Optimization of an Atom-Atom Quantum Network Link using Machine Learning

Sebastian Eppelt

Munich, March 2021

Optimization of an Atom-Atom Quantum Network Link using Machine Learning

Sebastian Eppelt

Munich, the 08th of March 2021

Master's Thesis
Faculty of Physics
Ludwig-Maximilians-Universität
Munich

handed in by
Sebastian Eppelt
from Ulm

supervised by
Prof. Dr. Harald Weinfurter

Optimierung eines Atom-Atom-Quantennetzwerklinks mit maschinellem Lernen

Sebastian Eppelt

München, den 08. März 2021

Masterarbeit
Fakultät für Physik
Ludwig-Maximilians-Universität
München

vorgelegt von
Sebastian Eppelt
aus Ulm

betreut von
Prof. Dr. Harald Weinfurter

Abstract

While quantum mechanics is a research field of physics with fundamental questions about its nature still pending, its possible applications in complex computation or secure communication are already being investigated. This thesis is based on an experiment where long-distance entanglement between two single rubidium-87 atoms, separated by 400 m, can be generated and used to reject local hidden variable theories through a loophole-free Bell test and to demonstrate key ingredients for future quantum networks.

The main goal of this thesis was to investigate, using fast online machine learning techniques, if, how, and to what extent two major aspects of the experiment can be improved. First, the initial state preparation of the atom via optical pumping, that is necessary for the lambda level scheme that creates atom-photon-entanglement as a first step in the atom-atom-entanglement procedure is studied. Second, we look into the readout process of the atomic state that consists of a Zeeman-level selective ionization scheme. Both processes rely on the interaction between the atom and external light fields. Their respective shape can be controlled by acousto-optic modulators which we will operate using an arbitrary waveform generator.

For finding the best way of steering the two processes, we have developed online optimizers that employ machine learning techniques, in particular differential evolution and Gaussian process regression. The optimizers are not only capable of global optimization and work well even in noisy environments, but are also capable of generating statistical models for the processes that help us understand the relation between input parameters to the experiment and the returned experimental performance.

Contents

1. Introduction

Quantum mechanics (QM) is a fundamental theory in physics, whose development began around the turn of the 20th century when it was first realized that classical mechanics would fail on the (sub)atomic scale [1]. Notable early problems that could only be solved by requiring the quantization of nature at the microscopic level were Max Planck's radiation law in 1900 [2, 3] as well as Albert Einstein's explanation of the photo-electric effect in 1905 [4]. One of the defining features quantum systems is entanglement, a term derived from the German *Verschränkung*. It is highly non-classical, as we will see soon, is used in quantum information processing and communication [5], and has applications in lithography and metrology [6]. If a system of particles is entangled, it is impossible to describe its constituents independently, which has a profound effect on the correlations of measurements with entangled particles. Some claimed that "the quantum mechanical description of physical reality [...] is not complete" [7] as phenomena resulting from entanglement seemed to violate local-realism. As a consequence, an almost three-decades long debate about the possibility and necessity of extending QM to a local hidden variable (LHV) theory began [8, 9]. Eventually, John Bell succeeded in showing that in certain experimental scenarios involving pairs of entangled particles LHV theories would predict different outcomes than quantum theory [10]. Bell's theorem can be formulated as an inequality whose most prominent form is the CHSH inequality [11].

