phase shifts of all pulses, which would be experimentally eliminated if known exactly in the first place, an exhaustive simulation of all contributing effects (for example using a Monte Carlo wavefunction approach [156]) would need the input of the branching ratio between three-body aba/abb and ab Feshbach losses, which is currently hard to obtain experimentally (see section 3.7). While we expect the channel ab to be pretty much harmless [157] besides the modification of the parameter $\Lambda \propto \sqrt{N}$, all asymmetric losses kick the imbalance z and thus the angular momentum of our inverted pendulum (see section 2.1.9) at random times [157–159]. We note that AC phase noise, for example from the microwave generator and the magnetic fields during the evolution, also jiggles this phase to some extent. To precisely take these noise sources into account requires further independent characterizations, since no detailed insight can be gained from a model by introducing free parameters which cause similar effects.

In Fig. 5.5 we compare our simplified model to some of the Husimi reconstructions for 25 ms and to non-Gaussian distributions obtained at an evolution time of 26 ms at a tomography angle of 58° (for $N = 380$ atoms this is the angle of largest observed Fisher information). The systematic coarse structure of the Husimi distributions is reproduced very well, while the discussed broadening is also apparent here. The black lines in the distribution plots for 26 ms are model results convolved with a Gaussian distribution corresponding to our photon shot noise¹. They already show the characteristic asymmetry observed in the experiment. Strikingly, a simple further Gaussian convolution with $\Delta z_{\text{add}} = 10/N$ allows to match the observed distributions with the model (red lines). At first sight, this seems to be a big effect,

¹Here, the small additional θ -rotation pulse is included in the model. Otherwise, the $\langle z \rangle$ -position would not match.

since it corresponds to additional variance in $N_b - N_a$ of 100. However, compared to the various conceivable effects discussed above, this is rather small, since we have also lost ~ 100 atoms at this time. One contribution that certainly matters is the phase noise of the tomography pulse, since we rotate a large portion ($\sin(58^\circ) \approx 85\%$) of it into the readout direction. In this respect, Δz_{add} corresponds for 380 atoms to a phase noise of $\Delta\phi \approx 1.8^\circ$, which is well in the realm of our technical noise. Taking for example the suppression of DC magnetic field fluctuations by the spin-echo pulse characterized in [130] for 15 ms to $S \sim 5$ dB, we arrive at $\Delta f = \Delta\phi / (2\pi t_{\text{evo}} S) \approx 0.36$ Hz for $t_{\text{evo}} = 25$ ms, which corresponds to a magnetic field stability of ≈ 34 μ G. The requirement is even stricter, since the expected performance of the spin-echo pulse is worse at 25 ms due to the nonlinear evolution. Finally, we point out that already the specified phase noise performance of the employed microwave generator is in the range of $\Delta\phi$ (see Appendix B.4) and we characterized the phase noise of the preparation pulse to be already $\approx 0.5^\circ$.

6 Fisher information extraction

Here, we show experimental results for the extraction of the Fisher information with the help of our statistical Hellinger distance method, which was theoretically analyzed in section 2.4.3. We first summarize the method, visualize its application with an example set of non-Gaussian experimental probability distributions and show selected extractions of the Fisher information including a reference measurement with the initial coherent spin state. We then detail the results for Gaussian spin squeezed states and the transition to the non-Gaussian regime.

6.1 Hellinger distances and extraction method

The Fisher information for a particular transformation is the parameter that characterizes the statistical speed (or rate of change) of the Hellinger distance

$$d_{\text{H}}^2(\theta) = \frac{1}{2} \sum_z \left(\sqrt{P_z(\theta)} - \sqrt{P_z(0)} \right)^2 \quad (6.1)$$

as a function of the parameter θ , i.e. $\partial d_{\text{H}}/\partial\theta = \sqrt{F/8}$. Furthermore, we have obtained the corresponding Taylor expansion

$$d_{\text{H}}^2(\theta) = \frac{F}{8}\theta^2 + \mathcal{O}(\theta^3), \quad (6.2)$$

which is the basis of our estimation of the Fisher information. We theoretically examined in 2.4.3 the properties of the extension of the Hellinger distance to experimental probability (frequency) distributions $\mathcal{F}_z(\theta)$ and showed a method to deal in a consistent way with the arising statistical bias due to the finite number of samples in the experimental histograms using a Jackknife resampling method. To simplify the notation, we make in the following no distinction between $\mathcal{F}_z(\theta)$ and $P_z(\theta)$. Here, we only deal with experimental probabilities.

For each Fisher information estimation we experimentally acquire a row of 9 histograms of the atomic imbalance z for distinct angle settings θ_{set} , realized with different lengths of the final pulse (rotation illustrated in the inset of Fig. 6.1). To obtain consistent conditions and make switch-on effects of the pulse negligible, we work at an offset-angle of $\theta_{\text{off}} = 6^\circ$. In the data analysis, we correct for the technical rounding of the pulse length to full μs using the calibrated Rabi frequency and then subtract the offset angle, such that $\theta = \theta_{\text{set}}^{(\text{corr})} - \theta_{\text{off}}^{(\text{corr})}$. The histograms are then normalized to their individual total number of counts to obtain the experimental probability distributions $P_z(\theta)$. We experimentally sample the reference setting $\theta = 0$ approximately four times as often, since its error contributes to all calculated Hellinger distances. For each $P_z(\theta)$ in combination with $P_z(0)$ we calculate the squared Hellinger distance $d_{\text{H}}^2(\theta)$ (Eq. 6.1). To subtract the statistical bias (offset) and estimate the statistical error

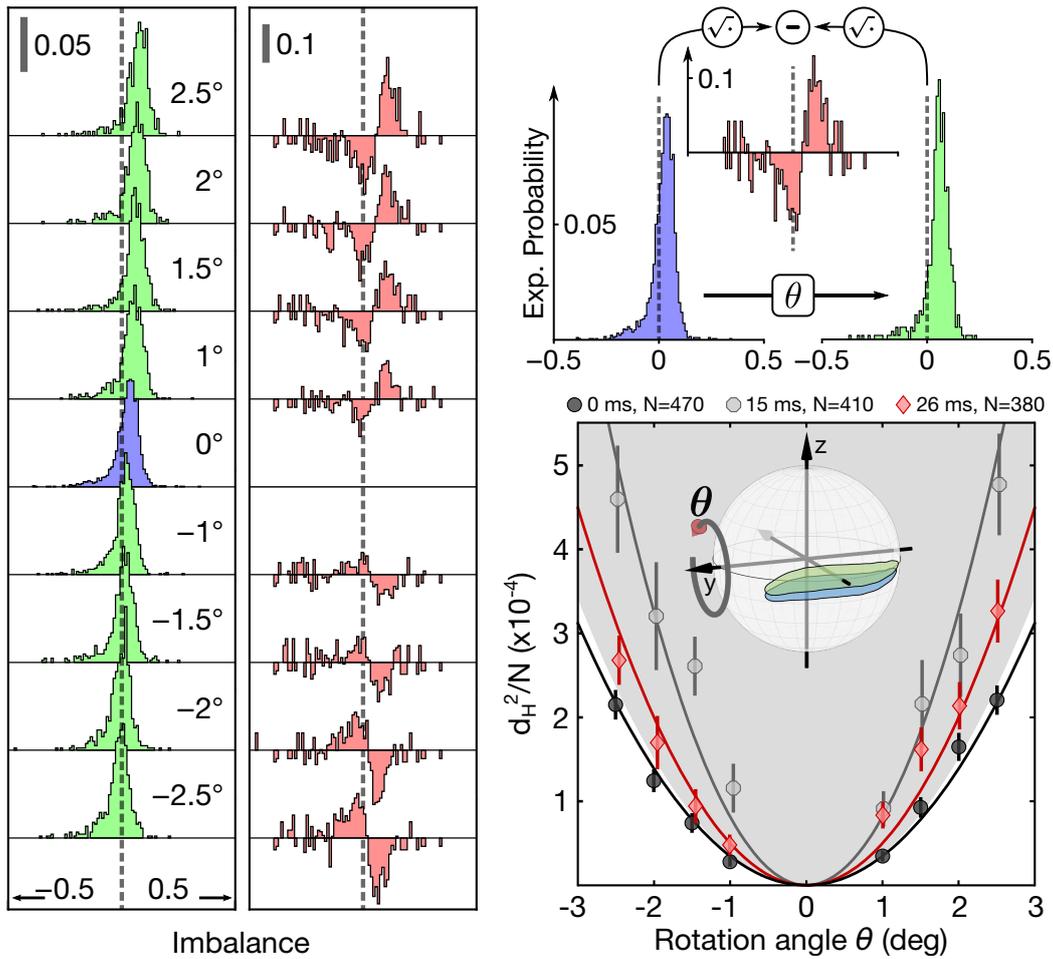


Figure 6.1 Extraction of the Fisher information with the Hellinger distance method. **Left row:** Examples of experimental probability distributions $P_z(\theta)$ of the imbalance z obtained with 26 ms of evolution time, a tomography angle $\alpha = 58^\circ$ and the indicated θ -rotation angles. The reference distribution $P_z(0)$ is colored in blue. The bottom right inset shows a sketch of the corresponding situation on the Bloch sphere. The main action of the additional θ -rotation is a shift of the distributions towards higher imbalances. **Middle row:** Differences $\sqrt{P_z(\theta)} - \sqrt{P_z(0)}$ of the probability amplitudes between the corresponding distributions and the reference. **Right, top:** Sketch of the procedure for extraction of the squared Hellinger distance (example of $\theta = 1.5^\circ$). The probability amplitudes $\sqrt{P_z}$ are subtracted for each bin and these differences are summed in quadrature to obtain the squared Hellinger distance. **Right, bottom:** Squared Hellinger distances for three different scenarios. The curvature of the respective quadratic fit is proportional to the extracted normalized Fisher information F/N . The values of a reference measurement with the initial coherent spin state (black circles) lie slightly outside the non-classical region (gray shaded area) and show that the initial state performs close to the standard quantum limit [$F(0\text{ ms})/N = 0.91(4)$]. The spin squeezed state [gray octagons, $F(15\text{ ms})/N = 2.2(2)$] surpasses the classical limit. After 26 ms (red diamonds), the state is non-Gaussian and features a Fisher information $F(26\text{ ms})/N = 1.31(9) > 1$, which demonstrates entanglement in a regime where no spin squeezing is present. The corresponding total atom numbers N are chosen according to the time scale of atom loss and are indicated in the legend.

on d_{H}^2 we average the block Jackknife results for mean and variance up to block size $h = 20$ (see Eq. 2.103 and 2.104). We verified with Monte-Carlo simulations (see section 2.4.5) and experimental data that this consistently reduces the statistical offset of d_{H}^2 to a negligible level and we fit quadratically without additional offset to estimate F from the curvature and its error from the 68% confidence region of the least-squares fit, taking into account the inverse Jackknife variances for $d_{\text{H}}^2(\theta)$ as fit weights.

The extraction procedure is illustrated in Fig. 6.1 together with several examples. To compare different evolution times and corresponding atom numbers, we normalize the squared Hellinger distances to the atom number, which allows to define the non-classicality region shown in gray. The standard quantum limit for the curvature of d_{H}^2/N is $1/8$, corresponding to $F = N$. Points inside the gray region are only accessible with $F > N$ and thus with entanglement in the system.

We performed a reference measurement with the initial coherent spin state with the result $F/N = 0.91(4)$, which is outside the non-classical region. While the ideal value is $F/N = 1$ in this case, in the presence of our detection noise we expect here a value of 0.93, which is consistent with the extracted Fisher information. For the squeezed state at 15 ms, we extract $F/N = 2.2(2)$, which confirms entanglement, while the spin-squeezing analyzed with the same measurement would correspond to a value of $1/\xi^2 \approx 2.6$. We underestimate the Fisher information here due to the chosen finite rotation angles for the extraction (detailed below in section 6.2). For the non-Gaussian state at 26 ms (experimental probability distributions shown in Fig. 6.1) we extract $F/N = 1.31(9)$, which shows entanglement in a regime where no spin-squeezing is present (see Fig.6.4). The behavior of the Fisher information during the transition from Gaussian to non-Gaussian states is detailed in the following section 6.3.

We note that the Hellinger distance method does not rely on symmetry assumptions about the quantum state. It directly characterizes the distinguishability of observable distributions under the action of a global transformation, which can be precisely calibrated by Rabi flopping. While still requiring a lot of statistics on the scale of BEC experiments (we typically acquire ~ 2000 realizations for $\theta = 0$ and ~ 500 realizations at $\theta \neq 0$), it is a quite efficient method, for example in view of the limits of quantum state reconstruction, and offers an intuitive interpretation. Its efficiency can be largely enhanced by additional information about the quantum state in order to choose the best transformation. In our case this was enabled by a combination of the intuition gained from the classical trajectories and the corresponding time evolution of quasiprobability distributions in the ideal two-mode model. Experimentally, mean and variance properties of z provided an important orientation in the parameter space of evolution time and tomography angle.

6.2 Gaussian spin-squeezed states

For a proper comparison of the obtained Fisher information results with the spin squeezing we have to take into account the reduction of the normalized mean spin length due to the increased extension of the quantum state, which fundamentally limits the obtainable contrast (visibility) in a Ramsey experiment [100]. For this we acquire interleaved experimental realizations with the tomography angle α that results in the biggest variance of z , i.e. the long axis of the state (see Fig. 5.4). We convert the obtained imbalances into angles via $\vartheta = \arccos z$ and then

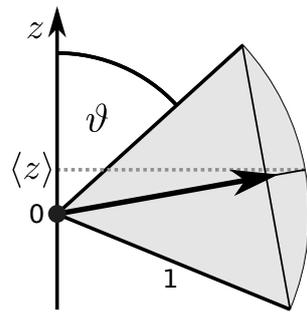


Figure 6.2 Geometric picture for the evaluation of the mean spin length. Due to the increased extension of the quantum state along its long axis (shown as gray angular spread), the mean normalized spin length (arrow) is reduced, which fundamentally limits the obtainable visibility \mathcal{V} of an otherwise perfect Ramsey experiment [100]. We evaluate the mean length of the arrow as $\langle \cos[\vartheta - \langle \vartheta \rangle] \rangle$.

calculate the estimate of the mean spin length as $\mathcal{V}_{\text{long}} = \langle \cos[\vartheta - \langle \vartheta \rangle] \rangle$. The geometry of this procedure is shown in Fig. 6.2. Here, we use the symmetric two-mode approximation, since we want to compare the Fisher information with the fundamental limit for spin squeezing. This means, we do not take into account any further technical or physical reduction of the mean spin length. The presented experimental squeezing values are calculated as $\xi^2 = \xi_N^2 / \mathcal{V}_{\text{long}}^2$ with the number squeezing factor ξ_N^2 (Eq. 5.1) and averaged over all angles θ of the Hellinger distance method. Note that ξ^2 is still a function of the tomography angle α in our case.

In Fig. 6.3 we show a comparison of the normalized Fisher information F/N extracted with the Hellinger distance method to the inverse spin squeezing factor $1/\xi^2$ for 15 ms of evolution time versus tomography angle α . This angle was varied around the axis of best number squeezing to obtain a systematic characterization. Both parameters lie consistently above the standard quantum limit and share the same optimal tomography angle.

Theoretically, for Gaussian spin squeezed states the two values agree, i.e. $F/N = 1/\xi^2$ (see section 2.6). However we observe here that the values of the Fisher information are systematically smaller. For the investigation of this behavior, we performed full rotations of θ and extracted the amplitude \mathcal{V} of the resulting fringes using a sinusoidal fit (see also section 7.1). The results of this measurement are $\mathcal{V} = \{97.1(2)\%, 96.6(2)\%, 96.0(2)\%\}$ for $N = \{380, 410, 440\}$ respectively, which are only marginally smaller than the results $\{98.4\%, 98.5\%, 98.3\%\}$ of the corresponding long axis measurement. The thin line in Fig. 6.3 shows the original sinusoidal fit (thick line) corrected by $\mathcal{V}^2 / \mathcal{V}_{\text{long}}^2$, which has a relatively minute effect. The remaining deviation can be explained by the finite θ -angle range of the Fisher information extraction in the following way. The functional form of the squared Hellinger distance of two Gaussian distributions is also Gaussian and obtains the form

$$d_{\text{H}}^2(\theta) = 1 - \exp\left(-\frac{F}{8}\theta^2\right). \quad (6.3)$$

We plug in the expected $F = N/\xi^2$, evaluate Eq. 6.3 at the experimental angle settings θ_i and fit quadratically for every tomography angle to obtain the curvature. This constitutes a noise-free 'simulation' of the experimental Hellinger method with the assumed Fisher information corresponding to the measured spin squeezing factor. The dashed lines in Fig. 6.3 show the results of this procedure, which agree very well with the results of the extracted Fisher

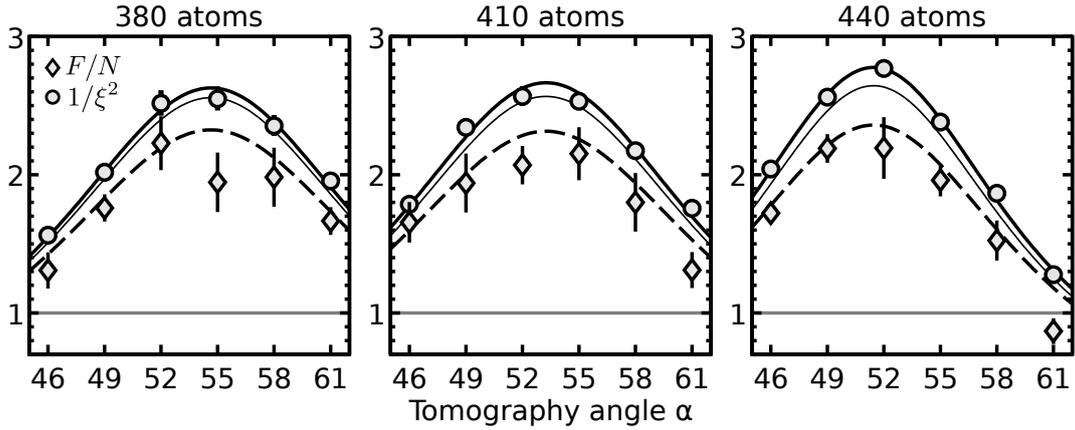


Figure 6.3 Fisher information of spin-squeezed states Here, we systematically compare the extracted Fisher information with the corresponding inverse spin squeezing factor for spin-squeezed states at 15 ms of evolution time. Both parameter surpass the standard quantum limit, which demonstrates entanglement, and share the same optimal tomography angle α . The thick solid line is a sinusoidal fit to $\xi^2 = \xi_N^2 / \mathcal{V}_{\text{long}}^2$. The thin solid line additionally includes the experimentally measured contrast during a full θ -rotation, which is only $\lesssim 2\%$ smaller than the prediction of the long axis. The smaller values of the Fisher information are the result of the finite θ -angle range of the extraction. Taking this into account allows to predict the behavior of the Fisher information from the spin squeezing results (dashed lines, see text).

information. The reason for this behavior is the reduced curvature of the Gaussian farther away from its maximum (minimum of the squared Hellinger distance), which leads to an underestimation of the infinitesimal curvature by the quadratic fit. This is a generic feature for peaked distributions (see also Fig. 2.10).

The results of the Hellinger distance method are still slightly smaller than this extrapolation, which shows that the extracted results for the Fisher information are conservative lower bounds. The detection of higher Fisher information values requires to measure with increasingly smaller angles θ_i . Experimentally, we have to make a tradeoff between signal and noise due to the finite number of samples for the Hellinger distances. If this inherent statistical noise is not the limiting factor, the smallest technologically reasonable angles should be chosen.

For non-Gaussian states an analytic form like Eq. 6.3 is not available, since the shape of the Hellinger distance depends on the shape of the probability distributions and their change in position and shape as a function of θ . Thus, a general statement about the required resolution in θ is not possible, but the Gaussian limit provides a good orientation.

6.3 Transition to non-Gaussian states

The developed methods for Fisher information extraction allow to experimentally characterize quantitatively the transition to the non-Gaussian regime. Here, the systematic variation of the tomography angle α is especially important, since the most sensitive direction (the direction of largest Fisher information) is expected to deviate from the direction of minimal variance, as discussed in section 2.6. Figure 6.4 compares and summarizes three such detailed experiments

performed around the evolution time of 25 ms, which yielded non-Gaussian states in the discussed tomographic characterization (see Fig.5.1). At the evolution time of 23 ms, the spin squeezing has already decreased compared to the optimal squeezing time of 15 ms. Both spin squeezing and Fisher information are still beyond the standard quantum limit. In the following evolution the spin squeezing decreases further, which is the result of the bending of the states. Increasingly larger tomography angles are required to reach minimal variance, which can be intuitively understood since we rotate clockwise while the state twists counterclockwise. The minimal variance is obtained when the additional 'tails' have the smallest projection along the readout direction (see also Figs. 5.2 and 2.12), while the highest Fisher information is reached, when the projection of the 'middle part' of the state has the smallest projection and the distributions feature significant tails (see Fig. 6.1). The maximal Fisher information at 25 ms is significantly larger than the maximally observed squeezing and the relative shift of the two optimal angles begins to become apparent. For 26 ms, this shift has increased further and the spin squeezing has completely vanished while the Fisher information is still beyond the standard quantum limit and signals entanglement. Figure 6.4 also shows the systematic behavior of the optimal angles with the total atom number at fixed evolution time, which we have discussed in the tomographic squeezing analysis (section 5.2).