Experimental test of Bell's theorem need to fulfill a set of assumptions, that otherwise turn into loopholes for LHVs. Early works by Freedman and Clauser [12] in 1972 and Aspect, Dalibard, and Roger [13, 14] in 1982 left two major loopholes open, that were closed individually by Weihs et al. (1998) and Rowe et al. (2001). Closing both loopholes in one experiment was first accomplished in Delft [15], Vienna [16], and Boulder [17] as well as by Wenjamin Rosenfeld, Harald Weinfurter et al. at LMU Munich [18]. The latter experiment employed heralded, long-distance entanglement over almost 400 m of two single rubidium-87 atoms. It can also serve as a demonstrator for quantum technologies like quantum network links and quantum cryptography, in particular device independent quantum key distribution (DIQKD). The experiment creates entanglement between the two individually trapped atoms through a two-step process. First, the hyperfine state of each atom is entangled with the polarization of a photon. Then, the entanglement is swapped onto the two atoms by performing a joint Bell state measurement (BSM) on the two photons. This thesis will investigate two aspects of the experiment: the initial state preparation, that is needed before atom-photon entanglement can be generated via spontaneous decay, and the atomic state readout, that is required to verify the atom-photon or atom-atom-entanglement. In this thesis we will explore if and how, by modifying the involved laser fields, it is possible to improve the performance of initial state preparation and readout. In particular, we investigate if this can be

achieved by optimizing the radio-frequency (RF) input of acousto-optic modulators (AOM) using an arbitrary waveform generator.

AOMs are widely used across quantum optics experiments. They offer fast and efficient switching with rates of several MHz. AOMs have been employed in real-time holographic display [19], synthesized aperture radar [20], or for performing fast, high-fidelity quantum state tomography [21]. Driving an AOM simultaneously with multiple RF frequencies can be interesting on a technical level, for example for improving the pointing stability of the output beam during intensity ramps [22]. It also facilitates the generation of defect-free 1D [23] and 2D [24] arrays of cold atoms. These methods provide a pathway towards a programmable quantum simulator that could be used to explore quantum many-body phenomena [25].

For the purpose of optimizing the laser fields in a closed-loop configuration we will employ machine learning (ML) techniques, in particular Gaussian process regression and differential evolution. ML is a branch of artificial intelligence that produces algorithms capable of improving their performance over time by studying data rather than being explicitly programmed to do so. Today, ML has a prominent role in society and industry as well as science. ML algorithms have gained traction in physical research in recent years [26], with applications in statistical physics [27], cosmology [28] and more. ML can be used for quantum optical control [29], both in theory [30, 31] or experiment [32, 33]. Our optimization scheme is based on the open-source M-LOOP package which has been developed specifically for the use in scientific experiments and has previously demonstrated the optimized production of Bose-Einstein-condensates [34].

This thesis will be structured as follows: Chapter 2 describes the quantum mechanical concepts behind and experimental methods and setup used for the creation of (heralded) atom-atom entanglement. Chapter 3 introduced Gaussian process regression and differential evolution, the algorithms used in the optimizations of the initial state preparation and readout process, that are presented in chapter 4.

2. Entangling Single Rubidium Atoms

2.1. Quantum Mechanics and Entangled Two Qubit Systems

The mathematical framework for quantum mechanics was created in the mid-1920s when matrix mechanics was first introduced by Heisenberg [35] and later fully formulated by Born, Jordan and Heisenberg [36]. Its full description would go beyond the scope of this thesis but is readily available for the interested reader for example in books by Sakurai [37] or Griffiths [38]. Nevertheless, I will attempt to summarize some key aspects of the the above mentioned works.

2.1.1. Quantum States and Density Matrices

Mathematically, in quantum mechanics a physical system is described within a Hilbert space $\mathscr{H}$ of dimension D. D can be either finite or infinite depending on the nature of the physical system at hand. The particular state of systems might be represented by a single complex vector in $\mathscr{H}$, that is either one single basis state or a superposition of multiple basis states, in which case it is referred to as pure. This principle of superposition sets quantum systems apart from classical states.It is best manifested in the famous double-slit experiment. While for classical particles we can reconstruct, which slit the particle has passed before it hit the detection screen, this information cannot be known, even in principle, in quantum mechanics [39].

For the more general case of a statistical mixture of more than one pure quantum state, one often uses the density operator or density matrix formalism to describe the system. Following the notation proposed by Dirac in 1939 [40], this thesis will denote pure states as either ket $|\phi\rangle$ or bra $\langle\phi|$ living in $\mathscr{H}$ or its dual space $\mathscr{H}^*$, respectively. A density operator takes the form of a sum over its $1 \leq N \leq \dim(\mathscr{H})$ constituent states,

$$\rho = \sum_{i=1}^{N} p_i \left|\phi_i\right\rangle \left\langle\phi_i\right| , \tag{2.1}$$

where the coefficients $p_i \leq 1$ are the probabilities to find the systems in state $|\phi_i\rangle$. Their sum and thus the trace of ρ needs to be equal to one. For the special case of $N = 1$, ρ represents a pure state. While the concrete representation of ρ is not unique, it is per definition Hermitian, thus diagonalizable and can be, in theory, brought into a form as in equation 2.1.

2.1.2. Two Level Systems

Systems with two possible distinct, i.e. orthogonal, internal states are of particular interest for various applications in quantum technologies. They are referred to as quantum bits or just qubits. In our particular setup we use two hyperfine states of the ground state of Rubidium 87 to encode a qubit as described in section 2.3.1. The underlying Hamiltonian of the system has two orthogonal eigenstates, denoted $|0\rangle$ and $|1\rangle$, respectively. An arbitrary state $|\psi\rangle$ in the two-dimensional Hilbert space $\mathscr{H}_2$ can be expressed (unambiguously up to an arbitrary global phase factor) as superposition of the two basis states

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle \in \mathscr{H}_2, \tag{2.2}$$

where the coefficients or amplitudes $\alpha, \beta \in \mathbb{C}$ need to fulfill the normalization condition $|\alpha|^2 + |\beta|^2 = 1$. Thus, we parameterize $|\psi\rangle$ by the two angles $\theta \in [0, \pi]$ and $\varphi \in [0, 2\pi]$,

$$|\psi\rangle = \cos(\theta/2)\,|0\rangle + e^{i\varphi}\sin(\theta/2)\,|1\rangle\,. \tag{2.3}$$

This lets us express a given state as a unique three dimensional vector with unit length according to

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} \sin(\theta)\cos(\varphi) \\ \sin(\theta)\sin(\varphi) \\ \cos(\theta) \end{bmatrix}. \tag{2.4}$$

All possible superposition states form the so called Bloch sphere which is shown in figure 2.1 with some important states. Note that the Bloch sphere is mathematically equivalent to the Poincaré sphere which is often used to represent the polarization states of electromagnetic waves [41].

While pure states of a two-level system occupy the surface of the Bloch sphere, mixed states lie within it. Generally, for a given state, pure or mixed, described by the 2×2 density matrix ρ one finds the representation [42]

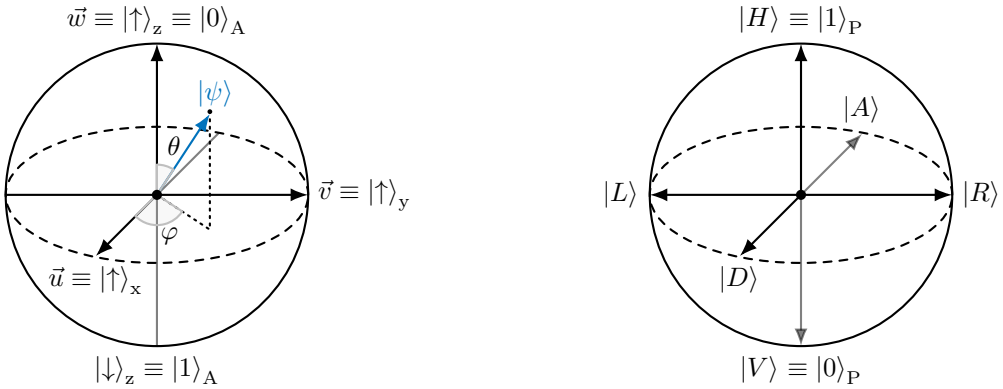
$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 2\,\mathrm{Re}(\rho_{10}) \\ 2\,\mathrm{Im}(\rho_{10}) \\ \rho_{00} - \rho_{11} \end{bmatrix}, \tag{2.5}$$

where ρ_{ij} with $i, j = 0, 1$ are the components of the density matrix.