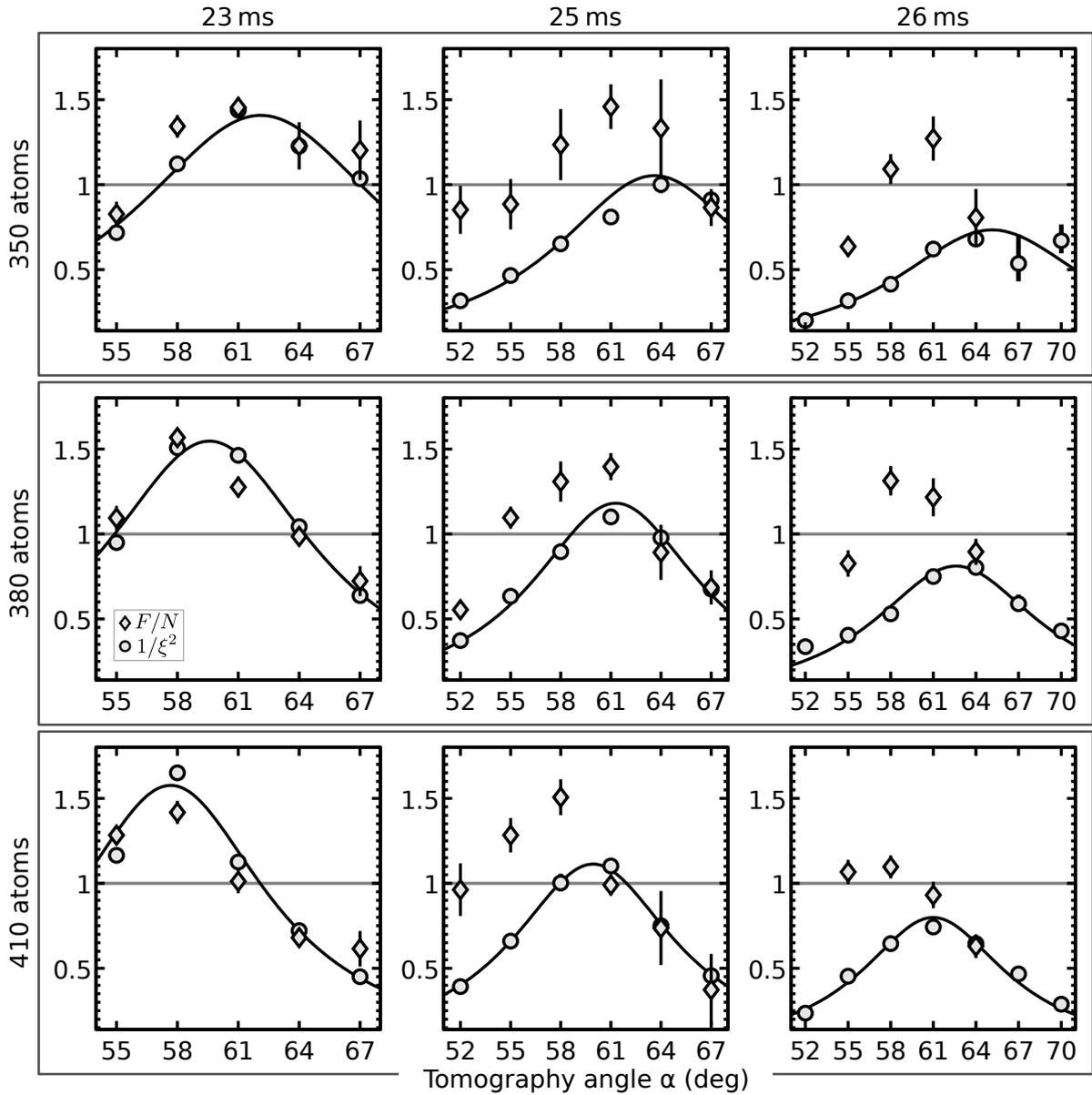


Figure 6.4 Fisher information at the transition to non-Gaussian states Here, we show the comparison between $1/\xi^2$ (circles with sinusoidal fits) and the normalized Fisher information F/N (diamonds) for three measurements performed at different evolution times in the vicinity of the transition to non-Gaussian states. The maximal Fisher information is found at increasingly smaller angles as the optimal squeezing, which is characteristic for the bend non-Gaussian states. While for 23 ms of evolution time spin squeezing and Fisher information are beyond the standard quantum limit ($F/N > 1$ and $1/\xi^2 > 1$), the squeezing vanishes in the course of the further evolution, whereas the Fisher information still signals entanglement in the non-Gaussian states.

7 Phase estimation

The extracted Fisher information of the generated states is a measure for their expected performance in a phase estimation task with unknown phase shift θ . In a conventional interferometric scheme, the phase shift inside the interferometer (e.g. the phase evolution during the Ramsey interrogation time) causes a shift of the mean value of the observable z on the readout axis after a projective measurement (in our case z is the atomic imbalance). With direct averaging of z from m repetitions, the precision for θ is given by the Gaussian error propagation

$$\Delta\theta_G = \frac{\Delta z}{\sqrt{m}} \left| \frac{\partial \langle z \rangle}{\partial \theta} \right|^{-1}. \quad (7.1)$$

Due to the central limit theorem, this is also true for non-Gaussian states in the case of many repetitions, where the distribution of the mean $\langle z \rangle$ becomes a Gaussian with the width $\sigma_{\langle z \rangle} = \Delta z / \sqrt{m}$. This is the general basis of classical statistical sampling to extract a mean value, which can be performed almost irrespective of the underlying distribution in the case of independent samples. In an interferometric context corresponding to our rotations on the Bloch sphere, the position $\langle z \rangle$ is a sinusoidal function of the phase, i.e. $\langle z \rangle = \mathcal{V} \sin(\theta - \theta_0)$ and the visibility \mathcal{V} of the interference fringe is the maximal speed of $\langle z \rangle$ as a function of the phase, i.e. $|\partial \langle z \rangle / \partial \theta| \leq \mathcal{V}$, with equality at the zero-crossings of the fringe. For states, whose probability distribution of single realizations is Gaussian (including spin squeezed states), Eq. 7.1 is already the limiting case since they are fully characterized by mean $\langle z \rangle$ and variance $(\Delta z)^2$. However, the probability distributions of non-Gaussian states can contain more information in which case the right-hand side of the Cramér-Rao-bound

$$\Delta\theta \geq \frac{1}{\sqrt{mF}} \quad (7.2)$$

can be smaller than $\Delta\theta_G$. To take advantage of the additional structure requires more advanced methods since averaging of z would lead back to the Gaussian limit $\Delta\theta_G$. In section 2.3 we have discussed two probabilistic methods to achieve this, namely the method of maximum likelihood and Bayesian analysis.

Here, we first demonstrate the Gaussian performance of a non-Gaussian state generated with 25 ms of nonlinear evolution time under direct averaging of the observable and then present a proof-of-principle implementation of a Bayesian phase-estimation protocol to overcome the limitations of the Gaussian protocol.

7.1 Gaussian performance

For the characterization of the Gaussian sensitivity after 25 ms of nonlinear evolution, we performed a full scan of the interferometry phase θ using the tomography angle $\alpha = 61^\circ$,

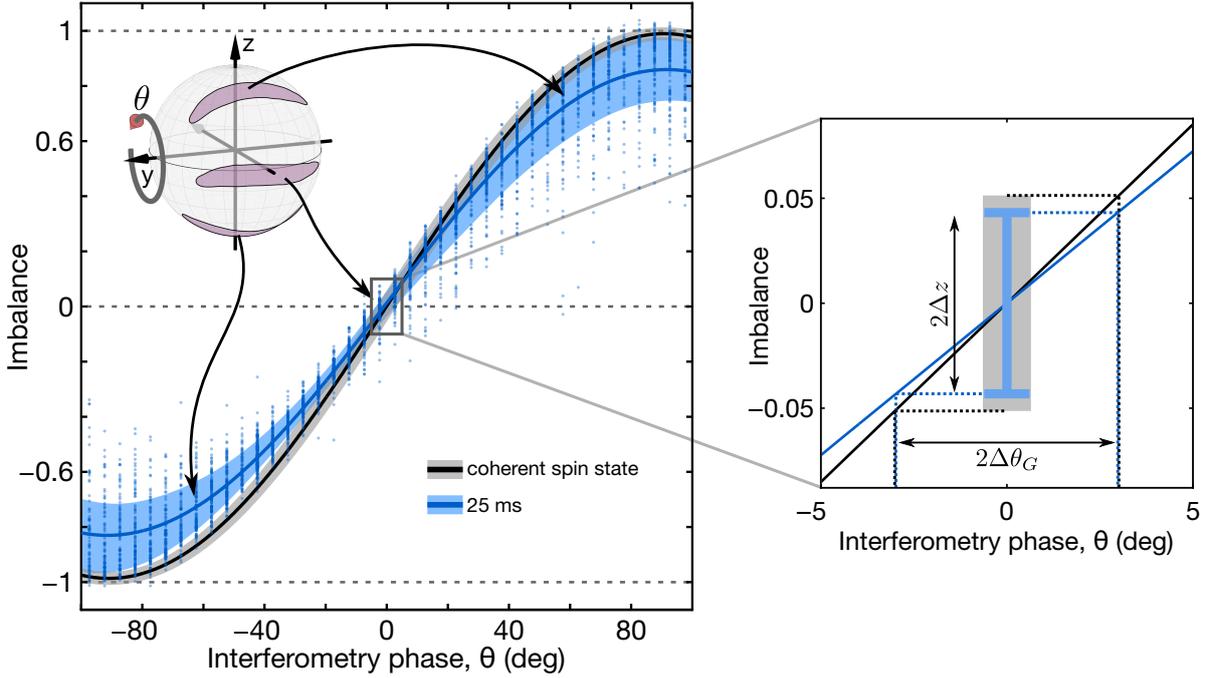


Figure 7.1 Gaussian sensitivity of a non-Gaussian state. Here, we show experimental results of a full scan of the phase θ , which is equivalent to a Ramsey fringe with phase shift θ acquired during the interrogation time (see Fig A.1). Due to the extension of the non-Gaussian quantum state, the visibility of the fringe as extracted from a sinusoidal fit is limited to 84.5(2)%. The blue points are raw experimental data and the blue region is their 68% confidence interval. Although the non-Gaussian state is still slightly number squeezed, its performance does not overcome the standard quantum limit in the Gaussian protocol. The right panel depicts the corresponding error propagation at the position of the steepest slope of the fringe, in which the dotted vertical lines mark the Gaussian phase uncertainty region. The black line and gray band is the result of a reference measurement with a coherent spin state of the same atom number. For this state, the obtained visibility is 98.8(2)%.

where we have detected the smallest variance (i.e. the largest value of $1/\xi^2$, see Fig. 6.4). We concentrate here on the total atom number $N = 380$. The results of this measurement are shown in Fig. 7.1 together with a reference measurement with a coherent spin state of the same atom number at 9.2 G, farther away from the Feshbach resonance, where influences of the nonlinearity are negligible. The coherent spin state obtains a visibility of $\mathcal{V} = 98.8(2)\%$ as extracted from a sinusoidal fit, whereas the extended non-Gaussian state is limited to $\mathcal{V} = 84.5(2)\%$, which reduces the steepness of the slope at the zero-crossing. Although its variance is slightly below the coherent state level (number squeezing factor $\xi_N^2 = -1.5(5)$ dB in this measurement), the phase estimation performance is barely at the standard quantum limit ($\xi_N^2/\mathcal{V}^2 = 0.0(5)$ dB). The zoom of Fig. 7.1 shows this situation graphically. For the coherent spin state, we have subtracted the detection noise of $\Delta z_{\text{det}} \approx 6.2/N$ here, with which its number squeezing factor is at 0.0(2) dB. Due to the slight technical reduction of the visibility and the additional detection noise the coherent spin state performs at +0.5(2) dB, which is 6% worse than the standard quantum limit. In this respect, the number squeezing of this particular non-Gaussian state still allows to reach the standard quantum limit in a Gaussian protocol (by averaging of many realizations) in the presence of our detection noise, but not to go beyond.

In terms of Gaussian phase sensitivity, this corresponds to $\Delta\theta_G = 2.9(2)^\circ/\sqrt{m}$, while the coherent spin state with detection noise is at $3.11(7)^\circ/\sqrt{m}$.

7.2 Bayesian phase sensitivity

The general Bayesian scheme (see section 2.3) employs previous knowledge about the underlying probability distributions to efficiently infer the most probable phase shift from a row of independently acquired experimental samples $\{z_1 \dots z_m\}$. In particular, it boils down to the calculation of the (log-)likelihood function

$$\log \mathcal{L}(\theta) = \log \prod_{j=1}^m P_{z_j}(\theta) = \sum_{j=1}^m \log P_{z_j}(\theta). \quad (7.3)$$

This corresponds to a probabilistic assessment of the measured values with the help of the knowledge of the probabilities of z as a function of the phase: $\mathcal{L}(\theta)$ is the total probability to obtain $\{z_1 \dots z_m\}$ in the case of the phase setting θ . For increasing m , the likelihood function obtains a Gaussian shape $\mathcal{L} \propto \exp(-(\theta - \theta_c)^2/2\sigma^2)$, peaked at the most probable phase θ_c (see section 2.3). Thus, the log-likelihood function becomes quadratic:

$$\log \mathcal{L}(\theta) \rightarrow -\frac{(\theta - \theta_c)^2}{2\sigma^2} + \text{const.} \quad (7.4)$$

The Bayesian uncertainty σ is related to the Fisher information and we expect $\sigma^2 \rightarrow 1/mF$, which we want to verify. Our previous knowledge which we want to employ for phase estimation consists of the experimental probability distributions $\mathcal{F}_z(\theta_i)$ for the small discrete phases θ_i , which have been acquired for the extraction of the Fisher information with the Hellinger distance method and constitute our best guess for the underlying probabilities $P_z(\theta_i)$. To simplify the notation, we make no distinction between $\mathcal{F}_z(\theta_i)$ and $P_z(\theta_i)$ in the following.

A problem in using this approach is that the measured histograms also contain bins with zero counts, which are interpreted as zero probability. In calculating the likelihood function, this leads to zero total probability or $\log \mathcal{L}(\theta_i) = -\infty$ for the whole sequence $\{z_1, \dots, z_m\}$, which we have to avoid. Our approach is to selectively discard these realizations while *not* decreasing the value of m for the evaluation of the sensitivity. In this way, we obtain a model-independent scheme for Bayesian phase estimation.

Our recipe thus consists of the following steps:

1. Acquire m experimental realizations $\{z_1 \dots z_m\}$.
2. Discard all z_j for which $P_{z_j}(\theta_i) = 0$ for particular θ_i .
3. Calculate $\log \mathcal{L}(\theta_i) = \sum_{j=1}^m \log P_{z_j}(\theta_i)$ for all θ_i where experimental probability distributions are available.
4. Fit $\log \mathcal{L}(\theta_i)$ quadratically to extract σ^2 .

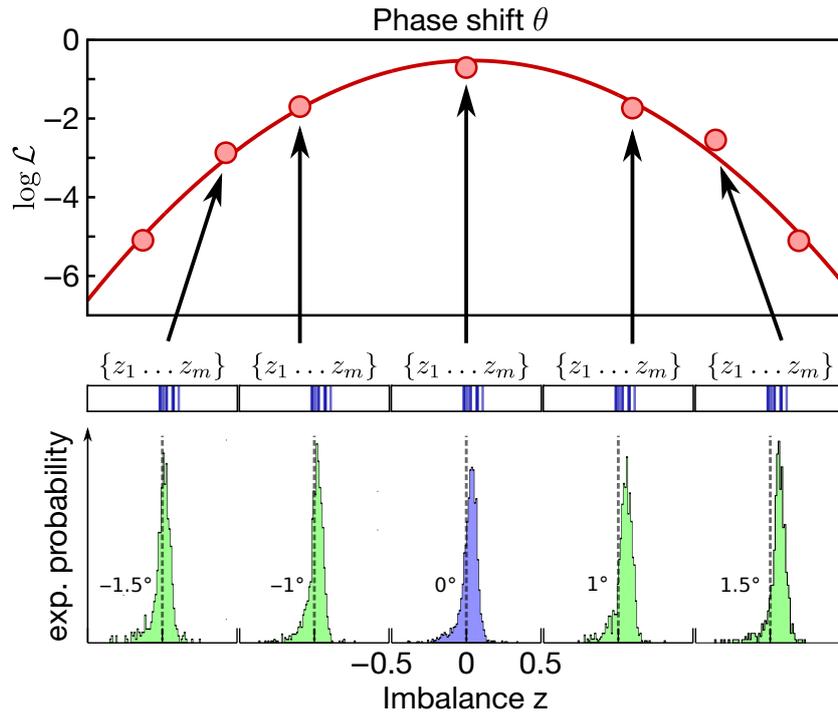


Figure 7.2 Bayesian phase estimation scheme. Here, we sketch our model-independent scheme for Bayesian phase estimation. With the help of the experimental probability distributions (bottom row) an independently acquired m -sequence (middle row for $m = 10$) with unknown phase setting can be probabilistically assessed by calculating the total experimental probability (likelihood $\mathcal{L}(\theta)$) to obtain $\{z_1 \dots z_m\}$ under the assumption of the individual phase settings. For m sufficiently large, $\mathcal{L}(\theta)$ becomes a Gaussian function and the corresponding Bayesian sensitivity can be extracted for each m -sequence by a quadratic fit to $\log \mathcal{L}(\theta)$ (depicted in the upper panel). By repeating this procedure, the center and the curvature of the fit statistically fluctuate. The shown example was obtained for $N = 380$, $\alpha = 58^\circ$ and 26 ms of nonlinear evolution time.

This procedure is graphically depicted in Fig. 7.2. What we have neglected thus far is the possible inclusion of previous knowledge of the phase shift in Bayesian estimation. In our case, this does not make sense, since we want to experimentally simulate the situation without previous knowledge, even though in our case we know with which phase setting $\{z_1 \dots z_m\}$ was acquired.

For our proof-of-principle demonstration, we randomly choose 1000 samples of the reference setting $\theta_{\text{ref}} = 0$ from the Hellinger distance measurements, which was experimentally sampled about $4\times$ as often as the other θ_i . The rest of these samples is used for the experimental probability distribution $P_z(0)$. The 1000 samples are then randomly divided into m -sequences $\{z_1 \dots z_m\}$ to extract the Bayesian sensitivity for each sequence. We typically have to discard less than 10% of the realizations to keep the likelihood finite for all θ_i . We note, that it is important to discard *after* the subdivision into m -sequences, since otherwise all remaining samples would become valid, which is not the realistic situation. In our dismissal approach, the reduced number of valid samples leads to a reduced sensitivity, since we judge the result as coming from a full m -sequence.

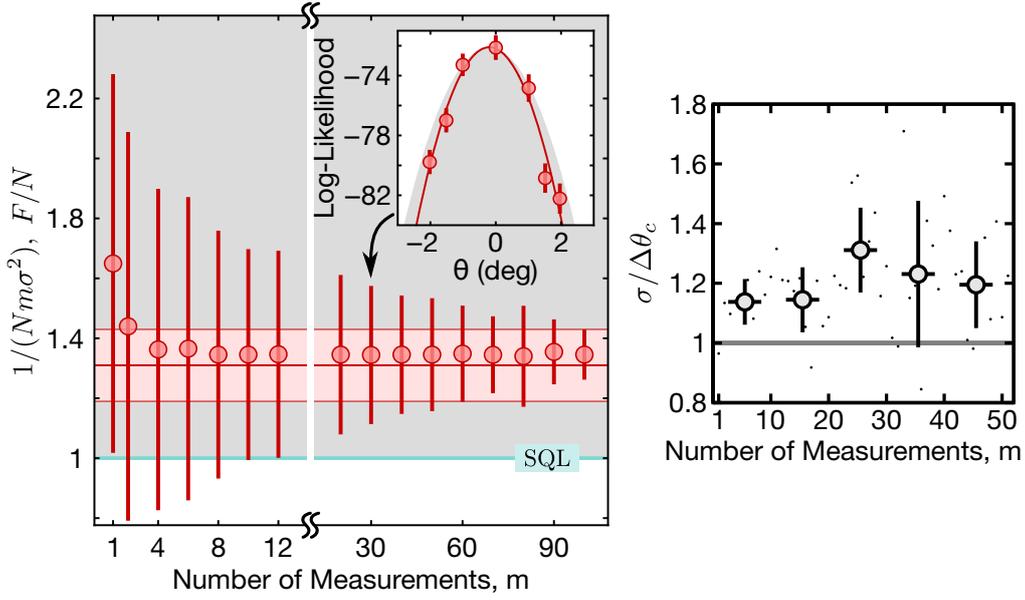


Figure 7.3 Bayesian phase estimation sensitivity. Here, we employ our Bayesian scheme in the non-Gaussian regime for $N = 380$ atoms and the tomography rotation angle $\alpha = 58^\circ$ and 25 ms of evolution time. The left panel shows the extracted sensitivities versus size of the m -sequences, where we normalized out the classical scaling $\sigma \propto \sqrt{m}$. We find a fast convergence of $1/m\sigma^2$ to the independently extracted Fisher information (red line with 68% confidence region), which is beyond the standard quantum limit (SQL, marked as line). On average, the Bayesian scheme allows to harness the non-Gaussian resource already for $m \gtrsim 2$. The error bars indicate the standard deviation of a single m -sequence. Consistent values can be obtained by first averaging the log-likelihoods and then performing the quadratic fit (shown in the inset for $m = 30$). All gray shaded areas are only accessible with entanglement in the system. The right panel shows a comparison of the mean Bayesian sensitivity σ to the statistical scatter of the center positions θ_c of the single fits. Their ratio is flat versus m and σ is consistently bigger than the statistical spread $\Delta\theta_c$ and thus a robust estimate of the sensitivity.

For small m , the center θ_c and the curvature $1/\sigma^2$ of the fits fluctuate according to the Bayesian uncertainty. They narrow and settle as m is increased. To compare this systematic behavior with the Fisher information, we normalize out the classical scaling $\sigma \propto \sqrt{m}$ and compare the values

$$\frac{1}{Nm\sigma^2} \leftrightarrow \frac{F}{N} \quad (7.5)$$

as a function of m . This is shown in Fig. 7.3 for the tomography angle $\alpha = 58^\circ$ and 25 ms of evolution time. For each value of m we extract the mean and standard deviation of σ^2 from the $\sim 1000/m$ repetitions of the Bayesian estimation. We always work with the same 1000 samples to obtain consistent conditions, especially to employ the same set of probability distributions for all m . Thus, the points in Fig. 7.3 are not statistically independent. We find a fast convergence close to the result $F/N = 1.31(12)$ extracted with the Hellinger distance method at this point. Note that with this choice of the tomography angle α , the Gaussian sensitivity is worse than the standard quantum limit (see Fig. 6.4). The performance here is $\sim 15\%$ beyond the standard quantum limit in the presence of all technical and physical noise sources, equivalent to a phase sensitivity of $2.5(1)^\circ/\sqrt{m}$.

In Fig. 7.3 (right) we additionally compare the statistical 68% confidence interval $2\Delta\theta_c$ for the fitted centers θ_c of the $1000/m$ repetitions with the mean of the Gaussian width σ of the likelihood function. The fraction $\sigma/\Delta\theta_c$ is flat versus m and robustly above one, which means that the Bayesian uncertainty σ is larger than the statistical scatter of θ_c and thus represents a robust estimate of the sensitivity.

We have employed this procedure for other interesting points and atom numbers (see Fig. 6.4), where we expected non-Gaussian performance beyond the standard quantum limit and got consistent results. We point out two examples: $\{N, \alpha, F/N\} = \{380, 58^\circ, 1.31(9)\}$ with 26 ms of evolution time, where we got a corresponding Bayesian sensitivity of $1/Nm\sigma^2 = 1.24(4)$ and $\{N, \alpha, F/N\} = \{410, 58^\circ, 1.51(10)\}$ at 25 ms, where we obtained $1/Nm\sigma^2 = 1.67(3)$.

In summary, we have reached Bayesian phase sensitivity of our model-independent protocol beyond the standard quantum limit in accordance with the independently extracted Fisher information in a regime where this is not attainable with a Gaussian protocol.

8 Outlook

In this thesis, we have demonstrated the feasibility of a novel method based on the statistical distance of experimental probability distributions to extract the Fisher information, which is a lower bound for the Quantum Fisher information. Using this method we have verified entanglement in Gaussian as well as non-Gaussian spin states of binary Bose-Einstein condensates with a Fisher information that exceeds the number of atoms in the detected ensembles. In the absence of spin squeezing, we have confirmed the implied resource of non-Gaussian states for quantum-enhanced phase estimation with a Bayesian protocol. It showed a sensitivity beyond the standard quantum limit in accordance with the extracted Fisher information, whereas this was not attainable with the Gaussian protocol, i.e. averaging of the observable.

Unlike state reconstruction approaches our method is not limited to small particle numbers and is broadly applicable to other many-body entanglement scenarios, since it does not depend on the special shape of the probability distributions. Very recently our work has inspired a remarkable theoretical proposal to characterize the entanglement close to quantum phase transitions [160] and was applied to a novel many-ion experiment in Penning traps [161].

Currently, the generation and application of atomic non-Gaussian many-body states is at the stage of proof-of-principle demonstrations [28, 108, 109, 162] and the route towards metrological sensitivity beyond the level of the best spin-squeezed states is challenging. For example, we have seen theoretically that in our system even tiny detection inefficiencies on the single particle level are sufficient to decrease the advantage of the non-Gaussian states considerably, even in an otherwise perfect system. An interesting pathway to mitigate this detection problem is the use of time-reversal sequences [18] to disentangle the state before detection, since for metrology the entanglement is only required at the stage of the phase transformation [17]. Besides the challenge to implement such schemes with sufficient fidelity, atomic losses during state preparation and the linear phase shift are still considerable limitations.

An interesting alternative method for the generation of non-Gaussian states with BECs is the process of spin-changing collisions [163–166], which is analogous to parametric down-conversion in optics. Interestingly, it resembles the two-mode bifurcation scenario [167] and allows to produce two-mode squeezed vacuum states and three-mode non-Gaussian states [168]. The advantage is, that the influence of losses is reduced. However, the applicability of the non-Gaussian states for metrology is hindered by the complex three-mode situation.

From a fundamental point of view, our experimental system and the developed methods offer future possibilities to study the entanglement generation close to classically unstable points or chaotic regions [169] and the role of decoherence in such situations [71]. The discussed bifurcation scenario is connected to a quantum phase transition in a one-dimensional situation [170], where the relation between parameter estimation and state distinguishability could elucidate the connection between criticality and the growth of entanglement [160, 171].

A Supplementary

A.1 Quantum Fisher Information

Here, we detail the derivation of the Quantum Fisher information for general mixed states and give its explicit form in the case of unitary transformations. In a quantum parameter estimation setting, the state of a quantum mechanical system described by the statistical operator $\hat{\rho}_0$ (input state) is changed by some transformation into ρ (the output state), parametrized by the parameter θ . We are interested in the precision of an estimator of θ , which is described by a quantum mechanical operator $\hat{\theta}_{\text{est}}$. In general, the estimator can be biased, such that its expectation value $\langle \hat{\theta}_{\text{est}} \rangle$ differs from the true value by

$$b(\theta) = \langle \hat{\theta}_{\text{est}} \rangle - \theta = \text{Tr} \left[\hat{\rho} (\hat{\theta}_{\text{est}} - \theta) \right]. \quad (\text{A.1})$$

Differentiation with respect to θ and using $\text{Tr} [\hat{\rho}] = 1$ yields

$$(1 + b'(\theta)) = \text{Tr} \left[(\hat{\theta}_{\text{est}} - \theta) \frac{\partial \hat{\rho}}{\partial \theta} \right]. \quad (\text{A.2})$$

For the following a self-adjoint operator \hat{L} is introduced, which is called Symmetric Logarithmic Derivative (SLD), as the solution of the operator equation

$$\frac{\partial \hat{\rho}}{\partial \theta} = \frac{\hat{\rho} \hat{L} - \hat{L} \hat{\rho}}{2}. \quad (\text{A.3})$$

With this, one obtains

$$\begin{aligned} (1 + b'(\theta))^2 &= \left(\frac{1}{2} \text{Tr} \left[((\hat{\theta}_{\text{est}} - \theta) \hat{L} \hat{\rho})^\dagger \right] + \frac{1}{2} \text{Tr} \left[(\hat{\theta}_{\text{est}} - \theta) \hat{L} \hat{\rho} \right] \right)^2 \\ &= \left(\text{Re Tr} \left[(\hat{\theta}_{\text{est}} - \theta) \hat{L} \hat{\rho} \right] \right)^2 \leq \left| \text{Tr} \left[\hat{L} \sqrt{\hat{\rho}} \sqrt{\hat{\rho}} (\hat{\theta}_{\text{est}} - \theta) \right] \right|^2 \end{aligned} \quad (\text{A.4})$$

since $\text{Tr} [A] + \text{Tr} [A^\dagger] = 2 \text{Re Tr} [A]$, where Re denotes the real part. Since $\text{Tr} [A^\dagger B]$ is a proper choice as inner product of Hilbert space operators (called Hilbert–Schmidt inner product), the Cauchy-Schwarz inequality holds which reads $|\text{Tr} [A^\dagger B]|^2 \leq \text{Tr} [A^\dagger A] \text{Tr} [B^\dagger B]$. Applied to Eq. A.4 this yields

$$\left| \text{Tr} \left[\hat{L} \sqrt{\hat{\rho}} \sqrt{\hat{\rho}} (\hat{\theta}_{\text{est}} - \theta) \right] \right|^2 \leq \text{Tr} \left[\hat{L} \sqrt{\hat{\rho}} \sqrt{\hat{\rho}} \hat{L} \right] \text{Tr} \left[(\hat{\theta}_{\text{est}} - \theta) \sqrt{\hat{\rho}} \sqrt{\hat{\rho}} (\hat{\theta}_{\text{est}} - \theta) \right] \quad (\text{A.5})$$

$$= \text{Tr} \left[\hat{\rho} \hat{L}^2 \right] \text{Tr} \left[\hat{\rho} (\hat{\theta}_{\text{est}} - \theta)^2 \right] \quad (\text{A.6})$$

In this context, $\sqrt{\hat{\rho}}$ is a short-hand notation for the self-adjoint operator \hat{O} , which solves $\hat{O}^2 = \hat{\rho}$. The rightmost term corresponds to the variance $(\Delta\hat{\theta}_{\text{est}})^2 = \langle (\hat{\theta}_{\text{est}} - \theta)^2 \rangle$ of the estimator and $F_Q = \text{Tr} [\hat{\rho} \hat{L}^2]$ is called the Quantum Fisher Information. Together with Eq. A.4 the Quantum-Cramér-Rao bound

$$(\Delta\hat{\theta}_{\text{est}})^2 \geq \frac{(1 + b'(\theta))^2}{F_Q} \quad (\text{A.7})$$

is obtained. It characterizes the maximally achievable precision of the estimator given the input state ρ . If the estimator is unbiased, i.e. $b(\theta) = 0$, the variance is simply bound by $1/F_Q$. So far, the treatment is very general, as no special form (besides being self-adjoint) of the transformation was assumed. Now, we will focus on unitary transformations, as this is the form of the time evolution of a quantum mechanical system and derive the explicit form of \hat{L} and F_Q in this case. The transformation of the density operator is governed by the von Neumann equation

$$i \frac{\partial \hat{\rho}}{\partial \theta} = [\hat{H}, \hat{\rho}], \quad (\text{A.8})$$

which has the solution $\hat{\rho}(\theta) = \exp(-i\theta\hat{H})\hat{\rho}(0)\exp(i\theta\hat{H})$. With this, the condition for the symmetric logarithmic derivative becomes

$$\hat{\rho}\hat{L} + \hat{L}\hat{\rho} = 2i[\hat{\rho}, \hat{H}] \quad (\text{A.9})$$

To obtain an explicit form for both sides, we work in a basis that diagonalizes $\hat{\rho} = \sum_k p_k |k\rangle\langle k|$, in which we get the matrix form $\hat{L} = \sum_{l,m} L_{lm} |l\rangle\langle m|$ with the matrix elements $L_{lm} = \langle l | \hat{L} | m \rangle$. In this representation we get

$$\begin{aligned} \hat{\rho}\hat{L} + \hat{L}\hat{\rho} &= \sum_{j,k,l,m} (p_j |j\rangle\langle j| L_{lm} |l\rangle\langle m| + L_{lm} |l\rangle\langle m| p_k |k\rangle\langle k|) \\ &= \sum_{j,k,l,m} (p_j L_{lm} \delta_{jl} |j\rangle\langle m| + p_k L_{lm} \delta_{mk} |l\rangle\langle k|) \\ &= \sum_{l,m} (p_l + p_m) L_{lm} |l\rangle\langle m|. \end{aligned} \quad (\text{A.10})$$

In exactly the same way we calculate for the commutator

$$\hat{\rho}\hat{H} - \hat{H}\hat{\rho} = \sum_{l,m} (p_l - p_m) H_{lm} |l\rangle\langle m|. \quad (\text{A.11})$$

By comparing the coefficients, one therefore obtains together with Eq. A.9

$$L_{lm} = 2i \frac{(p_l - p_m)}{(p_l + p_m)} H_{lm} \quad (\text{A.12})$$

for $(p_l + p_m) \neq 0$. With the help of this identity, we can finally calculate the quantum Fisher information as

$$\begin{aligned}
F_Q &= \text{Tr} [\hat{\rho} \hat{L}^2] = \sum_j \langle j | \hat{\rho} \hat{L} \hat{L} | j \rangle \\
&= \sum_{j,k,l,m,r,s} \langle j | p_k | k \rangle \langle k | L_{lm} | l \rangle \langle m | L_{rs} | r \rangle \langle s | | j \rangle \\
&= \sum_{j,k,l,m,r,s} p_k L_{lm} L_{rs} \delta_{jk} \delta_{kl} \delta_{mr} \delta_{sj} = \sum_{km} p_k L_{km} L_{mk} \\
&= 4 \sum_{k,m} p_k \frac{(p_k - p_m)^2}{(p_k + p_m)^2} |H_{km}|^2 \tag{A.13}
\end{aligned}$$

$$= 2 \sum_{k,m} (p_k + p_m) \frac{(p_k - p_m)^2}{(p_k + p_m)^2} |H_{km}|^2 \tag{A.14}$$

Thus, we obtain the Quantum Fisher information in the usual form [9, 95, 172]

$$F_Q[\hat{\rho}, \hat{H}] = 2 \sum_{k,m} \frac{(p_k - p_m)^2}{(p_k + p_m)} |\langle k | \hat{H} | m \rangle|^2 \tag{A.15}$$

The same form can be obtained by optimizing the Fisher information with respect to all positive operator valued measurements (POVM) allowed by quantum mechanics [95]. As such, it is an upper bound for the achievable precision and the Fisher information for a specific POVM \hat{E} is always smaller or equal, i.e. $F[\hat{\rho}, \hat{H}, \hat{E}] \leq F_Q[\hat{\rho}, \hat{H}]$. If the transformation is not restricted to being unitary and the probabilities p_k can also be subject of change, the most general form [95] reads

$$F_Q[\hat{\rho}, \hat{H}] = \sum_k \frac{1}{p_k} \left(\frac{\partial p_k}{\partial \theta} \right)^2 + 2 \sum_{k,m} \frac{(p_k - p_m)^2}{(p_k + p_m)} |\langle k | \hat{H} | m \rangle|^2 \tag{A.16}$$

The additional term has the form of the "classical" Fisher information and limits the attainable precision in the case of a classical system, where $\hat{H} = 0$. There, the probability distribution has to change in order to extract information about the parameter.

For pure states, i.e. $\hat{\rho} = |\psi\rangle\langle\psi|$, the QFI can be further simplified. In this case, the eigenbasis of $\hat{\rho}$ consists of $|1\rangle = |\psi\rangle$ with eigenvalue 1 and the degenerate orthogonal subspace $\sum_{k>1} |k\rangle\langle k| = \mathbb{1} - |\psi\rangle\langle\psi|$ with eigenvalue 0. Therefore, Eq. A.15 can be split in two parts with $(p_k = 0, p_m = 1)$ and vice versa. One obtains

$$\begin{aligned}
F_Q[\hat{\rho}, \hat{H}] &= 2 \sum_{m>1} \langle m | \hat{H} | 1 \rangle \langle 1 | \hat{H} | m \rangle + 2 \sum_{k>1} \langle 1 | \hat{H} | k \rangle \langle k | \hat{H} | 1 \rangle \\
&= 4 \sum_{m>1} \langle 1 | \hat{H} | m \rangle \langle m | \hat{H} | 1 \rangle \\
&= 4 \langle \psi | \hat{H} (\mathbb{1} - |\psi\rangle\langle\psi|) \hat{H} | \psi \rangle \tag{A.17}
\end{aligned}$$

$$= 4(\langle \hat{H}^2 \rangle - \langle \hat{H} \rangle^2) = 4(\Delta \hat{H})^2 \tag{A.18}$$

Thus, for pure states it is clear that high Fisher information is obtained for transformations, whose generators have high variance. For example, to be most sensitive to rotations around the z -axis (generator $\hat{H} = \hat{J}_z$), a pure state has to maximize the variance of \hat{J}_z .

A.2 Coupled basis and singlet states

Here, we list a full coupled basis set for the Hilbert space of 3 qubits and the two orthogonal singlet states of 4 qubits. This illustrates the rich structure besides the fully symmetric subspace. The coupled basis is obtained by successively coupling spin-1/2 to bigger spins with the help of Clebsch-Gordan coefficients. We begin with two qubits, which can be coupled to

$$S_2 = 0 \left\{ m_{S_2} = 0 \quad \frac{1}{\sqrt{2}}(|\downarrow\uparrow\rangle - |\uparrow\downarrow\rangle) \right. \quad (\text{A.19})$$

$$S_2 = 1 \left\{ \begin{array}{l} m_{S_2} = +1 \quad |\uparrow\uparrow\rangle \\ m_{S_1} = 0 \quad \frac{1}{\sqrt{2}}(|\downarrow\uparrow\rangle + |\uparrow\downarrow\rangle) \\ m_{S_1} = -1 \quad |\downarrow\downarrow\rangle \end{array} \right. \quad (\text{A.20})$$

S_2 in turn can be coupled to the third qubit in the following ways:

$$S_2 = 0, S_3 = 1/2 \left\{ \begin{array}{l} m_{S_3} = +\frac{1}{2} \quad \frac{1}{\sqrt{2}}(|\downarrow\uparrow\uparrow\rangle - |\uparrow\downarrow\uparrow\rangle) \\ m_{S_3} = -\frac{1}{2} \quad \frac{1}{\sqrt{2}}(|\downarrow\uparrow\downarrow\rangle - |\uparrow\downarrow\downarrow\rangle) \end{array} \right. \quad (\text{A.21})$$

$$S_2 = 1, S_3 = 1/2 \left\{ \begin{array}{l} m_{S_3} = +\frac{1}{2} \quad \sqrt{\frac{2}{3}}|\uparrow\uparrow\downarrow\rangle - \sqrt{\frac{1}{6}}(|\uparrow\downarrow\uparrow\rangle + |\downarrow\uparrow\uparrow\rangle) \\ m_{S_3} = -\frac{1}{2} \quad \sqrt{\frac{2}{3}}|\downarrow\downarrow\uparrow\rangle - \sqrt{\frac{1}{6}}(|\downarrow\uparrow\downarrow\rangle + |\uparrow\downarrow\downarrow\rangle) \end{array} \right. \quad (\text{A.22})$$

$$S_2 = 1, S_3 = 3/2 \left\{ \begin{array}{l} m_{S_3} = +\frac{3}{2} \quad |\uparrow\uparrow\uparrow\rangle \\ m_{S_3} = +\frac{1}{2} \quad \frac{1}{\sqrt{3}}(|\downarrow\uparrow\uparrow\rangle + |\uparrow\downarrow\uparrow\rangle + |\uparrow\uparrow\downarrow\rangle) \\ m_{S_3} = -\frac{1}{2} \quad \frac{1}{\sqrt{3}}(|\uparrow\downarrow\downarrow\rangle + |\downarrow\uparrow\downarrow\rangle + |\downarrow\downarrow\uparrow\rangle) \\ m_{S_3} = -\frac{3}{2} \quad |\downarrow\downarrow\downarrow\rangle \end{array} \right. \quad (\text{A.23})$$

Thus, for three qubits, there are two orthogonal subspaces with $S_3 = 1/2$ but no singlet state. For four qubits however, there are two orthogonal singlet states:

$$S_2 = 0, S_3 = 1/2, S_4 = 0 \left\{ m_{S_4} = 0 \quad \frac{1}{2}(|\downarrow\uparrow\downarrow\uparrow\rangle - |\uparrow\downarrow\downarrow\uparrow\rangle - |\downarrow\uparrow\uparrow\downarrow\rangle + |\uparrow\downarrow\uparrow\downarrow\rangle) \right. \quad (\text{A.24})$$

$$S_2 = 1, S_3 = 1/2, S_4 = 0 \left\{ \begin{array}{l} m_{S_4} = 0 \quad \frac{1}{\sqrt{3}}(|\downarrow\downarrow\uparrow\uparrow\rangle - |\uparrow\uparrow\downarrow\downarrow\rangle) - \frac{1}{\sqrt{12}}(|\downarrow\uparrow\downarrow\uparrow\rangle - |\uparrow\downarrow\uparrow\downarrow\rangle) \\ \quad - \frac{1}{\sqrt{12}}(|\uparrow\downarrow\downarrow\uparrow\rangle - |\downarrow\uparrow\uparrow\downarrow\rangle) \end{array} \right. \quad (\text{A.25})$$

A.3 SU(2) interferometer and rotations on the Bloch-sphere

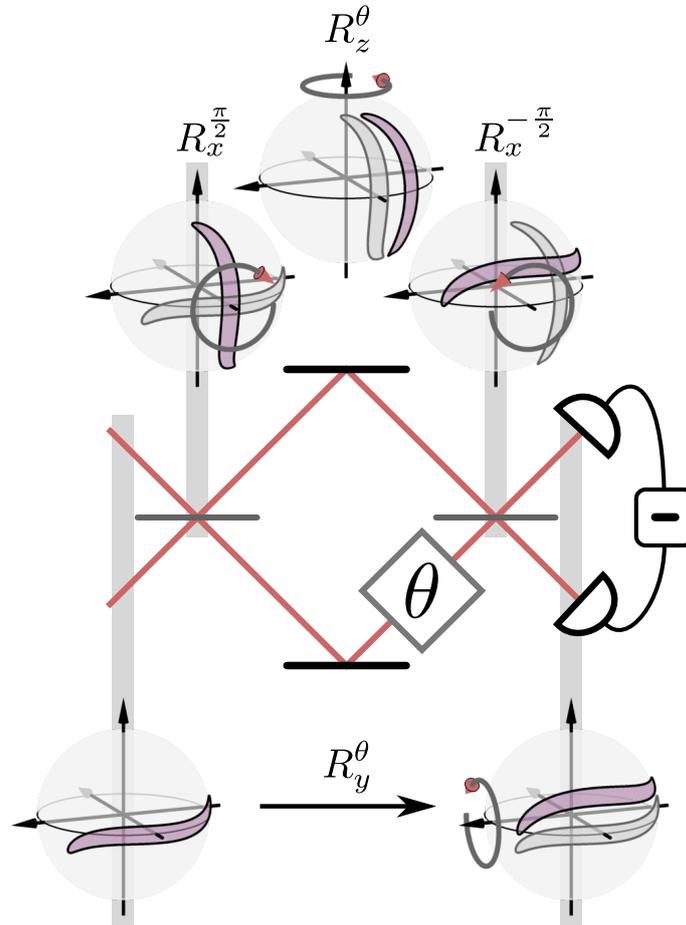


Figure A.1 Linear SU(2) interferometer and rotations on the Bloch sphere. The Mach-Zehnder interferometer (shown as sketch) is the optical analog of the atomic Ramsey sequence. Balanced beamsplitters ($\pi/2$ -pulses) perform rotations R_x by an angle of $\pi/2$ and the relative phase shift inside the interferometer performs a rotation R_z^θ by the angle θ . $R_{x,y,z}$ are the corresponding rotation matrices of SO(3). The sequence $R_x^{-\pi/2} R_z^\theta R_x^{\pi/2}$ is mathematically equivalent to R_y^θ . For a characterization of the interferometric sensitivity it is thus sufficient to apply R_y^θ only. (Figure from the supplementary information of our publication [21])

A.4 Hyperfine structure in a magnetic field

The Breit-Rabi formula is the solution to the eigenvalue problem of the hyperfine interaction Hamiltonian which obtains a surprisingly simple form [141] and reads for our particular case of nuclear spin $I = 3/2$

$$E_{\text{BR}} = -\frac{E_{\text{hfs}}}{8} + m_F g_I \mu_B B \pm \frac{E_{\text{hfs}}}{2} \sqrt{1 + m_F x + x^2}, \quad (\text{A.26})$$

with the argument $x = (g_J - g_I) \mu_B B / E_{\text{hfs}}$. For high fields, there is a sign problem, when the Zeeman shift of the state $m_F = -2$ is larger than the hyperfine splitting E_{hfs} . In this case, the square root has to be replaced by $(1 - x)$. A quadratic expansion of this formula for small x , that is small magnetic fields, yields (without the constant first term)

$$E_{\text{BR}} \approx m_F g_I \mu_B B \pm \frac{E_{\text{hfs}}}{2} \left(1 + \frac{m_F}{2} x + \frac{4 - m_F^2}{8} x^2 \right) \quad (\text{A.27})$$

and thus the linear and quadratic contributions to the Zeeman shift

$$E_Z \approx p_Z m_F B + q_Z (4 - m_F^2) B^2 \quad (\text{A.28})$$

with the parameters

$$p_Z = \pm \left(\frac{g_J - g_I}{4} \pm g_I \right) \mu_B \approx \begin{cases} +699.583 \text{ kHz/G} & \text{for } F = 2 \\ -702.369 \text{ kHz/G} & \text{for } F = 1 \end{cases} \quad (\text{A.29})$$

$$q_Z = \pm \left(\frac{g_J - g_I}{4} \right)^2 \frac{\mu_B^2}{E_{\text{hfs}}} \approx \pm 71.9 \text{ Hz/G}^2 \quad (\text{A.30})$$

For reference, we list the most recent constants with their sources in units of frequency:

μ_B	Bohr magneton	1.3996245042(86) MHz/G	CODATA 2014
g_S	Electron spin g-factor	2.00231930436182(52)	CODATA 2014
g_J	Bound electron g-factor in $5S_{1/2}$	$g_S \times (1 + 5.9(1) \times 10^{-6})$	[173]
g_I	Nuclear g-factor	$-g_J \times 4.9699147(50) \times 10^{-4}$	[174]
E_{hfs}	Zero-field hyperfine splitting	6 834 682 610.904 310(3) Hz	[175]

B Additional characterizations and technicalities

B.1 Characterization of the optical dipole traps

For optical trapping, we use focused laser beams with a wavelength of 1030 nm (waveguide and XDT) and an interference pattern of two crossing beams at 820 nm (one dimensional optical lattice). Both are red detuned to the D1 and D2 lines at 795 nm and 780 nm, respectively. The XDT is used to enhance thermalisation during evaporation to BEC and switched off completely once the BEC is loaded into the optical lattice.