2.1.3. Entanglement and Bell's Inequality

Quantum states that describe systems of multiple quantum particles can have another strictly non-classical property: entanglement. Mathematically, an entangled state can be best understood in contrast to its counter part, product states. Product states of N particles can be written in the form

$$|\psi_{\mathrm{product}}\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes ... \otimes |\psi_N\rangle \tag{2.6}$$



(a) Bloch sphere **(b)** Poincaré sphere

Figure 2.1.: Bloch sphere (a): The state $|\psi\rangle$ (blue) that is fully described by the polar angle φ and the azimuthal angle θ. Also indicated are the Bloch sphere representations of the atomic qubit in the three bases, relevant for later section. Poincaré sphere (b): the six major polarizations of the photon are indicated.

Any state that cannot be decomposed as in 2.6 is said to be entangled. While this definition is simple and straightforward, it is in general difficult to determine whether a given (mixed) state is entangled or not. We will now restrict ourselves to the discussion of entanglement in pure two qubit systems. While it is the simplest case of entanglement, it still captures many important concepts. A more general discussion can be found in [43].

Any state within the $2 \times 2 = 4$ dimensional Hilbert space of two qubits (basis states $|0\rangle$ and $|1\rangle$) can be written as

$$|\psi\rangle = c_{00}\,|00\rangle + c_{01}\,|01\rangle + c_{10}\,|10\rangle + c_{11}\,|11\rangle \tag{2.7}$$

where the amplitudes c_{ij} are properly normalized and we used the shorthand notation $|ij\rangle \equiv |i\rangle \otimes |j\rangle$. The orthonormal basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ can be linearily transformed into the so called Bell basis

$$\left|\Phi^+\right\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{2.8a}$$
$$\left|\Phi^-\right\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \tag{2.8b}$$
$$\left|\Psi^+\right\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \tag{2.8c}$$
$$\left|\Psi^-\right\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \tag{2.8d}$$

Those states are called maximally entangled, meaning that their von Neumann entropy $S = -\operatorname{Tr}\{\rho \ln \rho\}$ takes the maximum value of $S = 2 \ln 2$ for states in the four dimensions [44].

Entanglement has been at the center of many discussions about the nature of QM. The correlation between the measurement results of the components of an

entangled system are in contradiction to what is possible for classical measurement results and have led to the proposal of extending QM to a local hidden variable (LHV) theory. Bell's theorem of 1964 [10] put forward a criterion for testing if nature was described by such a theory. It is formulated as an inequality which is fulfilled by all LHV theories but violated by QM. It was later adapted by Clauser, Horne, Shimony, and Holt [11] whose inequality can be testes more easily in in experiments and it's violation can confirm that nature is not described by LHVs. It considers an abstract scenario in which two observers, often called Alice and Bob, each receive a qubit from a pair of qubits emitted by a source. Alice (Bob) has access to a device that can perform two different measurements depending on the input A (B) which can take the value 0 or 1 and gives an output X (Y) that is either +1 or -1. The experiment is evaluated based on the value of S_{CHSH} defined as

$$S_{\mathrm{CHSH}} := |E_{1,1} - E_{1,0}| + |E_{0,1} - E_{0,0}|, \tag{2.9}$$

where the correlators $E_{a,b}$ have the form

$$E_{a,b} = \frac{N_{a,b}^{X=Y} - N_{a,b}^{X\neq Y}}{N_{a,b}^{X=Y} + N_{a,b}^{X\neq Y}} \tag{2.10}$$

$N_{a,b}^{X=Y}$ and $N_{a,b}^{X\neq Y}$ denote the number of events with correlated and anti-correlated outputs, respectively, for inputs $A = a$ and $B = b$.

It can be shown that under a set of four assumptions (local measurements, independent measurements, independent inputs, unpredictable inputs[1]) and the correct measurements outlined in table Table 2.1 that S_{CHSH} has different bounds for LHV theories and quantum mechanics, namely

$$S_{\mathrm{CHSH}} \leq 2 \qquad \text{for LHV theories} \tag{2.11a}$$

$$\text{and } S_{\mathrm{CHSH}} = 2\sqrt{2} \qquad \text{for QM considering } \left|\Psi^-\right\rangle. \tag{2.11b}$$

For an experimental test of the CHSH inequality, the four assumptions from above turn in requirements that need to be guaranteed through the experimental design. Furthermore, each assumption opens up a "loophole" for local-realism to be true, despite a measured $S_{\mathrm{CHSH}} > 2$. Those loopholes are as follows:

1. **Locality loophole**
   No communication between Alice and Bob can be allowed and the measurement process must take place outside each other's light cone.

2. **Freedom of choice loophole**
   The inputs for Alice and Bob's measurements must be independent and ran-

[1] details found in [45, section 3.2.1]

	input	
	1	0
Alice	σ_z	σ_x
Bob	$\frac{1}{\sqrt{2}}(\sigma_z + \sigma_x)$	$\frac{1}{\sqrt{2}}(\sigma_z - \sigma_x)$

(a) Measurement operators.

	$A = 1$	$A = 0$
$B = 1$	$E_{1,1} = -1/\sqrt{2}$	$E_{0,1} = -1/\sqrt{2}$
$B = 0$	$E_{1,0} = -1/\sqrt{2}$	$E_{0,0} = +1/\sqrt{2}$

(b) Bell correlators.

Table 2.1.: Scenario for a measurement of the CHSH inequality. Inputs 1 and 0 lead to measurement of the operators shown in (a). Considering the singlet state $|\Psi^-\rangle$ from eq. 2.8d leads to the correlator in (b)
.

dom. For this typically a random number generator for each side of the experiment is employed, which in turn requires assumptions about the randomness of their outputs.

3. **Detection loophole**
   Implicitly the assumption that each created pair leads to one measurement outcome was made. In reality, detection efficiencies are smaller than 1. It can be shown that there exists a lower bound for the necessary detection efficiency depending on the experiment in question. For experiments with heralded entanglement, this bound can be shown to be 2/3 for each side.

4. **Memory loophole**
   Naturally, a given experiment can only produce finite statistics. One often assumes for the statistical analysis that the results or a repeated experiment are independent and identically distributed, which could be exploited by a LHV where the measurement outcomes depend on previous inputs and results [46]. The loophole can be addressed by a reformulation of the correlators [46] or by employing a game formalism [47, 48]

An experiment that succeeds in closing all four loophole at once is called *loophole-free*. One such experiment was the Bell test with widely separated, entangled atom-atom-pairs, that was conducted at LMU in 2016 [18]. This thesis presents work that has been done on the setup that was used for this experiment.

2.1.4. Time Evolution of a Quantum Mechanical System

After having discussed how physical systems are represented in quantum mechanics (sec. 2.1.1) it is natural to ask how such systems will evolve over time. This question was first investigated by Erwin Schrödinger [49], leading to the formulation of one of the most famous equations in physics, the time-dependent Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}\left|\Psi(t)\right\rangle = \mathcal{H}(t)\left|\Psi(t)\right\rangle \tag{2.12}$$

which completely describes the (necessarily) unitary time-evolution of a pure state $|\Psi(t)\rangle$ governed by the Hermitian Hamiltonian $\mathcal{H}$. $\mathcal{H}(t)$ itself might, in general, be time-dependent. Equation 2.12 can be generalized to also include mixed systems [44]

$$i\hbar\frac{\partial}{\partial t}\rho(t) = [\mathcal{H}(t), \rho(t)] := \mathcal{H}(t)\rho(t) - \rho(t)\mathcal{H}(t) \tag{2.13}$$

where we defined the commutator $[\cdot, \cdot]$. Equation 2.13 is known as the Von-Neumann- or Von-Neumann-Liouville equation.