Waveguide

The waveguide potential is created by a single focused laser beam with approximately Gaussian profile. It creates a potential of the form

$$V(x, r) = \frac{-V_0 w_0^2}{w^2(x)} \exp \left[-2 \frac{r^2}{w^2(x)} \right] \quad (\text{B.1})$$

with the radial coordinate $r = \sqrt{x^2 + y^2}$ and the x -dependent waist ($1/e^2$ -radius) $w(x) = w_0 \sqrt{1 + (x/x_R)^2}$. The Rayleigh range $x_R = \pi w_0^2 / \lambda$ marks the distance from the focus, where the beam waist has grown by a factor $\sqrt{2}$. Expanding quadratically in either direction around the minimum yields the harmonic approximation

$$V(x, r) \approx \frac{M}{2} (\omega_\perp^2 r^2 + \omega_x^2 x^2) \quad (\text{B.2})$$

with the trapping frequencies

$$\omega_\perp(x) = \frac{2}{w(x)} \sqrt{\frac{V_0}{M}} \quad \text{and} \quad \omega_x = \frac{\sqrt{2}}{x_R} \sqrt{\frac{V_0}{M}}. \quad (\text{B.3})$$

The aspect ratio $w_\perp(0)/\omega_x = \sqrt{2}x_R/w_0 = \sqrt{2}\pi w_0/\lambda$ depends only on the beam waist in the focus. For characterization, we performed a measurement of BEC dipole oscillations at reasonably high waveguide power. The longitudinal oscillation frequency was found by suddenly switching off the XDT. Due to the fact that the XDT focus and the equilibrium position of the BEC in the waveguide are not overlapped at this high power, the cloud starts spreading and oscillating, eventually leaving the field of view of the current imaging system. After each oscillation period, the cloud is visible again and allows to deduce the frequency to $2\pi \times 3.63(4)$ Hz after two periods. Figure B.1 shows the result of the corresponding

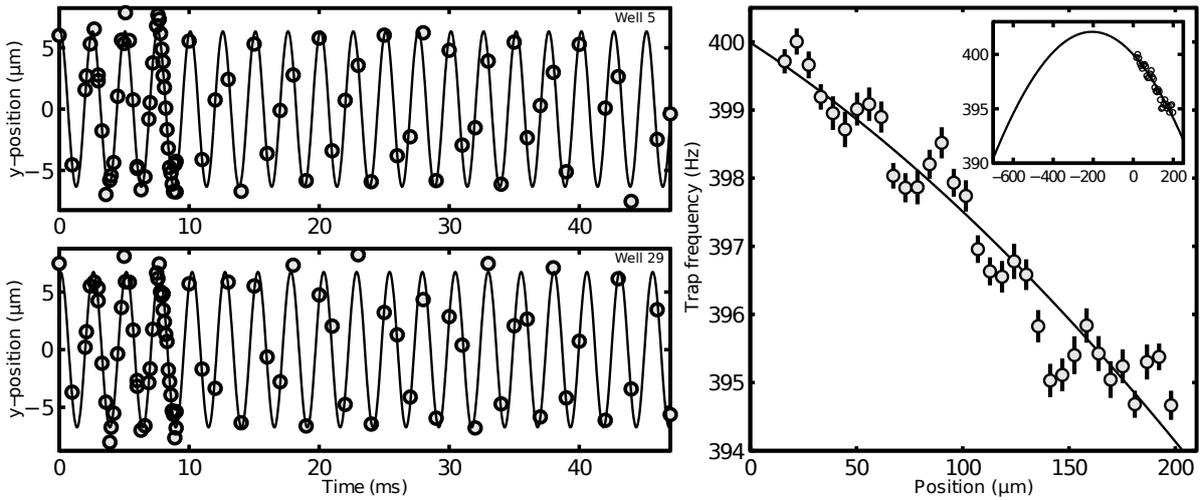


Figure B.1 Characterization of the waveguide trap Oscillations in transverse direction can be observed after displacement by a short period of free fall and subsequent retrapping. Extracting the oscillation frequency for each well yields a dependence on the horizontal position in the optical lattice (right). Including the independently characterized longitudinal trapping frequency of the waveguide beam with the same optical power (3.63(4) Hz) yields an estimate of the focus position, which appears to be located at $-206(16) \mu\text{m}$ with respect to the field of view of the imaging from 0 to $215 \mu\text{m}$ in object space.

measurement of ω_{\perp} . The waveguide power was ramped to its final value and briefly interrupted for $500 \mu\text{s}$. The short period in free fall results in a displacement from the equilibrium position and subsequent dipole oscillation in vertical direction. The slicing of the cloud by the lattice enables a local analysis of the frequency. We observe a clear dependence of the frequency on the well position in the lattice, which can be self-consistently fitted by the expected dependence $\omega_{\perp} = \omega_{\perp}(x_0) / \sqrt{1 + [(x - x_0)/x_R]^2}$ with the parameters $x_0 = -206(16) \mu\text{m}$ relative to the imaging field of view, $\omega_{\perp}(x_0) = 2\pi \times 402(1) \text{ Hz}$ and $x_R = 2.01(2) \text{ mm}$. This corresponds to a waist of $w_0 = 25.7(3) \mu\text{m}$ in the focus or an aspect ratio of $111(1)$. As a reference, Fig B.2 shows the calculated scaling of the trap frequencies in horizontal and vertical direction versus V_0 , which is proportional to the optical power. Including gravity, the vertical cut through the potential obtains the form

$$V(z) = Mgz - V_0 \exp \left[-2 \frac{z^2}{w^2} \right], \quad (\text{B.4})$$

which is plotted in the inset. It has two potential turning points located at $z = \pm w/2$ (zero of second derivative). The trap depth is vanishing if the first derivative at $z = -w/2$ is positive, which is the case for $V_0 < Mgw \exp[1/2]/2$ (marked as dotted line). At this point, the trap frequency also vanishes. The trap position obtains a downshift (gravitational sag) for decreasing V_0 ($\sim g/\omega_{\perp}^2$ in harmonic approximation). For the waveguide intensity used for all other measurements presented in this thesis, we performed an independent check using parametric excitation by sinusoidal modulation of the intensity, which yielded a clear resonance feature at $2\pi \times 261(1) \text{ Hz}$. One has to be careful with the interpretation of this observation. If the effect of gravity would be negligible, the strongest response would be expected at a

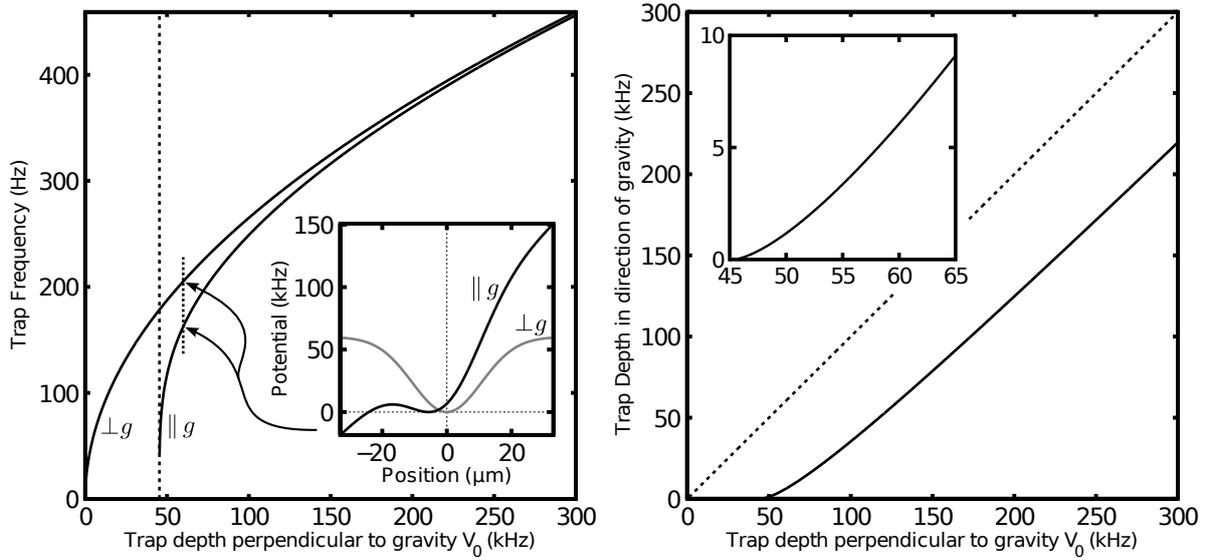


Figure B.2 Transversal trapping frequencies and trap depths of the waveguide deduced from the extracted parameters. The trap in vertical direction ($\parallel g$) is weakened by the additional linear gradient due to gravity, which also limits the minimal horizontal trap depth ($\perp g$) for the chosen beam waist to ~ 45 kHz before all atoms are lost. The parameter V_0 equals the trap depth in horizontal direction and is proportional to the applied optical power. The resulting trap frequencies in harmonic approximation at the potential minimum both have a square root like behavior. The inset shows a cut through the trapping potential for the example $V_0 = 2\pi\hbar \times 60$ kHz. Gravity weakens the trap, makes it shallower and shifts the minimum of the trapping potential vertically downwards by $\sim g/\omega_{\perp}^2$ (gravitational sag). The right panel shows a comparison of the two transversal trap depths and the inset a zoom into the region, where the vertical trap depth vanishes.

modulation frequency twice as big as the transverse trapping frequency [176]. In our case however, the modulation of the laser intensity couples very effectively to the dipole mode via the gravitational sag and thus the excitation is strongest when the modulation frequency is equal to the trapping frequency. The deduced transverse trapping frequency is consistent with the parameters of the independent characterization, from which its inhomogeneity at this setting can be constrained to $\lesssim 1\%$ over the whole BEC cloud.

Optical lattice

Here, we are interested in the local trapping frequencies at the minima of the lattice potential (maxima of the standing wave light intensity) along the longitudinal waveguide direction. It is possible to excite dipole oscillations in the lattice by abruptly shifting the optical phase of one of the two intersecting beams with an electro-optic modulator (EOM). This also shifts the phase of the sinusoidal trapping potential. Since it is not perfectly perpendicular to the waveguide trap, the potential shift also excites horizontal oscillations perpendicular to the lattice direction, which couple to the primarily excited mode. The resulting beat note can in principle be used to extract both frequencies. However, due to the high trapping frequencies and the small extension of the clouds in lattice direction, a direct observation of the oscillations is difficult, especially

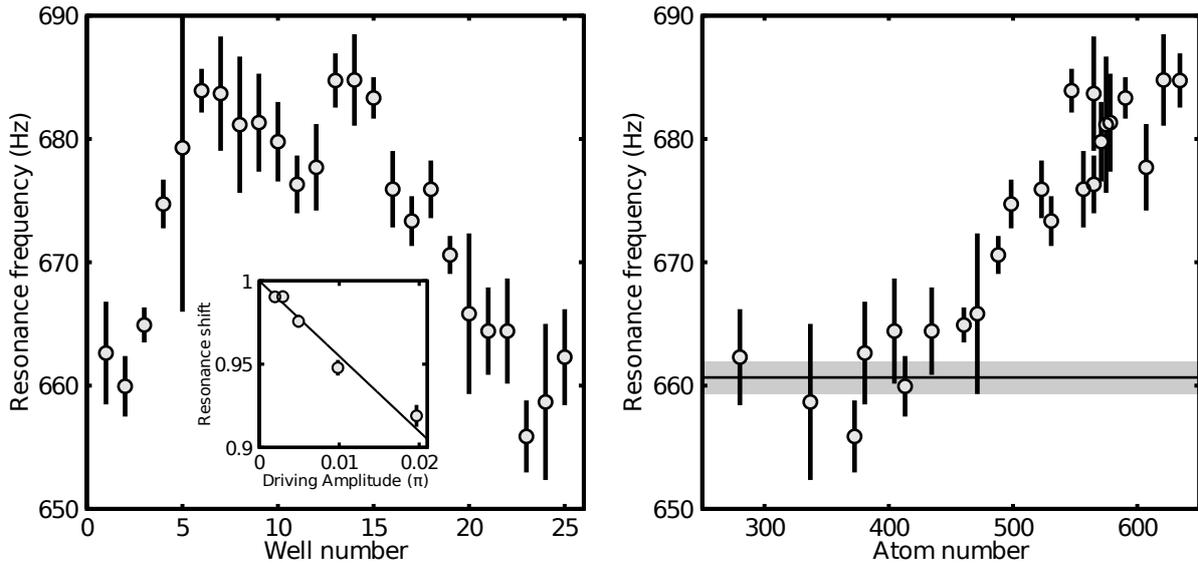


Figure B.3 Measurement of the trap frequency in the optical lattice potential The transverse width of the individual lattice sites after 1 s of lattice shaking (parametric heating) shows a resonant increase at the local longitudinal trapping frequency, which shifts with the driving amplitude (inset of left panel shows the mean relative shift vs. driving amplitude in units of the lattice phase). The final value for each well is obtained by linear extrapolation to zero driving amplitude. The left panel shows this extracted value vs. well number from left to right. The characteristic peak in the middle is caused by the atom number distribution in this measurement, which is revealed in the second panel. The upward trend for higher atom numbers is probably caused by interaction effects. As our estimate of the local trapping frequency, we average up to 450 atoms, which yields $661(2)$ Hz, shown as gray band. From this measurement, the variation of the optical power over the whole cloud can be roughly constrained to $\lesssim 2 \times 10/660 \approx 3\%$.

with small amplitudes. For bigger displacements, the oscillations probe parts of the potential with smaller curvature and thus the resulting frequency is reduced compared to the harmonic approximation at the potential minimum.

We therefore characterize the standing wave potential by parametric excitation [176]. For this, we modulate the phase of one beam with the EOM and look for a resonant enhancement of the vertical width of the cloud after a small time of flight. We extract it from the absorption image by a Gaussian fit to the density distribution of each well. The resonance feature vs. modulation frequency is fitted with a Gaussian for each well to obtain its center position. The observed resonance frequency shows a distinct downshift for bigger driving amplitudes, as expected for larger excursions. We use this systematic behavior to extrapolate linearly to zero driving amplitude. The results of this procedure are shown in Fig. B.3. Strikingly, a maximum of the resonance position is observed in the center of the total ensemble, which is reminiscent of the mean atom number distribution. This is revealed by a plot vs. atom number, which shows, that similar atom numbers obtain matching resonance frequencies, irrespective of the lattice well position. The observed shift presumably stems from the growing size of the BECs with atom number due to repulsive interactions which then again probe bigger distances away from the minimum. The frequency value settles for atom numbers < 450 and we thus take a weighted average of the results up to this point and obtain the estimate $2\pi \times 661(2)$ Hz.

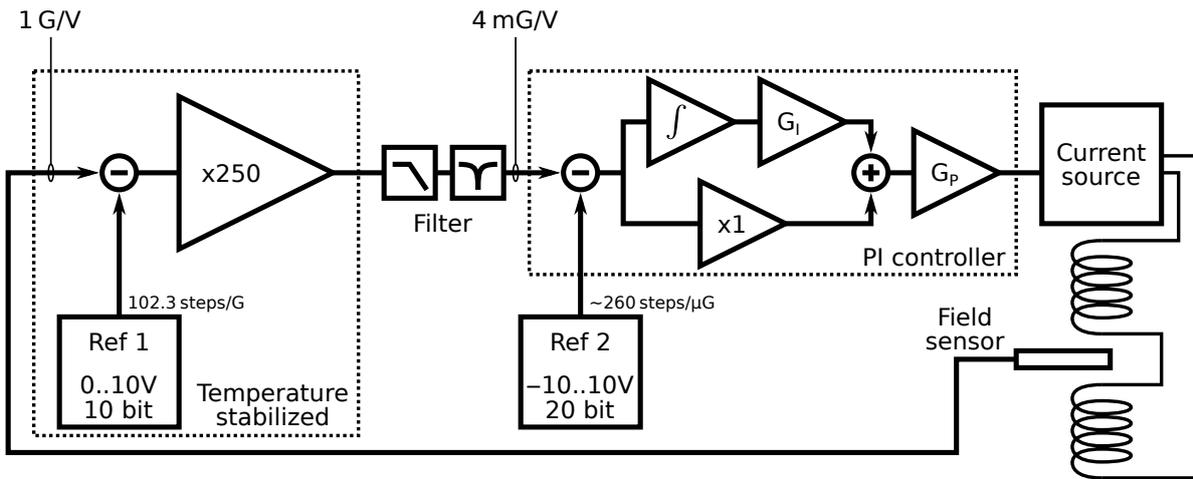


Figure B.4 Principle of the magnetic offset field stabilization The actual field is measured by a fluxgate sensor with a sensitivity of 1 V/G, from which a reference voltage (Ref 1) is subtracted and the result amplified by an instrumentation amplifier. This first stage is actively temperature stabilized. Subsequent low-pass and notch filters suppress the remaining ripple at the working frequency (~ 16 kHz) of the fluxgate. A finely adjustable set-point voltage (Ref 2) is subtracted at the input of the following PI controller. The controller output is connected to a voltage controlled current source, which drives a fine-tune coil pair wound on the same frame as the main coil.

Interestingly, the size of the resonance shift with the driving amplitude cannot be explained by simple time averaging of the modulated potential, which would also result in a depth and frequency reduction. The measured effect is about two orders of magnitude bigger.

B.2 Magnetic field stabilization

Our offset magnetic field of 9.13 G close to the Feshbach resonance at 9.1 G is actively stabilized to a fluxgate magnetic field sensor¹. Its expected noise characteristics are shown in Fig. B.5. From the specified noise floor value of $\sigma_{\text{spec}} = 0.1 \mu\text{G}/\sqrt{\text{Hz}}$ at 1 Hz we guess the shape of the noise power spectral density to

$$S(f) = \begin{cases} \sigma_{\text{spec}}^2 \frac{1\text{Hz}}{f} & \text{for } f \leq 1 \text{ Hz} \\ \sigma_{\text{spec}}^2 & \text{for } f > 1 \text{ Hz.} \end{cases} \quad (\text{B.5})$$

To estimate the reproducibility of a DC value for repeated experiments, we integrate $S(f)$ from a lower cutoff, given by the repetition time of $t_r \sim 40$ s to a highest frequency, estimated as the inverse of the probe time (where the atomic system is sensitive to the magnetic field). Usually the corresponding frequency is well above 1 Hz, so the part below the corner frequency gives a constant offset of $\sigma_1 = \sigma_{\text{spec}} \times \sqrt{1 \text{ Hz} \log(1 \text{ Hz} \times t_r)} \approx 0.2 \mu\text{G}$. When integrating down to ~ 0.5 hours, which is a typical timescale between automated Ramsey measurements for correcting magnetic field drifts, one gets a slightly bigger value of $\approx 0.3 \mu\text{G}$. The expected noise contribution above the corner frequency is $\sigma_2 = \sigma_{\text{spec}} \sqrt{(1/\tau - 1 \text{ Hz})}$ which evaluates to

¹Bartington Instruments Mag-03

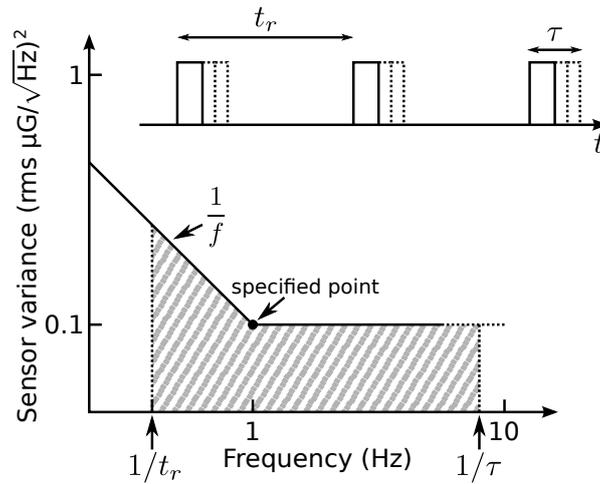


Figure B.5 Extrapolated specification of the magnetic field sensor and relevant timescales The only specified noise level of the employed Mag-03 sensor from Bartington instruments is $0.1 \mu\text{G}/\sqrt{\text{Hz}}$ at 1 Hz. Older versions of its datasheet suggest that this is the corner frequency for the transition from $1/f$ variance behavior (flicker noise, which is generically observed at low frequencies) to a white noise floor. The upper graph shows the experimentally relevant time scales repetition time $t_r \sim 40$ s and probe time $\tau \lesssim 30$ ms. The corresponding lower cutoff frequency lies below the corner frequency, the upper one usually well above. The expected reproducibility of subsequent experimental runs is estimated by the integral from the lower to the upper cutoff frequency, depicted as hatched area.