Nevertheless, equations 2.12 and 2.13 are only able to capture unitary dynamics, where phase coherence is maintained at all times. This would require a completely isolated system, a situation that can only be an approximation for any real physical system. More realistically, interactions between a system and its surrounding will dissipate the system's energy and lead to spontaneous decays and randomized phases. Such systems are referred to as open quantum systems in contrast to closed quantum systems, that can be described by the Schrödinger and Von-Neumann-Liouville equations.

For open systems the situation is more complex. The most general Markovian approach to describe an open system living in the N-dimensional Hilbert space $\mathscr{H}_N$, leads to the Lindblad (or Lindblad-Gorini-Kossakowski-Sudarshan) master equation for the system's density matrix [50]

$$\frac{\partial}{\partial t}\rho(t) = \frac{-i}{\hbar}[\mathcal{H}(t), \rho(t)] + \sum_{m,n=1}^{N^2-1} h_{mn}\left(A_n\rho(t)A_m^\dagger - \frac{1}{2}\{A_m^\dagger A_n, \rho(t)\}\right) \tag{2.14}$$

where A_m form a arbitrary orthonormal basis (ONB) of the Hilbert-Schmidt operators on $\mathscr{H}_N$ and where A_{N^2} is $\propto 1$. The positive semi-definite coefficient matrix h describes the incoherent dynamics of the system. Equation 2.14 may be diagonalized to yield

$$\frac{\partial}{\partial t}\rho(t) = \frac{-i}{\hbar}[\mathcal{H}(t), \rho(t)] + \sum_{i=1}^{N} L_i\rho(t)L_i^\dagger - \frac{1}{2}\{L_i^\dagger L_i, \rho(t)\} \equiv \mathcal{L}\rho(t), \tag{2.15}$$

where L_i are the jump operators.

When dealing with systems whose incoherent dynamics can solely be attributed to spontaneous decays, another form of the Lindblad equation can be used [51] which is particularly useful for numerical implementations, namely

$$\frac{\partial}{\partial t}\rho(t) = \frac{-i}{\hbar}[\mathcal{H}(t), \rho(t)] + \frac{1}{2}\sum_{i,j=1}^{N}\Gamma_{ij}\left(|i\rangle\langle j|\,\rho(t)\,|j\rangle\langle i| - \{|j\rangle\langle j|\,, \rho(t)\}\right), \tag{2.16}$$

where Γ_{ij} is the decay rate for decays from $|i\rangle$ to $|j\rangle$.

Both equations 2.13 and 2.15 define a system of N^2 partial differential equation of which a number can be neglected due to the hermiticity $\rho_{ij} = \rho^*_{ji}$ and trace of the density matrix, leaving $N^2 - \frac{N^2-N}{2} - 1$ equations to solve. Only for a small number of simple systems and/or by using sophisticated techniques like damping bases [52], this can be done analytically. More often it is necessary to rely on numerical approximations, as was done for simulations of the optical pumping (sec. 2.4.1) and readout process (sec. 2.4.3). The "quantum jump model" [53] describing the excitation process (see sec. 4.2.1 also is based on 2.16.

2.2. Entanglement of Distant Particles

Entanglement between two widely separated ^{87}Rb atoms lies at the heart of our experiment. While it would be in theory possible to separate the atoms after having generated entanglement between them, it is more practical to keep their positions fixed and perform the entanglement in two steps: First we generate entanglement between an atom and a photon on each side and then bring the photons together and measure them jointly, whereby we transfer the entanglement to create atom-atom-entanglement. The two steps will be explained in sections 2.2.1 and 2.2.2, respectively.

2.2.1. Atom-Photon-Entanglement

In our experiment entanglement between an atom and a photon is generated through a lambda level scheme. Lambda level schemes have wide application in quantum technologies and can be employed for coherent state transfer [54, 55], electromagnetically-induced transparency (EIT) [56], or, as in our case, hybrid entanglement to name just a few.

A Lambda level scheme consist of at least two (ground) levels that are each dipole-connected to a single (excited) level, but not to each other [56]. As the excited state is higher in energy, the three levels and the possible transitions form the Greek letter lambda (Λ) (e.g. the σ-transitions in fig. 2.2), hence the name. In our experiment, our excited state $|F' = 0, m_F = 0\rangle$ can decay into the stable, three-fold degenerate ground state $|F = 0, m_F = 0, \pm 1\rangle$ by emission of a single photon. It should be noted that these states are part of a larger manifold that can, however, be neglected here. Each of the three possible decay channels for $|F' = 0, m_{F'} = 0\rangle$ leads to a distinct combination of the polarization state of the emitted photon and the angular momentum in the ground state $|F = 0, m_F\rangle$ as shown in figure 2.2.

The decay is governed by the the linewidth Γ of the excited level. As we deal with a stable ground state, for long enough times ($t >> \tau$, where $\tau = 1/\Gamma$ is the lifetime

of the excited state) we can approximate that the entire system is in a superposition state of the three ground states and the state of the emitted photon

$$|\Psi_{\text{tot}}\rangle_{\text{AP}} = \frac{1}{\sqrt{3}}(\left|\Sigma^+\right\rangle_{\text{P}} \otimes |1,-1\rangle_{\text{A}} + |\Pi\rangle_{\text{P}} \otimes |1,0\rangle_{\text{A}} + \left|\Sigma^-\right\rangle_{\text{P}} \otimes |1,+1\rangle_{\text{A}}), \tag{2.17}$$

where we neglected any remaining population in the excited state[2]. This represents a well defined quantum state with entanglement between the projection of the angular momentum of the particle onto the quantization axis and the polarization mode ($|\Sigma^\pm\rangle, |\Pi\rangle$) of the emitted photon. The polarization mode can be decomposed into the spin eigenstate ($\sigma^\pm, \pi$) and spatial and frequency modes ($|\theta,\omega\rangle$ of the photon [57], yielding

$$\left|\Sigma^\pm\right\rangle = \sqrt{\frac{3}{16\pi}(1+\cos^2\theta)}\left|\sigma^\pm\right\rangle \otimes \sum_{\theta,\omega} g_\omega |\omega,\theta\rangle\,, \tag{2.18a}$$

and

$$|\Pi\rangle = \sqrt{\frac{3}{8\pi}}\sin\theta\,|\pi\rangle \otimes \sum_{\theta,\omega} g_\omega |\omega,\theta\rangle\,. \tag{2.18b}$$

If we now opt for collecting the emitted photons along the quantization axis where $\theta = 0$ we can eliminate the case where $m_F = 0$ and the photon is π-polarized. As the

[2] corresponding to the infinite time limit $t \to \infty$

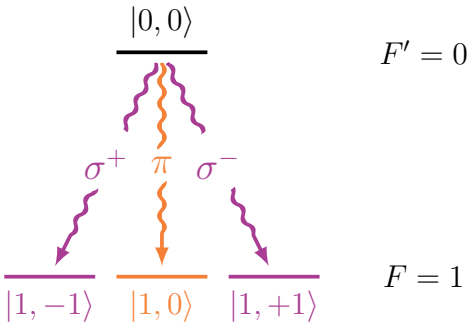


Figure 2.2.: Spontaneous decay of a quantum system from the excited state with vanishing angular momentum ($F' = 0$) into a ground state with total angular momentum $F = 1$ and thus three possible projections $m_F = 0, \pm 1$ along the quantization axis.