$\approx 0.6 \mu\text{G}$ and $\approx 3.8 \mu\text{G}$ for the two relevant time scales of the unstable fixed point evolution (~ 30 ms) and the interrogation time for Ramsey sequences ($\sim 700 \mu\text{s}$) respectively.

Both offset and scaling of the sensor are temperature dependent and are specified as $6 \mu\text{G}/^\circ\text{C}$ and $200 \text{ ppm}/^\circ\text{C}$ respectively. The scaling error translates into a sizable dependence of $2 \text{ mG}/^\circ\text{C}$ at our working field.

To stably reduce the sensitivity of the sensor ($1 \text{ V}/\text{G}$) to a manageable level, we subtract an offset voltage, which is derived from a buffered² voltage reference³ by a digitally programmable potentiometer⁴. The subtraction is performed by a low-noise instrumentation amplifier⁵ and amplified by a factor ~ 250 . The expected temperature drift of this stage is dominated by the digital potentiometer. One has to note, that Seebeck thermoelectric effects⁶ already play a role at the level of several μV , for example if air currents on the circuit cause the leads of the amplifier to have slightly different temperatures or there exist thermal gradients over its body. To reduce such effects, this first stage is firmly enclosed in a machined aluminum box, surrounded by a fitted styrofoam block and actively temperature stabilized via an attached Peltier element using feedback through a thermistor on a Wheatstone bridge. The obtained temperature stability is in the low mK region. Once set, the serial communication leads for the digital potentiometer are disconnected to avoid ground shifts and incoupled noise. After the amplification, the sensitivity is reduced to $4 \text{ mG}/\text{V}$, such that the temperature stability of the

²unity gain with OP27, offset voltage $\sim 0.5 \mu\text{V}/^\circ\text{C}$

³AD588JQ, specified temperature dependence $\sim 15 \mu\text{V}/^\circ\text{C}$

⁴AD5293BRUZ-20, $5 \text{ ppm}/^\circ\text{C} \sim 45 \mu\text{V}/^\circ\text{C}$

⁵AD524, input offset drift $\sim 2 \mu\text{V}/^\circ\text{C}$

⁶for example $\sim 5 \mu\text{V}/^\circ\text{C}$ for lead-tin solder joined copper [177]

lab is sufficient for the stable operation of the remaining electronics. After this stage a second reference voltage set by a precision digital-to-analog converter (DAC)⁷ on an optocoupled microcontroller is subtracted at the input stage of the following analog proportional-integral (PI) controller. The output of the PI controls an unidirectional current source, which drives current (~ 1 A) through a pair of coils. These are wound over two quadratic frames with an edge length of 1 m. On the same frames, wires of bigger cross section are wound which carry ~ 140 A⁸ to reach a magnetic field of 9.1 G. The size of the coils is chosen as big as possible with the space constraints around the experimental chamber, which is also why they were positioned at twice the distance of the ideal Helmholtz configuration⁹. The magnetic field in z -direction therefore has a local minimum located close to the position of the atoms. Big coils are used to increase the homogeneity in transverse direction, since the sensor is mounted at a distance $(y, z) \sim (3, 10)$ cm above the BEC cloud, whose long axis extends in x -direction.

We characterized the suppression of changing external fields in closed loop configuration during a slow ramp of a superconducting 15 T magnet located in the second floor of the building (our experiment is in the basement). This caused a change of ≈ 16 mG deduced from the output voltage of the PI controller. During the same period, the actual magnetic field at the place of the atoms as measured by repeated microwave Ramsey sequences changed by ≈ 0.3 mG. Thus, the suppression factor of static external magnetic fields in the direction of the stabilized field is ~ 50 (34 dB). We performed a corresponding test by changing the current through the big coils, which the loop control could compensate to 0.35 mG at a total change of 97 mG, which corresponds to a suppression of ~ 280 (49 dB).

The remaining long-term drifts of the magnetic field are compensated by changes of the setpoint (second voltage reference) based on the result of automated Ramsey measurements¹⁰. The changes are dominated by external magnetic fields and temperature drifts of the sensor signal, which can be seen in Fig. B.6. The air temperature close to the magnetic field sensor inside the curtain enclosed region around the experimental chamber is routinely logged by a wireless sensor appliance¹¹. The magnetic field control error as compensated after each reference measurement with atoms almost perfectly correlates with this temperature. However, the obtained slope of 3.9 mG/ $^{\circ}$ C is a factor of two larger than the specified temperature drift of the magnetic field sensor.

To set this into perspective, we want to shortly compare this to usual drifts of its surrounding. If we generously estimate position drifts of the sensor by thermal expansion of its holder (~ 40 cm of aluminum¹²) to ~ 10 μ m/ $^{\circ}$ C, the sensor would have to reside in a magnetic field gradient of 2 G/cm in z -direction to explain a change of 2 mG/ $^{\circ}$ C.

The closest source of magnetic field gradients is the magnet of the main ion sputter vacuum pump¹³. Typical values of its stray field and gradient¹⁴ are 200 mG and 10 mG/cm respectively

⁷AD5791 [178] with REF5050 voltage reference

⁸current controlled Delta Elektronika SM15-200D

⁹One of the coils is attached to the undersurface of the optical table

¹⁰For all characterizations of the magnetic field, we use Ramsey sequences on the linear Zeeman sensitive one-photon transition. During measurements on the pseudospin system we use the two-photon transition to keep track of all changes to its specific resonance.

¹¹AREXX Multilogger

¹²linear expansion coefficient ~ 25 ppm/ $^{\circ}$ C

¹³Varian VacIon 150

¹⁴corresponding plots can be found in the manual

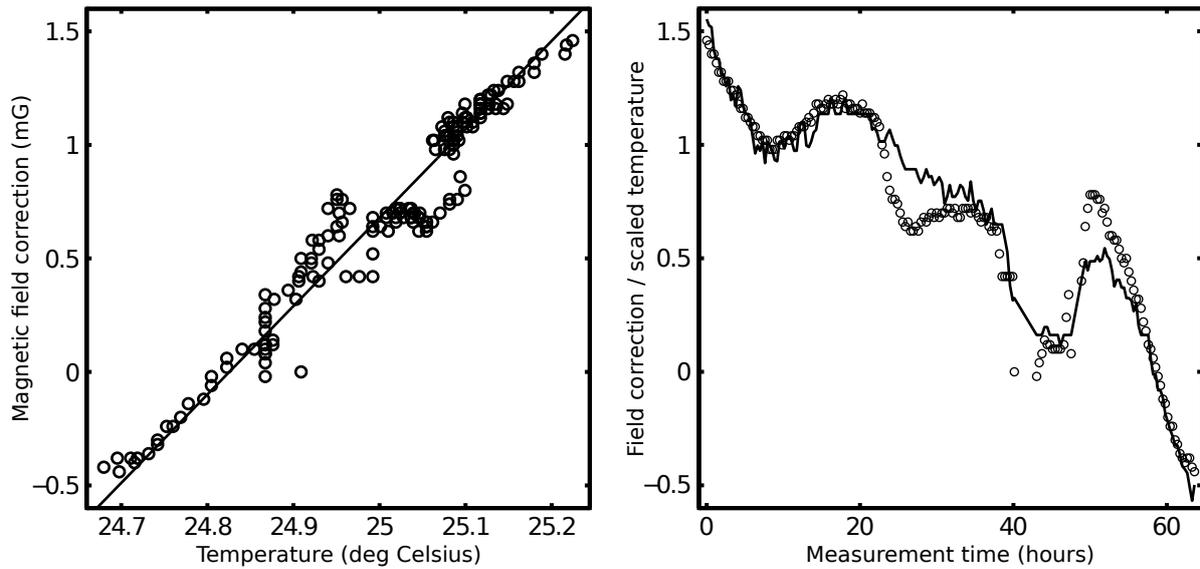


Figure B.6 Temperature drift of the magnetic field stabilisation The absolute error of the active stabilisation as determined and corrected by automated microwave Ramsey experiments strongly depends on the temperature close to the magnetic field sensor. A correlation plot on a typical measurement timescale (2.5 days) yields a dependence of $3.9 \text{ mG}/^\circ\text{C}$ (left, black line). Both values vs. time are shown in the right plot. The temperature (solid line) is rescaled by the linear correlation.

at the position of the magnetic field sensor ($\sim 40 \text{ cm}$ distance to the flange of the pump). This would translate into a required thermal sensor movement of $2 \text{ mm}/^\circ\text{C}$, which is two orders of magnitude larger than reasonable. In other words, the expected thermal position drifts translate into negligible $10 \mu\text{G}/^\circ\text{C}$.

Another source of temperature drift could be the one of the magnet of the pump itself. Typical ferrite hard magnetic materials have a temperature coefficient of their remanence (permanent field at the absence of excitation) of $-0.2\%/^\circ\text{C}$, which would translate into a change of the gradient of $\sim 20 \mu\text{G}/^\circ\text{C cm}$ and thus to an error of $\sim 0.2 \text{ mG}/^\circ\text{C}$ at the position of the atomic cloud.

All these influences are smaller than the observed effect, which makes it most plausible that the core or electronics of the sensor is causing this temperature dependence.

To linearize the sensor and enlarge its dynamic range, the external field is nulled by internal compensation coils around the fluxgate core in a closed loop control. The required compensation current is proportional to the external field, from which the sensor output voltage is derived. During the magnetic trapping phase of the experiment, this circuitry is saturated and the maximal power is dissipated. The large magnetic fields during this time also cause offset shifts due to a slight magnetization of the core material¹⁵. We found these effects to equilibrate if the experimental cycle time is kept constant and the experimental sequence for cooling and trapping is fixed. Cutting the sensor power supply during the time the field is not actively stabilized in order to reduce the power dissipation resulted in larger shot-to-shot fluctuations than keeping it powered all the time. We suspect that switch-on drifts of the sensor electronics dominated in this case.

¹⁵Bartington Instruments, private communication

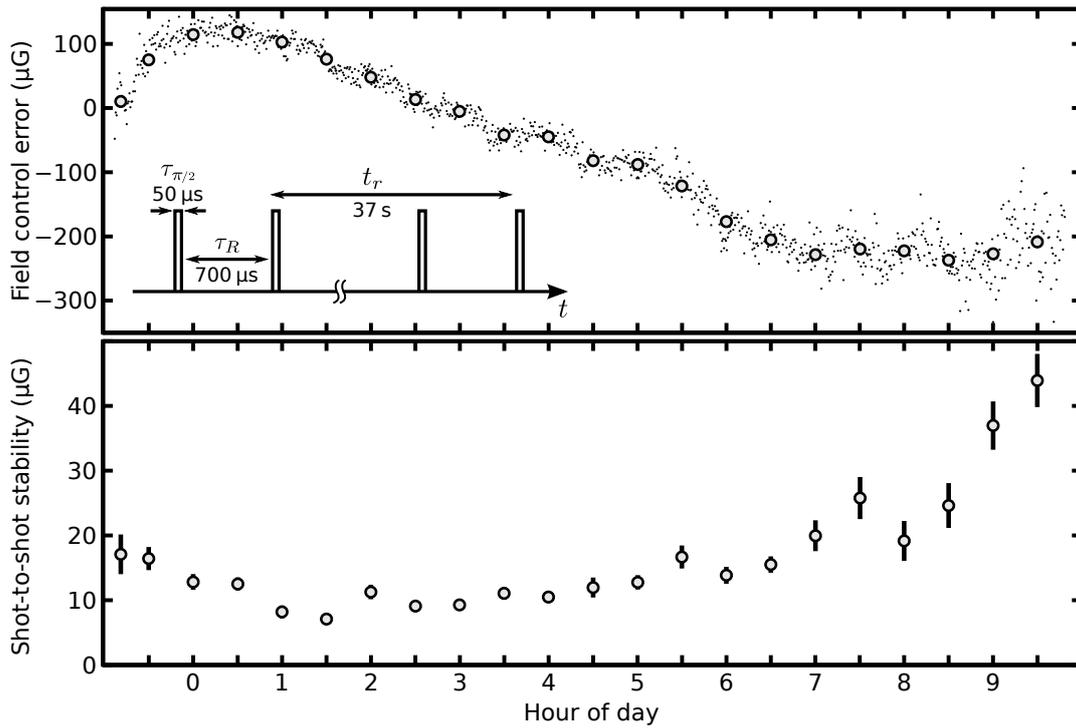


Figure B.7 Magnetic field stability obtained with the active control at an offset of 9.2 G. The magnetic field control error as measured with Ramsey sequences on the sensitive transition $|F, m_F\rangle = |1, 1\rangle \leftrightarrow |2, 0\rangle$ during a relatively quiet night shows a slow drift (300 μG , corresponding to a temperature change of $\sim 75 \text{ mK}$). The shot-to-shot fluctuations reach a minimum at about three o'clock in the morning, where the institute and its surroundings are typically most quiet. The inset schematically shows the employed microwave Ramsey sequence with pulse time $\tau_{\pi/2}$, Ramsey interrogation time τ_R and repetition time t_r . Every small point in the upper panel corresponds to one Ramsey experiment.

Thus, the increased thermal sensitivity is most probably caused by the influence of the harsh magnetic environment on the sensor and not by position drifts of any sort.

To reduce temperature drifts, we stabilized the air temperature of the lab¹⁶ by employing a digital PID controller (with a thermistor as room temperature sensor) connected to a solid state relay, which switches 4 kW of power from two fan heaters. This reduced thermal drifts typically to $< 0.1 \text{ }^\circ\text{C}/\text{day}$ such that automated Ramsey corrections, spaced by about half an hour during long-term measurements provided sufficiently stable conditions¹⁷.

Figure B.7 shows a typical time trace of the magnetic field with this improvement. The characterization of the control error was performed with repeated Ramsey sequences with an interrogation time of $\tau_R = 700 \mu\text{s}$ on the transition $|1, 1\rangle \leftrightarrow |2, 0\rangle$. A detuning of 357 Hz was chosen to obtain a zero-crossing of the imbalance after the second $\pi/2$ -pulse. To increase the dynamic range and to operate the Ramsey sequence at its highest sensitivity, the setpoint (Ref 2) of the controller was repeatedly adjusted according to the Ramsey result of the preceding

¹⁶This has the positive side effect of largely increased stability of the lasers and the optical setup

¹⁷The data of Fig. B.6 were taken before this step

three repetitions. This measurement yields the total error of the field control by combining the setpoint compensation value with the obtained atomic imbalance z :

$$B_{\text{err}} = U_{\text{Ref}2} \times \frac{4 \text{ mG}}{\text{V}} - \arcsin z \times \left[2\pi(\tau_R + 2\tau_{\pi/2}) 700 \frac{\text{Hz}}{\text{mG}} \right]^{-1} \quad (\text{B.6})$$

Here, $2\tau_{\pi/2}$ is the time of the two microwave pulses, which adds in our case $100 \mu\text{s}$ to the total phase evolution time. The analysis of the shot-to-shot repeatability employs a variant of the Allan deviation, routinely used to characterize the stability of clocks (see also B.4). One calculates the local differences

$$\delta B_{\text{err}}^{(j)} = B_{\text{err}}^{(j+1)} - B_{\text{err}}^{(j)} \quad (\text{B.7})$$

and from them the Allan-type variance as $\sigma_A^2 = \langle \delta B_{\text{err}}^2 \rangle / 2$. The use of the local differences has the advantage of separating out the actual shot-to-shot variability from the total change in magnetic field, which also includes slow drifts. In Fig. B.7 the averaging for σ_A was performed over a timescale of half an hour. The achieved shot-to-shot stability largely depends on the activity of the surrounding experiments and presumably also on the variability of the total power supplied to the institute. The stability significantly improves during the night and on weekends. Including slow drifts, we can achieve a typical stability of $30 \dots 40 \mu\text{G}$ on the relevant timescale of half an hour between automated correction measurements.

The magnetic field control also includes a synchronization of the whole experiment to the 50 Hz mains frequency. For this, the main clock of the experimental control is interrupted after the BEC preparation, $\sim 50 \text{ ms}$ before the beginning of the microwave sequence and restarted at a fixed phase of a 50 Hz square-wave signal derived from a small transformer. Thus, the experiment pauses for up to 20 ms at this stage.

The remaining amplitude of the 50 Hz component of the magnetic field as reduced by the control loop is $\sim 200 \mu\text{G}$, which we characterized by scanning the beginning of the Ramsey sequence. We compensated this with a feed-forward by adding a synchronized sinusoidal 50 Hz voltage with an amplitude of 50 mV from a frequency generator to the input of the PI loop. By proper adjustment of the phase, we could reduce this influence below the shot-to-shot variation. Higher harmonics of the line frequency are already suppressed by the loop to an unobservable level.

B.3 Microwave and radio frequency system

Figure B.8 schematically shows the employed generation and control system of radio frequency and microwave magnetic fields to drive single- as well as two-photon transitions in ^{87}Rb . All critical sources are frequency stabilized to an ovenized 10 MHz quartz oscillator, whose slow aging and temperature drifts are compensated through a comparison with the timing signal of the global positioning system (GPS). This part is crucial for long-term stability, since already the aging of the time-bases of the synthesizers ($\sim 5 \cdot 10^{-10}/\text{day}$) gives a sizable drift of 3.4 Hz/day at 6.8 GHz.

The microwave signal is amplified by a chain of power amplifiers to a total power up to 15 W, which is applied to a small 'antenna', which consists of the exposed solid copper core of

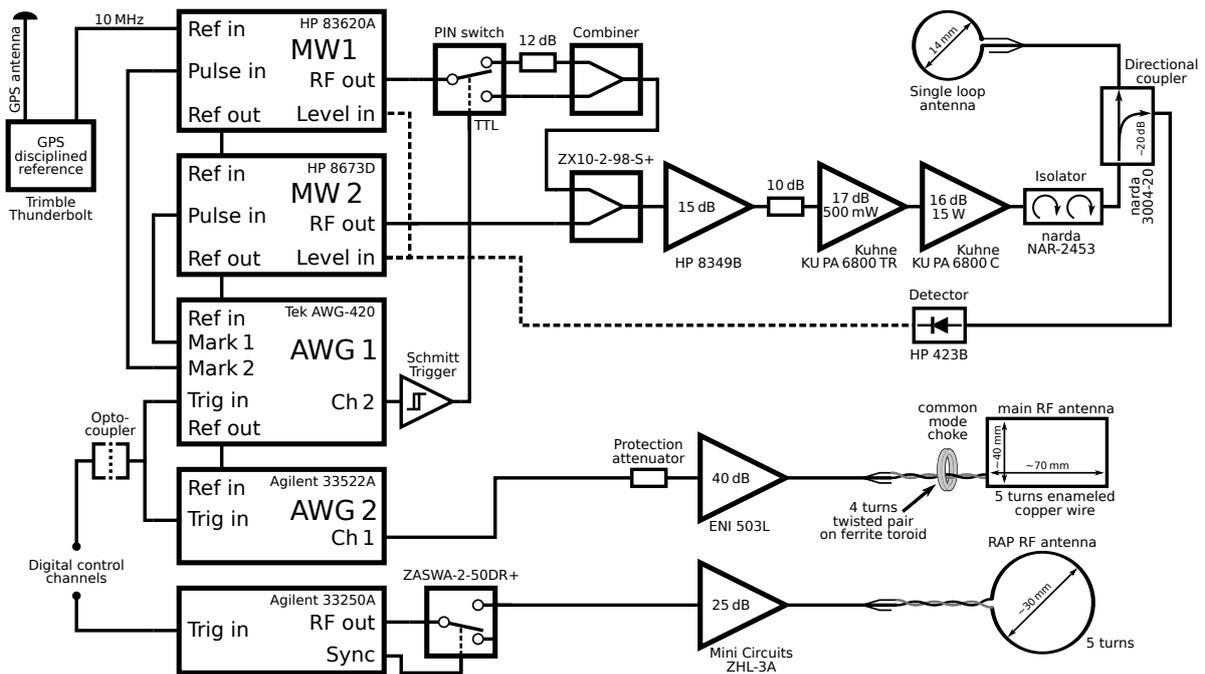


Figure B.8 Block circuit of the microwave and radio frequency system. The signals of two microwave synthesizers are combined and amplified to a level of up to 15 W. A double isolator protects the power amplifier from back-reflected power. A small fraction of the final power is coupled out and rectified. This signal can be used for power stabilization via the external leveling option of the synthesizers. The microwave sequence is controlled by an arbitrary waveform generator (AWG 1), which controls the pulse modulators of the synthesizers as well as a single pole double throw (SPDT) switch on MW 1, which can be used for quick and precise attenuation of its power by a fixed value. Radio frequency signals are generated by a second AWG with a larger waveform memory and a general purpose function generator, which provides the rapid adiabatic passage (RAP) frequency sweep for initial state preparation. All critical frequency sources are long-term stabilized on an ovenized 10 MHz reference oscillator, which is synchronized with the atomic clocks on the global-positioning-system satellites. Its antenna is mounted on the roof of the institute. The magnetic near fields of the loop antennas drive the atomic transitions.

the coaxial cable bent to a circle and soldered to the shield conductor of the coax. This loop is not impedance matched, such that an uncontrolled portion of the power gets reflected, which is then absorbed by the Faraday isolator.

A small fraction of the microwave power is coupled out by a directional coupler before the antenna and rectified by a diode detector, which is thermally contacted to a temperature stabilized aluminum holder. This signal can be used at the external leveling input of the microwave generators for power stabilization.

The timing of the pulse sequences is realized with two arbitrary waveform generators (AWG). AWG 1 drives the pulse modulators of the two microwave synthesizers and the switchable attenuator on MW 1. It is also capable of providing radio frequency signals with the other channel but is limited in communication speed and waveform memory depth. AWG 2 has a lower bandwidth than AWG 1 but can be programmed more quickly and supports longer waveforms. It is used to generate the phase-controlled radio-frequency part for the two-photon

transition. The RF of AWG 2 is amplified by a general purpose amplifier delivering up to 3 W of power.

The attached RF antenna has a rectangular loop shape with five windings. Interestingly, an estimate for an appropriate number of turns can be found without referring to much of the details of the loop. Assuming that the last amplifier has a real output impedance¹⁸ of $Z_0 = R_0 = 50 \Omega$ and the DC resistance of the loop can be neglected with respect to it, the value to optimize is the generated magnetic field $B \propto nI$, where I is the current and n is the number of windings of the coil. Ordinary impedance analysis is sufficient here, since the wavelength (~ 50 m) is still way larger than all the dimensions. Thus, the AC magnetic field amplitude scales with

$$|B|^2 \propto \left| \frac{nU_0}{R_0 + i\omega L} \right|^2 \propto \frac{n^2}{R_0^2 + (\omega n^2 L_1)^2} \propto \frac{1}{\Gamma} \frac{\Gamma^2}{\omega^2 + \Gamma^2} \quad (\text{B.8})$$

with $\Gamma = R_0/n^2 L_1$. Here, L_1 is the inductance for a single loop and U_0 the source voltage. This is optimized for $\Gamma_{\text{opt}} = \omega$, that is for $n_{\text{opt}} = \sqrt{R_0/\omega L_1}$. In our case L_1 is on the order of $0.1 \dots 0.3 \mu\text{H}$, so the choice of 5 windings is optimal for $3 \dots 1$ MHz respectively. It is interesting to note that the task is not to match the impedance of the antenna to those of the source (to create a resistive load of 50Ω). It is rather its reactance that has to match 50Ω to be optimal, that is $\omega L_{\text{opt}} = R_0$. For a given (too large) inductance of the antenna, this requirement can be met with a proper RF transformer inserted at the transition from the coaxial cable to the antenna. It is not important, that the majority of the power gets reflected and is dissipated in the amplifier, since we are only interested in the near field amplitude, not in the far field radiation.

We chose the straightforward approach to directly connect the end of the coax to the loop. Apparently, the antenna has significant coupling to ground connected metal parts of its surrounding, causing the current of inner conductor and shield of the coax to deviate from the ideal equal magnitude and 180° relative phase. This results from the grounding of the amplifier, which opens various return paths besides the inside of the coax shield. The current imbalance is possible since the outside of the shield acts as separate conductor for high frequencies¹⁹, and turns the coax itself into an antenna. In our case, this is especially troublesome for the closeby magnetic field sensor. The mismatch can be largely damped by a so called current- or choke-balun [179], which consists of several windings of twisted pair on a toroidal ferrite core. It introduces a large impedance for common mode currents, while differential currents can easily pass, balancing the currents on the coax. This step solved the crosstalk problem of the RF pulses on the magnetic field stabilization.

The microwave crosstalk is more subtle, but could be considerably reduced by careful positioning of a sheet of aluminum foil at the bottom end of the magnetic field sensor.

For robust initial state preparation from $|F = 1, m_F = -1\rangle$ to $|F = 1, m_F = 1\rangle$ we employ a rapid adiabatic passage (RAP), for which a separate RF chain with a general purpose signal generator delivers a frequency sweep over 1.2 MHz in 20 ms. The switch at the output of the generator solves the problem, that the sweep mode cannot be combined with the burst mode which would otherwise lead to continuous output before and after the sweep.

¹⁸The connecting coaxial cable does not change this.

¹⁹The skin effect is already significant, letting AC current only flow in a thin layer close to the surface of conductors (the $1/e$ skin depth is only $\sim 20 \mu\text{m}$ for copper at 10 MHz).

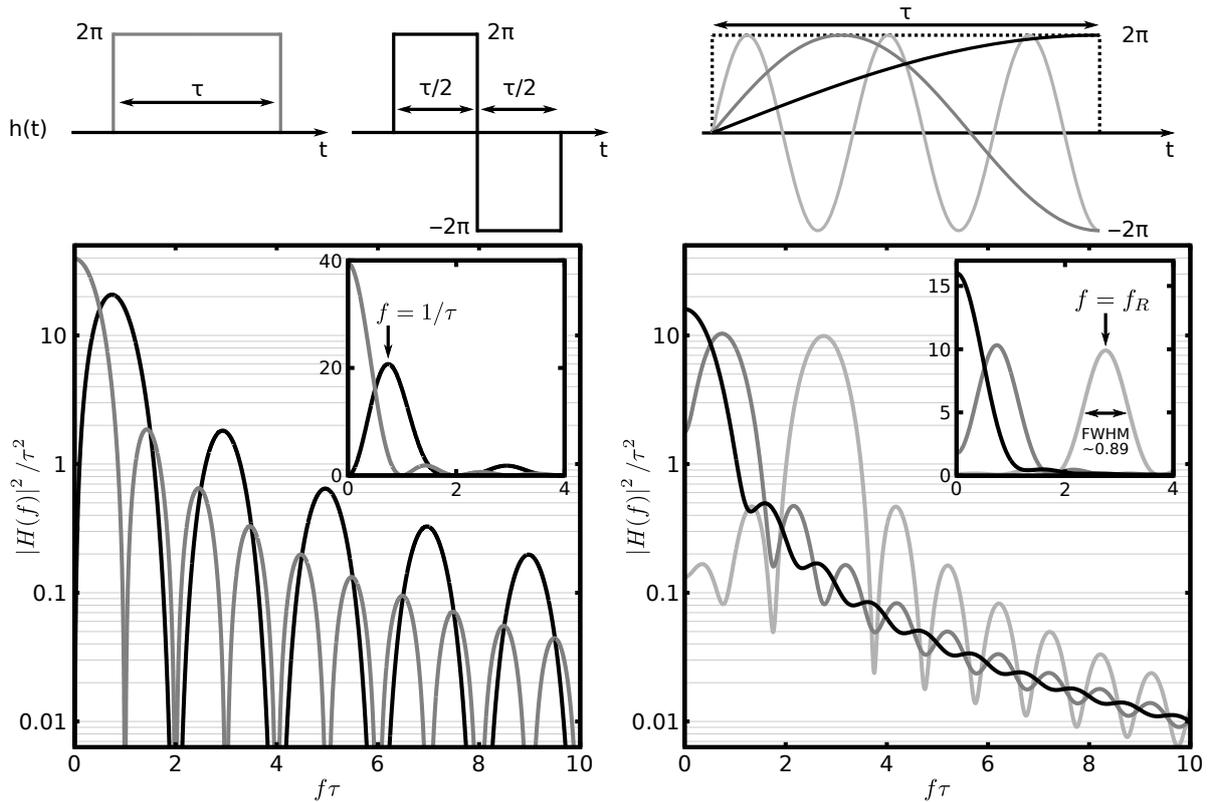


Figure B.9 Universal spectral transfer functions of pulse sequences. Here, three examples of transfer functions are summarized. The top row shows the timing sequence $h(t) = 2\pi s_x(t)$ for the three cases of piecewise free evolution on the equator of the Bloch sphere (left) and resonant Rabi flopping starting in a z -polarized state (right) for three different Rabi frequencies. The bottom row shows the Fourier transform of the above sequences, giving the spectral transfer functions $|H(f)|^2$. The free evolution (single pulse, gray) is most sensitive for frequencies close to zero, whereas this can be suppressed by a phase inversion in the middle of the evolution (spin echo, double pulse), which has the drawback of an increased sensitivity at higher frequencies. Rabi flopping (right) introduces the time scale of the Rabi frequency. A single $\pi/2$ -pulse (black) still has the maximum sensitivity at zero frequency, whereas a sequence of flops performs a series of phase inversions, which makes it most sensitive at the Rabi frequency with a spectral width scaling with the total pulse length. The three plotted cases are $f_R = 1/4\tau$ ($\pi/2$ -pulse, black), $f_R = 3/4\tau$ ($3\pi/2$ -pulse, gray) and $f_R = 11/4\tau$ ($11\pi/2$ -pulse, light gray).

B.4 Spectral sensitivity of pulse sequences

When using frequency sources for atomic spin manipulations, one is confronted with the problem of translating specified or measured numbers of frequency stability into a meaningful stability or variance measure for the atomic system after some pulse sequence or interrogation time. Specifically insightful in this context is the powerful frequency domain view of clock stability [180, 181], which relates spectral properties with time domain variances like the Allan variance. With a proper adaption of this formalism one can obtain a characterization of temporal pulse sequences in the Fourier domain via spectral transfer functions, which act as

filter for the fluctuations of the frequency source. Here, we will summarize this method and calculate some important and insightful examples.

The output voltage of a close to pure (monochromatic) frequency source is characterized by

$$U(t) = U_0[1 + \epsilon(t)] \sin[2\pi\nu_0 t + \phi(t)] \quad (\text{B.9})$$

Here, ν_0 is the carrier frequency and $\phi(t)$ is the statistically fluctuating phase with $|\phi(t)| \ll \pi/2$. We assume here that $\epsilon(t)$, which characterizes the power fluctuations, is small and phase fluctuations dominate. The instantaneous frequency deviation $\Delta\nu$ is the time derivative

$$\Delta\nu(t) = \frac{1}{2\pi} \frac{d\phi}{dt} \quad (\text{B.10})$$

In the stability analysis of oscillators and atomic clocks, one defines $y = \Delta\nu/\nu_0$, which is then invariant under frequency division and multiplication. Its statistical properties are conveniently characterized through variances over a specific integration time in the time domain or power spectral densities in the frequency domain, which are defined as Fourier transforms of autocorrelation functions. A variance in the time domain can be expressed as [182]

$$\sigma^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} g(\lambda)^2 d\lambda \quad (\text{B.11})$$

with $g(\lambda)$ defined as the convolution of a temporal sensitivity function $h(t)$ and the actual fluctuating quantity $x(t)$ as

$$g(\lambda) = \int_{-\infty}^{\infty} x(t)h(t - \lambda) dt \quad (\text{B.12})$$

$h(t)$ describes how $x(t)$ is selected and weighted in the time domain and can be interpreted as impulse response of a spectral filter. In our case, we are interested in the variance of the atomic azimuthal phase φ on the Bloch sphere after a specified timing sequence with periods of variable sensitivity on external detuning. We will concentrate here on two important cases, namely appended free evolution periods on the equator of the Bloch sphere and Rabi flopping starting in a z -polarized state. We consider the Hamiltonian

$$\hat{H} = \Omega \frac{\hat{s}_y}{2} + \delta(t) \frac{\hat{s}_z}{2}. \quad (\text{B.13})$$

Here, \hat{s}_j is used for the Pauli matrices to avoid confusion with the other σ s in this section. We will assume that the mean spin vector is approximately aligned with the x -axis at the end of the time evolution with small fluctuations around it. In this case, the fluctuations of the phase are those of the mean y -component $[\sigma_\varphi^2 = (\Delta\varphi)^2 \approx (\Delta\langle\hat{s}_y\rangle)^2]$. The brackets $\langle\cdot\rangle$ indicate the quantum mechanical average. From the Ehrenfest theorem, we get

$$\frac{d}{dt} \langle\hat{s}_y\rangle = i \langle[\hat{H}, \hat{s}_y]\rangle = \delta(t) \langle\hat{s}_x\rangle = \delta(t) s_x(t) \quad (\text{B.14})$$

The pickup of phase fluctuations thus depends on the mean spin length in x -direction. We assume that the external coupling is resonant on average, that means $\int \delta(t) dt = 0$ for long

integration times. For Rabi flopping with $\Omega \neq 0$ beginning in a polarized state the mean x -spin has a time dependence [$s_x(t) = \sin(\Omega t) = \sin(2\pi f_R t)$], while for free evolution on the equator of the Bloch sphere it is $s_x = \pm 1$. The detuning is directly connected to the frequency fluctuations of the source via $\delta(t) = 2\pi\Delta\nu(t)$ and thus the phase fluctuations can be cast in the above formalism via

$$g(\lambda) = \int_{-\infty}^{\infty} \Delta\nu(t) \underbrace{2\pi s_x(t-\lambda)}_{h(t-\lambda)} d\lambda. \quad (\text{B.15})$$

This convolution in the time domain can for stationary Gaussian processes be translated into a multiplication in the frequency domain [182], which yields the phase variance

$$\sigma_\varphi^2 = \int_0^{\infty} S_{\Delta\nu}(f) |H(f)|^2 df. \quad (\text{B.16})$$

Here, $S_{\Delta\nu}(f)$ is the one-sided power spectral density of the frequency fluctuations and $H(f)$ is the Fourier transform of $h(t)$. In this picture, the temporal sequence acts as a filter in the frequency domain. We now proceed to calculate three examples for the spectral filter functions, which are also depicted in Fig. B.9.

Single pulse This first example has the simple pulse shape

$$h(t) = 2\pi \text{rect}(t/\tau), \quad (\text{B.17})$$

which represents a single phase accumulation time of length τ on the equator of the Bloch sphere, for example the interrogation time of a Ramsey sequence. Its fourier transform is the omnipresent sinc-Function

$$|H(f)|^2 = |2\pi\tau \text{sinc}(f\tau)|^2 = 4 \frac{\sin^2(\pi f\tau)}{f^2}. \quad (\text{B.18})$$

Spin echo The second example is a sequence in which the spin direction is flipped in the middle of the interrogation time τ , corresponding to a spin echo pulse sequence. In this case

$$h(t) = 2\pi \left[\text{rect}\left(\frac{t + \tau/4}{\tau/2}\right) - \text{rect}\left(\frac{t - \tau/4}{\tau/2}\right) \right], \quad (\text{B.19})$$

which leads to the Fourier transform

$$|H(f)|^2 = \left| -4i \frac{\sin^2(\pi f\tau/2)}{f} \right|^2 = 16 \frac{\sin^4(\pi f\tau/2)}{f^2} \quad (\text{B.20})$$

The sensitivity to low frequency fluctuations is decreased substantially compared to the single pulse ($|H(0)|^2 = 0$) but has the disadvantage of an increased sensitivity to higher frequencies. It has a first maximum at $f = 1/\tau$ with a FWHM of $\approx 0.92/\tau$.

Rabi pulse The last example we want to look at is Rabi flopping with frequency f_R and duration τ beginning with a z -polarized state, so $s_x(0) = 0$. The impulse response in this case is

$$h(t) = 2\pi \sin(2\pi f_R t) \operatorname{rect}\left(\frac{t - \tau/2}{\tau}\right), \quad (\text{B.21})$$

whose Fourier transform is most effectively obtained by convolution of the Fourier transforms of the individual terms:

$$\begin{aligned} H(f) &= 2\pi \left[\frac{\delta(f - f_R) - \delta(f + f_R)}{2i} \right] * \left[e^{-i\pi\tau f} \tau \operatorname{sinc}(\tau f) \right] \\ &= \frac{\pi}{i} e^{-i\pi\tau f} \left[e^{i\pi\tau f_R} \tau \operatorname{sinc}[\tau(f - f_R)] - e^{-i\pi\tau f_R} \tau \operatorname{sinc}[\tau(f + f_R)] \right]. \end{aligned}$$

As before, the sinc-function is defined here as $\operatorname{sinc}(x) \equiv \sin(\pi x)/(\pi x)$. The square modulus of the transfer function yields

$$\begin{aligned} |H(f)|^2 &= (\pi\tau)^2 \left[\operatorname{sinc}^2[\tau(f - f_R)] + \operatorname{sinc}^2[\tau(f + f_R)] \right. \\ &\quad \left. - 2 \cos(2\pi\tau f_R) \operatorname{sinc}[\tau(f - f_R)] \operatorname{sinc}[\tau(f + f_R)] \right] \end{aligned}$$

One observes that the transfer function takes the form of two interfering wave packets in Fourier space with relative phase $2\pi\tau f_R$. The interference term vanishes for $\tau = (2n - 1)\pi/(4f_R)$, that is for odd multiples of a $\pi/2$ -pulse, while it is maximal for multiples of a π -pulse [$\tau = n\pi/(2f_R)$]. It is only in the latter case that the function obtains real zeros, as in the case of the other two examples with piecewise free phase evolution.

Power spectral densities To understand the meaning and scaling of the power spectral densities, it is instructive to have a look at a monochromatic frequency modulation of the form

$$\Delta\nu(t) = A \sin(2\pi f_m t), \quad (\text{B.22})$$

which has the power spectral density[181]

$$S_{\Delta\nu}(f) = \left(\frac{A}{\sqrt{2}} \right)^2 \delta(f - f_m) \quad (\text{B.23})$$

The value $A/\sqrt{2}$ is the RMS amplitude of the sine and can be replaced by the DC value for frequency zero. The spectral density is connected to other often encountered measures via

$$S_{\Delta\nu}(f) = \nu_0^2 S_y(f) = f^2 S_\phi(f) = 2f^2 \mathcal{L}(f) \quad (\text{B.24})$$

The value mostly reported in manufacturer datasheets and manuals for single-sideband phase noise is $10 \log_{10}(\mathcal{L}(f))$ in the unit dBc/Hz, which means 'dB below the carrier in a 1 Hz bandwidth at the frequency f away from the carrier'. Close to the carrier frequency ν_0 these spectral densities can often be very well approximated by power laws, which are characteristic for the dominant noise process in a specific frequency band [180]. Figure B.10 shows $\mathcal{L}(f)$ of our microwave sources as specified in the manual with piecewise approximations by power laws and additionally converted to a linear frequency scale to obtain a feeling of what it would look like on a spectrum analyzer around the carrier frequency.

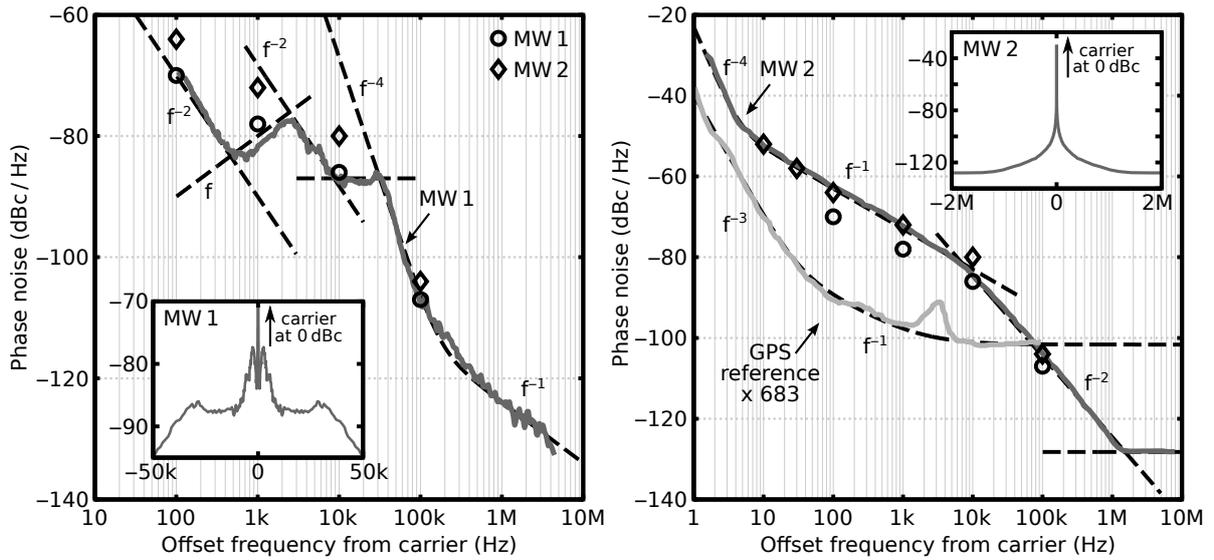


Figure B.10 Collected specifications and characterizations of the microwave sources Here, the single-sideband phase noise of our two HP microwave synthesizers is shown as specified in the manual. The corresponding 'typical phase noise plots', shown as gray solid lines, are shifted vertically to match best the specified numbers (points). To better grasp the meaning of this plots, in the insets they are converted to linear abscissa and mirrored as they would usually be viewed on a spectrum analyzer. Piecewise approximations by power laws are plotted as dashed lines. The right plot additionally shows a phase noise measurement of a Trimble GPS reference performed by John Miles [183] multiplied by the factor 683^2 , which is the minimum increase of the phase noise in the process of direct multiplication of the 10 MHz output to the relevant microwave frequency.

Phase variances With the help of the transfer functions and the phase noise spectral densities, estimates for the expected value of technical phase fluctuations of the different sequences can be obtained by integration of Eq. B.16 for different values of τ . For comparison with long-term measurements, it is interesting to note the connection to the Allan variance, which is a widely used clock characterization in the time domain. It is defined as

$$\text{AVAR}(\tau) = \sigma_y^2(\tau) = \frac{1}{2} \langle (y_{k+1} - y_k)^2 \rangle. \quad (\text{B.25})$$

It is the two-sample variance of average relative frequency deviations

$$y_k = \frac{1}{2\pi\nu_0} \frac{\Delta\phi}{\Delta t} = \frac{\phi([k+1]\tau) - \phi(k\tau)}{2\pi\tau\nu_0} \quad (\text{B.26})$$

over the integration time τ . Its frequency domain representation is [181]

$$\text{AVAR}(\tau) = \int_0^\infty 2 \frac{\sin^4(\pi\tau f)}{(\pi\tau f)^2} S_y(f) df. \quad (\text{B.27})$$

The transfer function has the same form as those of the spin echo sequence, the difference being the normalization and that it incorporates two periods of length τ . The connection reads

$$\sigma_\varphi^2(\tau) \Big|_{\text{SE}} = 2(\pi\tau\nu_0)^2 \text{AVAR}(\tau/2). \quad (\text{B.28})$$

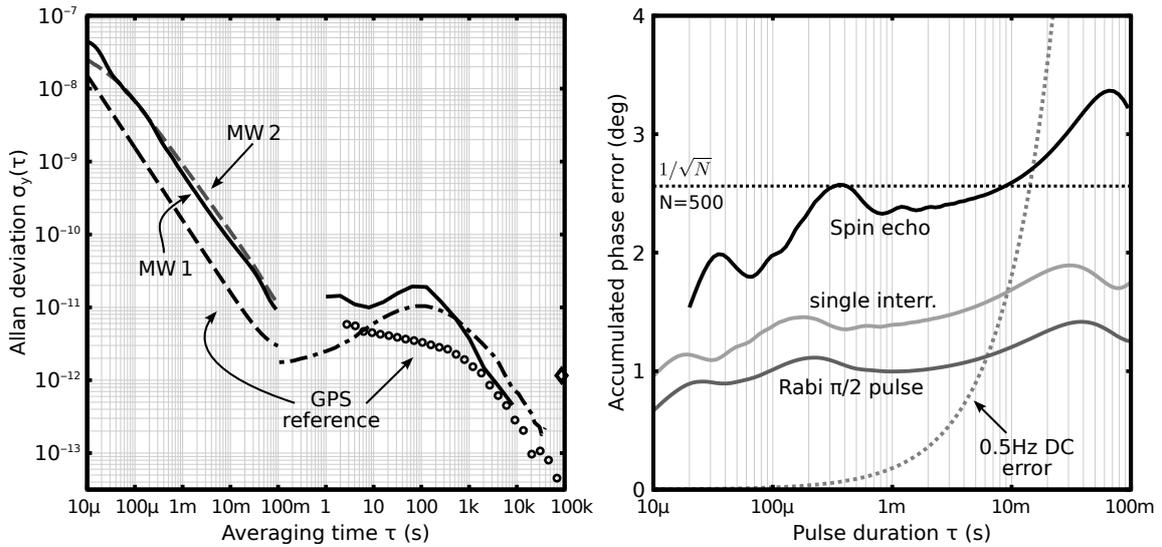


Figure B.11 Collection of Allan deviations and examples of inferred angle fluctuations The left graph shows the short time Allan deviation obtained from the specifications of the microwave generators and the GPS reference single side band (SSB) phase noise with a numerical integration from 10 Hz to 100 kHz. They approximately average down with $1/\tau$, which makes the accumulated phase noise only weakly dependent on τ . The three curves at long integration times are three characterizations of the GPS reference in the time domain. The solid black line is from the manufacturer datasheet, the dash-dotted line is a measurement of John Miles [183] and the points show the result of our measurement of the relative stability of two GPS references attached to the same antenna. The relative phase detector consisted of an RF mixer and a low-pass. The value shown as diamond point is the specified accuracy of the GPS reference (1.16×10^{-12} with one day averaging). The right graph shows the conversion of the short-term specifications of MW 1 to expected phase fluctuations of the atomic system. As comparison, the standard quantum limit for 500 atoms is shown as well as the accumulated error for a DC deviation of 0.5 Hz (\sim our magnetic field stability on the two-photon transition), which linearly grows with phase accumulation time. The Rabi pulse has the weakest phase noise sensitivity, since the spin length only gradually builds up during the pulse. The spin echo sequence is affected most, since the sidelobes of its transfer function decay more slowly than those of the single interrogation (the 'frequency leakage' is bigger).

The Allan deviation is defined as the square root of the Allan variance. In Fig. B.11 the Allan deviation for small τ is shown as inferred from the phase noise specifications and compared to direct measurements in the time domain for the GPS reference at long integration times. A scaling of σ_y with $1/\tau$ means that the phase fluctuations average down at the same rate as the integration time increases, thus keeping the phase error at the end of the sequence independent of τ . This is approximately the case for small τ due to the characteristics of the microwave and for very long τ due to the steering to the GPS clocks.

Our spin measurements show slightly better short term phase noise performance as the specification of the microwaves would suggest. For a $\pi/2$ -pulse with a length of 800μ s we get typical values of $\approx 0.5^\circ$ phase noise, extracted via the linear slope of the number squeezing parameter ξ_N versus atom number for a coherent spin state after an additional 90° rotation to convert phase noise into imbalance noise [130]. However, the understanding of the ballpark figures helps to appreciate the influence and importance of the frequency sources.

Bibliography

- [1] P. R. Berman. *Atom interferometry* (Academic Press, 1997).
- [2] A. D. Cronin, J. Schmiedmayer, and D. E. Pritchard. *Optics and interferometry with atoms and molecules*. Rev. Mod. Phys. **81**, 1051–1129 (2009).
- [3] G. M. Tino and M. A. Kasevich. *Atom Interferometry*. Proceedings of the International School of Physics "Enrico Fermi", Course CLXXXVIII, Varenna (2013).
- [4] J. Kitching, S. Knappe, and E. Donley. *Atomic Sensors – A Review*. Sensors Journal, IEEE **11**, 1749–1758 (2011).
- [5] W. D. Phillips. *Nobel Lecture: Laser cooling and trapping of neutral atoms*. Rev. Mod. Phys. **70**, 721–741 (1998).
- [6] J. R. Anglin and W. Ketterle. *Bose-Einstein condensation of atomic gases*. Nature **416**, 211–218 (2002).
- [7] N. F. Ramsey. *Experiments with separated oscillatory fields and hydrogen masers*. Rev. Mod. Phys. **62**, 541–552 (1990).
- [8] V. Giovannetti, S. Lloyd, and L. Maccone. *Quantum-Enhanced Measurements: Beating the Standard Quantum Limit*. Science **306**, 1330–1336 (2004).
- [9] L. Pezzé and A. Smerzi. *Entanglement, Nonlinear Dynamics, and the Heisenberg Limit*. Phys. Rev. Lett. **102**, 100401 (2009).
- [10] D. J. Wineland, J. J. Bollinger, W. M. Itano, F. L. Moore, and D. J. Heinzen. *Spin squeezing and reduced quantum noise in spectroscopy*. Phys. Rev. A **46**, R6797–R6800 (1992).
- [11] M. Kitagawa and M. Ueda. *Squeezed spin states*. Phys. Rev. A **47**, 5138–5143 (1993).
- [12] W. Muessel, H. Strobel, D. Linnemann, D. B. Hume, and M. K. Oberthaler. *Scalable Spin Squeezing for Quantum-Enhanced Magnetometry with Bose-Einstein Condensates*. Phys. Rev. Lett. **113**, 103004 (2014).
- [13] W. Muessel, H. Strobel, D. Linnemann, T. Zibold, B. Juliá-Díaz, and M. K. Oberthaler. *Twist-and-turn spin squeezing in Bose-Einstein condensates*. Phys. Rev. A **92**, 023603 (2015).
- [14] R. Schnabel, N. Mavalvala, D. E. McClelland, and P. K. Lam. *Quantum metrology for gravitational wave astronomy*. Nat Commun **1**, 121 (2010).

- [15] LIGO Scientific Collaboration. *A gravitational wave observatory operating beyond the quantum shot-noise limit*. Nat Phys **7**, 962–965 (2011).
- [16] C. M. Caves. *Quantum-mechanical noise in an interferometer*. Phys. Rev. D **23**, 1693–1708 (1981).
- [17] V. Giovannetti, S. Lloyd, and L. Maccone. *Quantum Metrology*. Phys. Rev. Lett. **96**, 010401 (2006).
- [18] E. Davis, G. Bentsen, and M. Schleier-Smith. *Approaching the Heisenberg limit without single-particle detection*. arXiv:1508.04110 (accepted in Phys. Rev. Lett.) (2015).
- [19] L. Pezzé, A. Smerzi, G. Khoury, J. F. Hodelin, and D. Bouwmeester. *Phase Detection at the Quantum Limit with Multiphoton Mach-Zehnder Interferometry*. Phys. Rev. Lett. **99**, 223602 (2007).
- [20] R. Krischek, C. Schwemmer, W. Wieczorek, H. Weinfurter, P. Hyllus, L. Pezzé, and A. Smerzi. *Useful Multiparticle Entanglement and Sub-Shot-Noise Sensitivity in Experimental Phase Estimation*. Phys. Rev. Lett. **107**, 080504 (2011).
- [21] H. Strobel, W. Muessel, D. Linnemann, T. Zibold, D. B. Hume, L. Pezzè, A. Smerzi, and M. K. Oberthaler. *Fisher information and entanglement of non-Gaussian spin states*. Science **345**, 424–427 (2014).
- [22] M. A. Nielsen and I. L. Chuang. *Quantum computation and quantum information* (Cambridge university press, 2010).
- [23] R. P. Feynman, F. L. Vernon, and R. W. Hellwarth. *Geometrical Representation of the Schrödinger Equation for Solving Maser Problems*. Journal of Applied Physics **28**, 49–52 (1957).
- [24] M. Schimpf and K. Svozil. *A glance at singlet states and four-partite correlations*. Mathematica Slovaca **60**, 701–722 (2010).
- [25] R. Blatt and D. Wineland. *Entangled states of trapped atomic ions*. Nature **453**, 1008–1015 (2008).
- [26] A. Sørensen and K. Mølmer. *Quantum Computation with Ions in Thermal Motion*. Phys. Rev. Lett. **82**, 1971–1974 (1999).
- [27] I. D. Leroux, M. H. Schleier-Smith, and V. Vuletić. *Implementation of Cavity Squeezing of a Collective Atomic Spin*. Phys. Rev. Lett. **104**, 073602 (2010).
- [28] F. Haas, J. Volz, R. Gehr, J. Reichel, and J. Estève. *Entangled States of More Than 40 Atoms in an Optical Fiber Cavity*. Science **344**, 180–183 (2014).
- [29] J. Schwinger. *On Angular Momentum*. Harvard University; Nuclear Development Associates, Inc. (1952).

- [30] P. Jordan. *Der Zusammenhang der symmetrischen und linearen Gruppen und das Mehrkörperproblem*. Zeitschrift für Physik **94**, 531–535 (1935).
- [31] F. T. Arecchi, E. Courtens, R. Gilmore, and H. Thomas. *Atomic Coherent States in Quantum Optics*. Phys. Rev. A **6**, 2211–2237 (1972).
- [32] J. M. Radcliffe. *Some properties of coherent spin states*. J. Phys. A: Gen. Phys. **4**, 313 (1971).
- [33] C. T. Lee. *Q representation of the atomic coherent states and the origin of fluctuations in superfluorescence*. Phys. Rev. A **30**, 3308–3310 (1984).
- [34] W. Vogel. *Nonclassical States: An Observable Criterion*. Phys. Rev. Lett. **84**, 1849–1852 (2000).
- [35] T. Kiesel, W. Vogel, B. Hage, and R. Schnabel. *Direct Sampling of Negative Quasiprobabilities of a Squeezed State*. Phys. Rev. Lett. **107**, 113604 (2011).
- [36] T. Kiesel, W. Vogel, S. L. Christensen, J.-B. Béguin, J. Appel, and E. S. Polzik. *Atomic nonclassicality quasiprobabilities*. Phys. Rev. A **86**, 042108 (2012).
- [37] U. Leonhardt and H. Paul. *Measuring the quantum state of light*. Progress in Quantum Electronics **19**, 89 – 130 (1995).
- [38] T. Kiesel and W. Vogel. *Nonclassicality filters and quasiprobabilities*. Phys. Rev. A **82**, 032107 (2010).
- [39] J. P. Dowling, G. S. Agarwal, and W. P. Schleich. *Wigner distribution of a general angular-momentum state: Applications to a collection of two-level atoms*. Phys. Rev. A **49**, 4101–4109 (1994).
- [40] A. Micheli, D. Jaksch, J. I. Cirac, and P. Zoller. *Many-particle entanglement in two-component Bose-Einstein condensates*. Phys. Rev. A **67**, 013607 (2003).
- [41] E. Nicklas. *A new tool for miscibility control: Linear coupling*. Ph.D. thesis. Universität Heidelberg (2013).
- [42] B. Vlastakis, G. Kirchmair, Z. Leghtas, S. E. Nigg, L. Frunzio, S. M. Girvin, M. Mirrahimi, M. H. Devoret, and R. J. Schoelkopf. *Deterministically Encoding Quantum Information Using 100-Photon Schrödinger Cat States*. Science **342**, 607–610 (2013).
- [43] G. Kirchmair, B. Vlastakis, Z. Leghtas, S. E. Nigg, H. Paik, E. Ginossar, M. Mirrahimi, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf. *Observation of quantum state collapse and revival due to the single-photon Kerr effect*. Nature **495**, 205–209 (2013).
- [44] D. Leibfried, E. Knill, S. Seidelin, J. Britton, R. B. Blakestad, J. Chiaverini, D. B. Hume, W. M. Itano, J. D. Jost, C. Langer, R. Ozeri, R. Reichle, and D. J. Wineland. *Creation of a six-atom ‘Schrödinger cat’ state*. Nature **438**, 639–642 (2005).

- [45] T. Monz, P. Schindler, J. T. Barreiro, M. Chwalla, D. Nigg, W. A. Coish, M. Harlander, W. Hänsel, M. Hennrich, and R. Blatt. *14-Qubit Entanglement: Creation and Coherence*. Phys. Rev. Lett. **106**, 130506 (2011).
- [46] C. A. Sackett, D. Kielpinski, B. E. King, C. Langer, V. Meyer, C. J. Myatt, M. Rowe, Q. A. Turchette, W. M. Itano, D. J. Wineland, and C. Monroe. *Experimental entanglement of four particles*. Nature **404**, 256–259 (2000).
- [47] D. Leibfried, B. DeMarco, V. Meyer, D. Lucas, M. Barrett, J. Britton, W. M. Itano, B. Jenlenkovic, C. Langer, T. Rosenband, and D. J. Wineland. *Experimental demonstration of a robust, high-fidelity geometric two ion-qubit phase gate*. Nature **422**, 412–415 (2003).
- [48] D. Leibfried, M. D. Barrett, T. Schaetz, J. Britton, J. Chiaverini, W. M. Itano, J. D. Jost, C. Langer, and D. J. Wineland. *Toward Heisenberg-Limited Spectroscopy with Multiparticle Entangled States*. Science **304**, 1476–1478 (2004).
- [49] G. Breitenbach, S. Schiller, and J. Mlynek. *Measurement of the quantum states of squeezed light*. Nature **387**, 471–475 (1997).
- [50] G. M. D’Ariano, M. G. A. Paris, and M. F. Sacchi. *On the parametric approximation in quantum optics*. Nuovo Cim. B **114**, 339–354 (1999).
- [51] M. O. Scully and M. S. Zubairy. *Quantum optics* (Cambridge University Press, 1997).
- [52] P. Meystre and M. Sargent. *Elements of quantum optics* (Springer Science & Business Media, 2007).
- [53] E. Nicklas, H. Strobel, T. Zibold, C. Gross, B. A. Malomed, P. G. Kevrekidis, and M. K. Oberthaler. *Rabi Flopping Induces Spatial Demixing Dynamics*. Phys. Rev. Lett. **107**, 193001 (2011).
- [54] H. Lipkin, N. Meshkov, and A. Glick. *Validity of many-body approximation methods for a solvable model: (I). Exact solutions and perturbation theory*. Nuclear Physics **62**, 188 – 198 (1965).
- [55] G. Mazza and M. Fabrizio. *Dynamical quantum phase transitions and broken-symmetry edges in the many-body eigenvalue spectrum*. Phys. Rev. B **86**, 184303 (2012).
- [56] A. Smerzi, S. Fantoni, S. Giovanazzi, and S. R. Shenoy. *Quantum Coherent Atomic Tunneling between Two Trapped Bose-Einstein Condensates*. Phys. Rev. Lett. **79**, 4950–4953 (1997).
- [57] S. Raghavan, A. Smerzi, S. Fantoni, and S. R. Shenoy. *Coherent oscillations between two weakly coupled Bose-Einstein condensates: Josephson effects, π oscillations, and macroscopic quantum self-trapping*. Phys. Rev. A **59**, 620–633 (1999).
- [58] I. Marino, S. Raghavan, S. Fantoni, S. R. Shenoy, and A. Smerzi. *Bose-condensate tunneling dynamics: Momentum-shortened pendulum with damping*. Phys. Rev. A **60**, 487–493 (1999).

- [59] M. Albiez, R. Gati, J. Fölling, S. Hunsmann, M. Cristiani, and M. K. Oberthaler. *Direct Observation of Tunneling and Nonlinear Self-Trapping in a Single Bosonic Josephson Junction*. Phys. Rev. Lett. **95**, 010402 (2005).
- [60] R. Gati and M. K. Oberthaler. *A bosonic Josephson junction*. Journal of Physics B: Atomic, Molecular and Optical Physics **40**, R61 (2007).
- [61] T. Zibold, E. Nicklas, C. Gross, and M. K. Oberthaler. *Classical Bifurcation at the Transition from Rabi to Josephson Dynamics*. Phys. Rev. Lett. **105**, 204101 (2010).
- [62] B. Josephson. *Possible new effects in superconductive tunnelling*. Physics Letters **1**, 251 – 253 (1962).
- [63] M. Albiez. *Observation of nonlinear tunneling of a Bose-Einstein condensate in a single Josephson junction*. Ph.D. thesis. Heidelberg University (2005).
- [64] R. Gati. *Bose-Einstein Condensates in a Single Double Well Potential*. Ph.D. thesis. University of Heidelberg (2007).
- [65] T. Anker, M. Albiez, R. Gati, S. Hunsmann, B. Eiermann, A. Trombettoni, and M. K. Oberthaler. *Nonlinear Self-Trapping of Matter Waves in Periodic Potentials*. Phys. Rev. Lett. **94**, 020403 (2005).
- [66] K. Ochs. *A comprehensive analytical solution of the nonlinear pendulum*. European Journal of Physics **32**, 479 (2011).
- [67] E.-M. Graefe, H. J. Korsch, and M. P. Strzys. *Bose-Hubbard dimers, Viviani’s windows and pendulum dynamics*. Journal of Physics A: Mathematical and Theoretical **47**, 085304 (2014).
- [68] A. Widom, G. Megaloudis, T. Clark, J. Mutton, R. Prance, and H. Prance. *The Josephson pendulum as a nonlinear capacitor*. Journal of Low Temperature Physics **57**, 651–658 (1984).
- [69] J. M. Martinis. *Course 13 Superconducting qubits and the physics of Josephson junctions*. in *Quantum Entanglement and Information Processing École d’été de Physique des Houches Session LXXIX*. Les Houches, Vol. 79. edited by D. Estève, J.-M. Raimond, and J. Dalibard (Elsevier, 2004) pp. 487 – 520.
- [70] W. P. Schleich. *Quantum optics in phase space* (John Wiley & Sons, 2011).
- [71] W. H. Zurek. *Decoherence and the transition from quantum to classical—REVISITED*. arXiv quant-ph/0306072 (2003).
- [72] B. Juliá-Díaz, T. Zibold, M. K. Oberthaler, M. Melé-Messeguer, J. Martorell, and A. Polls. *Dynamic generation of spin-squeezed states in bosonic Josephson junctions*. Phys. Rev. A **86**, 023615 (2012).
- [73] G. Sorelli, L. Pezzè, and A. Smerzi. *Fast generation of entanglement in Bosonic Josephson Junctions*. arXiv preprint arXiv:1511.01835 (2015).

- [74] R. Schmied and P. Treutlein. *Tomographic reconstruction of the Wigner function on the Bloch sphere*. New Journal of Physics **13**, 065019 (2011).
- [75] F. Y. Edgeworth. *On the Probable Errors of Frequency-Constants*. Journal of the Royal Statistical Society **71**, 381–397 (1908).
- [76] R. A. Fisher. *On the Mathematical Foundations of Theoretical Statistics*. Phil. Trans. R. Soc. A **222**, 309–368 (1922).
- [77] R. A. Fisher. *Theory of Statistical Estimation*. Mathematical Proceedings of the Cambridge Philosophical Society **22**, 700–725 (1925).
- [78] H. Cramér. *Mathematical Methods of Statistics* (Princeton University Press, 1946).
- [79] C. R. Rao. *Information and the Accuracy Attainable in the Estimation of Statistical Parameters*. Bulletin of Cal. Math. Soc. **37**, 81–91 (1945).
- [80] G. Casella and R. L. Berger. *Statistical Inference*. 2nd ed. (Duxbury, 2002).
- [81] C. W. Helstrom. *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
- [82] F. Nielsen. *Cramer-Rao Lower Bound and Information Geometry*. ArXiv e-prints (2013).
- [83] L. L. Cam. *Asymptotic Methods in Statistical Decision Theory* (Springer, 1986).
- [84] A. W. van der Vaart. *Asymptotic Statistics* (Cambridge University Press, 1998).
- [85] T. Moroder, P. Hyllus, G. Tóth, C. Schwemmer, A. Niggelbaum, S. Gaile, O. Gühne, and H. Weinfurter. *Permutationally invariant state reconstruction*. New Journal of Physics **14**, 105001 (2012).
- [86] C. Schwemmer, L. Knips, D. Richart, H. Weinfurter, T. Moroder, M. Kleinmann, and O. Gühne. *Systematic Errors in Current Quantum State Tomography Tools*. Phys. Rev. Lett. **114**, 080403 (2015).
- [87] W. K. Wootters. *Statistical distance and Hilbert space*. Phys. Rev. D **23**, 357–362 (1981).
- [88] A. DasGupta. *Probability for Statistics and Machine Learning* (Springer, 2011).
- [89] B. Efron. *The Jackknife, the Bootstrap and Other Resampling Plans* (Society for Industrial and Applied Mathematics, 1982).
- [90] B. Efron. *Bootstrap Methods: Another Look at the Jackknife*. Ann. Statist. **7**, 1–26 (1979).
- [91] R. G. Miller. *The Jackknife—A Review*. Biometrika **61**, 1–15 (1974).
- [92] V. Giovannetti, S. Lloyd, and L. Maccone. *Advances in quantum metrology*. Nat Photon **5**, 222–229 (2011).

- [93] G. Tóth and I. Apellaniz. *Quantum metrology from a quantum information science perspective*. Journal of Physics A: Mathematical and Theoretical **47**, 424006 (2014).
- [94] L. Pezzè and A. Smerzi. *Quantum theory of phase estimation*. in *Atom Interferometry, Proceedings of the International School of Physics "Enrico Fermi", Course 188, Varenna* (IOS Press, Amsterdam, 2014) p. 691. arXiv:1411.5164 [quant-ph] .
- [95] S. L. Braunstein and C. M. Caves. *Statistical distance and the geometry of quantum states*. Phys. Rev. Lett. **72**, 3439–3443 (1994).
- [96] J. J. Bollinger, W. M. Itano, D. J. Wineland, and D. J. Heinzen. *Optimal frequency measurements with maximally correlated states*. Phys. Rev. A **54**, R4649–R4652 (1996).
- [97] S. L. Braunstein, C. M. Caves, and G. Milburn. *Generalized Uncertainty Relations: Theory, Examples, and Lorentz Invariance*. Annals of Physics **247**, 135 – 173 (1996).
- [98] L. Mandelstam and I. Tamm. *The Uncertainty Relation between Energy and Time in Non-Relativistic Quantum Mechanics*. J. Phys. (USSR) **9**, 249 (1945).
- [99] A. D. Ludlow, M. M. Boyd, J. Ye, E. Peik, and P. O. Schmidt. *Optical atomic clocks*. Rev. Mod. Phys. **87**, 637–701 (2015).
- [100] D. J. Wineland, J. J. Bollinger, W. M. Itano, and D. J. Heinzen. *Squeezed atomic states and projection noise in spectroscopy*. Phys. Rev. A **50**, 67–88 (1994).
- [101] A. S. Sørensen and K. Mølmer. *Entanglement and Extreme Spin Squeezing*. Phys. Rev. Lett. **86**, 4431–4434 (2001).
- [102] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. *Quantum entanglement*. Rev. Mod. Phys. **81**, 865–942 (2009).
- [103] A. Sørensen, L.-M. Duan, J. I. Cirac, and P. Zoller. *Many-particle entanglement with Bose-Einstein condensates*. Nature **409**, 63–66 (2001).
- [104] G. Tóth, C. Knapp, O. Gühne, and H. J. Briegel. *Optimal Spin Squeezing Inequalities Detect Bound Entanglement in Spin Models*. Phys. Rev. Lett. **99**, 250405 (2007).
- [105] P. Hyllus, W. Laskowski, R. Krischek, C. Schwemmer, W. Wieczorek, H. Weinfurter, L. Pezzé, and A. Smerzi. *Fisher information and multiparticle entanglement*. Phys. Rev. A **85**, 022321 (2012).
- [106] C. Gross, T. Zibold, E. Nicklas, J. Estève, and M. K. Oberthaler. *Nonlinear atom interferometer surpasses classical precision limit*. Nature **464**, 1165–1169 (2010).
- [107] C. Gross. *Spin squeezing and non-linear atom interferometry with Bose-Einstein condensates*. Ph.D. thesis. University of Heidelberg (2010).
- [108] R. McConnell, H. Zhang, J. Hu, S. Cuk, and V. Vuletic. *Entanglement with negative Wigner function of almost 3,000 atoms heralded by one photon*. Nature **519**, 439–442 (2015).

- [109] G. Barontini, L. Hohmann, F. Haas, J. Estève, and J. Reichel. *Deterministic generation of multiparticle entanglement by quantum Zeno dynamics*. *Science* **349**, 1317–1321 (2015).
- [110] B. Lücke, J. Peise, G. Vitagliano, J. Arlt, L. Santos, G. Tóth, and C. Klempt. *Detecting Multiparticle Entanglement of Dicke States*. *Phys. Rev. Lett.* **112**, 155304 (2014).
- [111] F. Benatti and D. Braun. *Sub-shot-noise sensitivities without entanglement*. *Phys. Rev. A* **87**, 012340 (2013).
- [112] W. Dür, G. Vidal, and J. I. Cirac. *Three qubits can be entangled in two inequivalent ways*. *Phys. Rev. A* **62**, 062314 (2000).
- [113] S. M. Tan, D. F. Walls, and M. J. Collett. *Nonlocality of a single photon*. *Phys. Rev. Lett.* **66**, 252–255 (1991).
- [114] S. Ashhab, K. Maruyama, and F. Nori. *Observing quantum nonlocality in the entanglement between modes of massive particles*. *Phys. Rev. A* **75**, 022108 (2007).
- [115] M. C. Tichy, F. Mintert, and A. Buchleitner. *Essential entanglement for atomic and molecular physics*. *Journal of Physics B: Atomic, Molecular and Optical Physics* **44**, 192001 (2011).
- [116] U. Marzolino and A. Buchleitner. *Quantum teleportation with identical particles*. *Phys. Rev. A* **91**, 032316 (2015).
- [117] F. Benatti, R. Floreanini, and U. Marzolino. *Entanglement and squeezing with identical particles: ultracold atom quantum metrology*. *Journal of Physics B: Atomic, Molecular and Optical Physics* **44**, 091001 (2011).
- [118] N. Killoran, M. Cramer, and M. B. Plenio. *Extracting Entanglement from Identical Particles*. *Phys. Rev. Lett.* **112**, 150501 (2014).
- [119] R. Prevedel, G. Cronenberg, M. S. Tame, M. Paternostro, P. Walther, M. S. Kim, and A. Zeilinger. *Experimental Realization of Dicke States of up to Six Qubits for Multiparty Quantum Networking*. *Phys. Rev. Lett.* **103**, 020503 (2009).
- [120] J.-W. Pan, Z.-B. Chen, C.-Y. Lu, H. Weinfurter, A. Zeilinger, and M. Żukowski. *Multiphoton entanglement and interferometry*. *Rev. Mod. Phys.* **84**, 777–838 (2012).
- [121] C. Silberhorn, P. K. Lam, O. Weiß, F. König, N. Korolkova, and G. Leuchs. *Generation of Continuous Variable Einstein-Podolsky-Rosen Entanglement via the Kerr Nonlinearity in an Optical Fiber*. *Phys. Rev. Lett.* **86**, 4267–4270 (2001).
- [122] K. Wagner, J. Janousek, S. Armstrong, J.-F. Morizur, P. K. Lam, and H.-A. Bachor. *Asymmetric EPR entanglement in continuous variable systems*. *Journal of Physics B: Atomic, Molecular and Optical Physics* **47**, 225502 (2014).

- [123] V. Handchen, T. Eberle, S. Steinlechner, A. Sambrowski, T. Franz, R. F. Werner, and R. Schnabel. *Observation of one-way Einstein-Podolsky-Rosen steering*. Nat Photon **6**, 596–599 (2012).
- [124] T. Eberle, V. Händchen, J. Duhme, T. Franz, R. F. Werner, and R. Schnabel. *Strong Einstein-Podolsky-Rosen entanglement from a single squeezed light source*. Phys. Rev. A **83**, 052329 (2011).
- [125] N. Bar-Gill, C. Gross, I. Mazets, M. Oberthaler, and G. Kurizki. *Einstein-Podolsky-Rosen Correlations of Ultracold Atomic Gases*. Phys. Rev. Lett. **106**, 120404 (2011).
- [126] Q. Y. He, M. D. Reid, T. G. Vaughan, C. Gross, M. Oberthaler, and P. D. Drummond. *Einstein-Podolsky-Rosen Entanglement Strategies in Two-Well Bose-Einstein Condensates*. Phys. Rev. Lett. **106**, 120405 (2011).
- [127] P. Hyllus, L. Pezzé, A. Smerzi, and G. Tóth. *Entanglement and extreme spin squeezing for a fluctuating number of indistinguishable particles*. Phys. Rev. A **86**, 012337 (2012).
- [128] A. Weller. *Dynamics and Interaction of Dark Solitons in Bose-Einstein Condensates*. Ph.D. thesis. University of Heidelberg (2009).
- [129] T. Zibold. *Classical Bifurcation and Entanglement Generation in an Internal Bosonic Josephson Junction*. Ph.D. thesis. University of Heidelberg (2012).
- [130] W. Muessel. *Scalable Spin Squeezing for Quantum-Enhanced Magnetometry with Bose-Einstein Condensates*. Ph.D. thesis. University of Heidelberg (2014).
- [131] E. G. M. van Kempen, S. J. J. M. F. Kokkelmans, D. J. Heinzen, and B. J. Verhaar. *Interisotope Determination of Ultracold Rubidium Interactions from Three High-Precision Experiments*. Phys. Rev. Lett. **88**, 093201 (2002).
- [132] M. Erhard, H. Schmaljohann, J. Kronjäger, K. Bongs, and K. Sengstock. *Measurement of a mixed-spin-channel Feshbach resonance in ^{87}Rb* . Phys. Rev. A **69**, 032705 (2004).
- [133] A. Widera, O. Mandel, M. Greiner, S. Kreim, T. W. Hänsch, and I. Bloch. *Entanglement Interferometry for Precision Measurement of Atomic Scattering Properties*. Phys. Rev. Lett. **92**, 160406 (2004).
- [134] A. M. Kaufman, R. P. Anderson, T. M. Hanna, E. Tiesinga, P. S. Julienne, and D. S. Hall. *Radio-frequency dressing of multiple Feshbach resonances*. Phys. Rev. A **80**, 050701 (2009).
- [135] G. Reinaudi, T. Lahaye, Z. Wang, and D. Guéry-Odelin. *Strong saturation absorption imaging of dense clouds of ultracold atoms*. Opt. Lett. **32**, 3143–3145 (2007).
- [136] C. F. Ockeloen, A. F. Tauschinsky, R. J. C. Spreeuw, and S. Whitlock. *Detection of small atom numbers through image processing*. Phys. Rev. A **82**, 061606 (2010).

- [137] W. Muessel, H. Strobel, M. Joos, E. Nicklas, I. Stroescu, J. Tomkovič, D. B. Hume, and M. K. Oberthaler. *Optimized absorption imaging of mesoscopic atomic clouds*. Applied Physics B **113**, 69–73 (2013).
- [138] O. Morsch and M. Oberthaler. *Dynamics of Bose-Einstein condensates in optical lattices*. Rev. Mod. Phys. **78**, 179–215 (2006).
- [139] W. Zwerger. *Mott–Hubbard transition of cold atoms in optical lattices*. J. Opt. B **5**, S9 (2003).
- [140] R. Grimm, M. Weidemüller, and Y. N. Ovchinnikov. *Optical dipole traps for neutral atoms*. Adv. At. Mol. Opt. Phys. **42**, 170 (2000).
- [141] D. A. Steck. *Rubidium 87 D line data*. rev. **2.1.4** (2010).
- [142] Hélène Perrin. *Spins and fields*. Les Houches lectures on adiabatic potentials (2013).
- [143] C. Cohen-Tannoudji and D. Guéry-Odelin. *Advances in Atomic Physics: An Overview* (World Scientific, 2011).
- [144] N. Lundblad, M. Schlosser, and J. V. Porto. *Experimental observation of magic-wavelength behavior of ^{87}Rb atoms in an optical lattice*. Phys. Rev. A **81**, 031611 (2010).
- [145] R. Chicireanu, K. D. Nelson, S. Olmschenk, N. Lundblad, A. Derevianko, and J. V. Porto. *Differential Light-Shift Cancellation in a Magnetic-Field-Insensitive Transition of ^{87}Rb* . Phys. Rev. Lett. **106**, 063002 (2011).
- [146] F. Dalfovo, S. Giorgini, L. P. Pitaevskii, and S. Stringari. *Theory of Bose-Einstein condensation in trapped gases*. Rev. Mod. Phys. **71**, 463–512 (1999).
- [147] S. Tojo, T. Hayashi, T. Tanabe, T. Hirano, Y. Kawaguchi, H. Saito, and M. Ueda. *Spin-dependent inelastic collisions in spin-2 Bose-Einstein condensates*. Phys. Rev. A **80**, 042704 (2009).
- [148] S. Tojo, Y. Taguchi, Y. Masuyama, T. Hayashi, H. Saito, and T. Hirano. *Controlling phase separation of binary Bose-Einstein condensates via mixed-spin-channel Feshbach resonance*. Phys. Rev. A **82**, 033609 (2010).
- [149] J. Söding, D. Guéry-Odelin, P. Desbiolles, F. Chevy, H. Inamori, and J. Dalibard. *Three-body decay of a rubidium Bose–Einstein condensate*. Applied Physics B **69**, 257–261 (1999).
- [150] C. Chin, R. Grimm, P. Julienne, and E. Tiesinga. *Feshbach resonances in ultracold gases*. Rev. Mod. Phys. **82**, 1225–1286 (2010).
- [151] E. L. Hahn. *Spin Echoes*. Phys. Rev. **80**, 580–594 (1950).
- [152] J. Řeháček, Z. Hradil, and M. Ježek. *Iterative algorithm for reconstruction of entangled states*. Phys. Rev. A **63**, 040303 (2001).

- [153] A. I. Lvovsky. *Iterative maximum-likelihood reconstruction in quantum homodyne tomography*. Journal of Optics B: Quantum and Semiclassical Optics **6**, S556 (2004).
- [154] J. Řeháček, Z. c. v. Hradil, E. Knill, and A. I. Lvovsky. *Diluted maximum-likelihood algorithm for quantum tomography*. Phys. Rev. A **75**, 042108 (2007).
- [155] Y. S. Teo, B. Stoklasa, B.-G. Englert, J. Řeháček, and Z. c. v. Hradil. *Incomplete quantum state estimation: A comprehensive study*. Phys. Rev. A **85**, 042317 (2012).
- [156] K. Mølmer, Y. Castin, and J. Dalibard. *Monte Carlo wave-function method in quantum optics*. JOSA B **10**, 524–538 (1993).
- [157] Y. Li. *Spin squeezing in Bose-Einstein condensates*. Ph.D. thesis. ENS Paris (2010).
- [158] A. Sinatra, J.-C. Dornstetter, and Y. Castin. *Spin squeezing in Bose-Einstein condensates: Limits imposed by decoherence and non-zero temperature*. Frontiers of Physics **7**, 86–97 (2012).
- [159] Y. Li, Y. Castin, and A. Sinatra. *Optimum Spin Squeezing in Bose-Einstein Condensates with Particle Losses*. Phys. Rev. Lett. **100**, 210401 (2008).
- [160] P. Hauke, M. Heyl, L. Tagliacozzo, and P. Zoller. *Measuring multipartite entanglement via dynamic susceptibilities*. arXiv:1509.01739 (2015).
- [161] J. G. Bohnet, B. C. Sawyer, J. W. Britton, M. L. Wall, A. M. Rey, M. Foss-Feig, and J. J. Bollinger. *Quantum spin dynamics and entanglement generation with hundreds of trapped ions*. arXiv:1512.03756 (2015).
- [162] S. L. Christensen, J.-B. Béguin, E. Bookjans, H. L. Sørensen, J. H. Müller, J. Appel, and E. S. Polzik. *Quantum interference of a single spin excitation with a macroscopic atomic ensemble*. Phys. Rev. A **89**, 033801 (2014).
- [163] B. Lücke, M. Scherer, J. Kruse, L. Pezzé, F. Deuretzbacher, P. Hyllus, O. Topic, J. Peise, W. Ertmer, J. Arlt, L. Santos, A. Smerzi, and C. Klempt. *Twin Matter Waves for Interferometry Beyond the Classical Limit*. Science **334**, 773–776 (2011).
- [164] C. Gross, H. Strobel, E. Nicklas, T. Zibold, N. Bar-Gill, G. Kurizki, and M. K. Oberthaler. *Atomic homodyne detection of continuous-variable entangled twin-atom states*. Nature **480**, 219–223 (2011).
- [165] C. D. Hamley, C. S. Gerving, T. M. Hoang, E. M. Bookjans, and M. S. Chapman. *Spin-nematic squeezed vacuum in a quantum gas*. Nat Phys **8**, 305–308 (2012).
- [166] J. Peise, I. Kruse, K. Lange, B. Lücke, L. Pezze, J. Arlt, W. Ertmer, K. Hammerer, L. Santos, A. Smerzi, and C. Klempt. *Satisfying the Einstein-Podolsky-Rosen criterion with massive particles*. Nat Commun **6** (2015).
- [167] T. M. Hoang, C. S. Gerving, B. J. Land, M. Anquez, C. D. Hamley, and M. S. Chapman. *Dynamic Stabilization of a Quantum Many-Body Spin System*. Phys. Rev. Lett. **111**, 090403 (2013).

- [168] C. Gerving, T. Hoang, B. Land, M. Anquez, C. Hamley, and M. Chapman. *Non-equilibrium dynamics of an unstable quantum pendulum explored in a spin-1 Bose-Einstein condensate*. Nat Commun **3**, 1169 (2012).
- [169] J. Tomkovič, W. Muessel, H. Strobel, S. Löck, P. Schlagheck, R. Ketzmerick, and M. K. Oberthaler. *Observing the emergence of chaos in a many-particle quantum system*. arXiv:1509.01809 (2015).
- [170] E. Nicklas, M. Karl, M. Höfer, A. Johnson, W. Muessel, H. Strobel, J. Tomkovič, T. Gasenzer, and M. K. Oberthaler. *Observation of Scaling in the Dynamics of a Strongly Quenched Quantum Gas*. Phys. Rev. Lett. **115**, 245301 (2015).
- [171] P. Zanardi, M. G. A. Paris, and L. Campos Venuti. *Quantum criticality as a resource for quantum estimation*. Phys. Rev. A **78**, 042105 (2008).
- [172] R. P. Krischek. *Experimental multi-photon entanglement and quantum enhanced metrology*. Ph.D. thesis. Universität München (2011).
- [173] E. Arimondo, M. Inguscio, and P. Violino. *Experimental determinations of the hyperfine structure in the alkali atoms*. Rev. Mod. Phys. **49**, 31–75 (1977).
- [174] C. W. White, W. M. Hughes, G. S. Hayne, and H. G. Robinson. *Determination of g-Factor Ratios for Free Rb⁸⁵ and Rb⁸⁷ Atoms*. Phys. Rev. **174**, 23–32 (1968).
- [175] Y. B. Ovchinnikov, K. Szymaniec, and S. Edris. *Measurement of rubidium ground-state hyperfine transition frequency using atomic fountains*. Metrologia **52**, 595 (2015).
- [176] S. Friebel, C. D’Andrea, J. Walz, M. Weitz, and T. W. Hänsch. *CO₂-laser optical lattice with cold rubidium atoms*. Phys. Rev. A **57**, R20–R23 (1998).
- [177] T. Kim and D. D. L. Chung. *Thermoelectric Behavior of Solder*. ASME. J. Electron. Packag. **125**, 161–162 (2003).
- [178] ANALOG DEVICES. *Circuit Note CN-0191. 20-Bit, Linear, Low Noise, Precision Bipolar ±10 V DC Voltage Source*. (2011).
- [179] R. W. Lewallen. *Baluns: What they do and how they do it*. ARRL Antenna Compendium **1**, 157–164 (1985).
- [180] E. Rubiola. *Phase Noise and Frequency Stability in Oscillators*. The Cambridge RF and Microwave Engineering Series (Cambridge University Press, 2010).
- [181] J. Rutman and F. Walls. *Characterization of frequency stability in precision frequency sources*. Proceedings of the IEEE **79**, 952–960 (1991).
- [182] D. Allan, M. Weiss, and J. L. Jespersen. *A frequency-domain view of time-domain characterization of clocks and time and frequency distribution systems*. in *Proceedings of the 45th Annual Symposium on Frequency Control* (1991) pp. 667–678.
- [183] J. Miles. *Notes on Trimble Thunderbolt performance and modifications*. <http://www.ke5fx.com/tbolt.htm> (2011).

Acknowledgements – Danksagung

So many people have contributed to make the past years an unforgettable experience that an exhaustive portrayal is close to impossible in a concise form.

First and most of all, I thank Markus for his constant support, the trust in my abilities from the beginning, his infectious excitement about physics as well as experimental challenges and his commitment to intense and detailed discussions while at the same time being occupied with so many other important things. I thank him for sparking my interest in coherent matter waves and giving me the great opportunity to work hands-on with ultracold atoms and being part of an ever-changing group of talented people with a vivid team spirit. I also thank him for the opportunity to present our work at several international conferences where I had the chance to meet and discuss with many interesting people and to broaden my perspective.

Many thanks to Carsten Klempt for the immediate and uncomplicated readiness to act as second referee and taking the travel effort and time to come to the oral examination.

I am grateful to Luca Pezzè and Augusto Smerzi for proposing the statistical distance method, for their generous hospitality during my stay in Firenze and in general for the fruitful collaboration during which I learned many things about and related to 'the Fisher' and 'the Hellinger' and had the pleasure of many in-depth discussions, often spiced with an irresistibly rousing temper.

Without my splendid colleagues on the BEC experiment this work would not have been possible. Together we have cultivated a collective problem-solving approach with constant open discussions and scrutiny, which proved successful in dealing with complex problems as well as seemingly 'simple' tasks. Especially when working on vital elements of the experiment, where small changes can have long-lasting implications, immediate feedback and profound second thoughts are invaluable. I thank all involved people for their commitment and the basis of friendship and trust, which is essential for such a close teamwork. We had and still have a great time inside and outside the lab.

Auch im Namen aller Beteiligten möchte ich hier ganz speziell Wolfgang danken. Noch während des Studiums und ganz besonders seit der Diplomarbeit hat er einen wichtigen Platz in meinem Leben. Auf ihn konnte ich mich jederzeit blind verlassen. Wir haben so manche Widrigkeit und Durststrecke zusammen gemeistert und uns dabei entweder gut ergänzt oder gemeinsam unsere Defizite bekämpft. Unser gegenseitiger Respekt hielt auch so manchen derben Sprüchen und Neckereien von beiden Seiten spielend stand. Ich danke dir für die schöne und immer humorvolle Zeit und dafür, dass du ein echter Freund bist. Wir wünschen dir viel Erfolg bei deinen neuen Herausforderungen und Glück auf!

Die ganze Matterwave-Gruppe ist ein fantastisch bunter und lebendiger Haufen, dass es mir eine Ehre ist, ein Teil davon zu sein. Vielen Dank für die vielen physikalischen, technischen, weltanschaulichen und grundsätzlichen Diskussionen, die interessanten Frühstücksbesprechungen und Journal-Clubs, die offene Hilfsbereitschaft und die gelebte Völkerverständigung. Mit euch hat man einfach immer etwas zu lachen, selbst wenn einem erst nicht danach zumute ist. Da möchte ich doch noch ein Zitat einstreuen: „*Und als ich meinen Teufel sah, da fand ich ihn ernst, gründlich, tief, feierlich; es war der Geist der Schwere – durch ihn fallen alle Dinge. Nicht durch Zorn, sondern durch Lachen tötet man. Auf, laßt uns den Geist der Schwere töten!*“ (F. Nietzsche, *Also sprach Zarathustra*).

Ein großer Dank gilt unseren Teamassistentinnen Christiane und Dagmar, die das organisatorische Rückgrat der Gruppe bilden und uns stets mit Rat und Tat zur Seite stehen.

Vielen Dank der Verwaltungsabteilung des KIP für die unkomplizierte Erledigung so vieler bürokratischer Aufgaben und der EDV-Abteilung für die stabile Infrastruktur und die geduldige Einrichtung immer neuer Backup-Laufwerke für unsere Labordaten.

Eine wichtige Anlaufstelle bei technischen Problemen des Laboralltags ist die Elektronik-Abteilung. Insbesondere 'unser' Jürgen hat stets ein offenes Ohr und gute Tipps parat.

Ein herzlicher Dank geht an die feinmechanische Werkstatt für die zuverlässig präzise Arbeit und ganz besonders an Herrn Spiegel für seine Geduld, die prompte Hilfsbereitschaft und seine wertvollen Tricks und Ideen.

Ganz besonders danke ich meiner Mutter, die mir das Studium ermöglicht hat, die immer zu mir hält und mich unterstützt. Durch ihren selbstlosen und starken Einsatz für meine Großeltern hat sie mir während der Promotion entscheidend den Rücken freigehalten.

Ganz zum Schluss aber dafür umso mehr danke ich meiner lieben Clarissa, die mich schon so lange durchs Leben begleitet und mich (nach eigenen Angaben) immer noch so liebt wie ich bin. Ich bin froh, dass ich dich hab!

Erklärung

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den 04.02.2016

.....