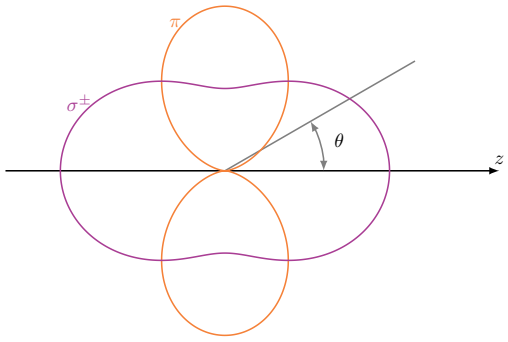


Figure 2.3.: Emission characteristics for π- and $\sigma^\pm$-polarized light, represented by the square of the expressions in 2.18. The emission is rotationally symmetric around the quantization axis. Due to the transverse nature of EMR, emission of π-polarized light along the quantization axis is forbidden.

energy difference of $|1,+1\rangle$ and $|1,-1\rangle$ is much smaller than the energy-uncertainty of the excited state, we finally arrive at the expression for a maximally entangled two-qubit state[3], formed of a circularly polarized photon and an atom which we know from section 2.1.3:

$$|\Psi\rangle_{\mathrm{AP}} = \frac{1}{\sqrt{2}}\left(\left|\sigma^+\right\rangle_{\mathrm{P}}|1,-1\rangle_{\mathrm{A}} + \left|\sigma^-\right\rangle_{\mathrm{P}}|1,+1\rangle_{\mathrm{A}}\right) \tag{2.19}$$

2.2.2. Atom-Atom-Entanglement

After having successfully created entanglement between an atom and a photon on each side, one is ready for the next step: entangling the two atoms via the entanglement swapping process. Experimentally, by means of a joint measurement of the two photons, this process enables to swap the entanglement from the two atom-photon pairs to a combined state of the two atoms [53, 58–60].

2.2.2a. Heralded Entanglement via Entanglement Swapping

We will consider two entangled qubit pairs

$$|\Psi\rangle_{\mathrm{AP},1(2)} = \frac{1}{\sqrt{2}}\left(|1\rangle_{\mathrm{A},1(2)}|0\rangle_{\mathrm{P},1(2)} + |0\rangle_{\mathrm{A},1(2)}|1\rangle_{\mathrm{AP},1(2)}\right) \tag{2.20}$$

where the subscripts denote the particle A (P) for atom (photon) and pair 1 (2) . This is equivalent to equation 2.19, with the more general notation $|0\rangle$ and $|1\rangle$ for that atomic and photonic state.

The combined state

$$\begin{aligned}|\Psi\rangle_{\mathrm{all}} &= |\Psi\rangle_{\mathrm{AP},1}|\Psi\rangle_{\mathrm{AP},2} \\ &= \frac{1}{2}\left(|10\rangle_{\mathrm{AP},1}|10\rangle_{\mathrm{AP},2} + |10\rangle_{\mathrm{AP},1}|01\rangle_{\mathrm{AP},2} + |10\rangle_{\mathrm{AP},1}|01\rangle_{\mathrm{AP},2} + |01\rangle_{\mathrm{AP},1}|01\rangle_{\mathrm{AP},2}\right)\end{aligned} \tag{2.21}$$

can be decomposed into Bell states (2.8) of two photons and two atoms to yield

$$|\Psi\rangle_{\mathrm{all}} = \left|\Psi^+\right\rangle_{\mathrm{AA}}\left|\Psi^+\right\rangle_{\mathrm{PP}} + \left|\Psi^-\right\rangle_{\mathrm{AA}}\left|\Psi^-\right\rangle_{\mathrm{PP}} + \left|\Phi^+\right\rangle_{\mathrm{AA}}\left|\Phi^+\right\rangle_{\mathrm{PP}} + \left|\Phi^-\right\rangle_{\mathrm{AA}}\left|\Phi^-\right\rangle_{\mathrm{PP}}. \tag{2.22}$$

Projection of the combined state onto a photonic Bell state

$$I_{AA} \otimes |\Psi_{\mathrm{Bell}}\rangle\langle\Psi_{\mathrm{Bell}}|_{PP}\ |\Psi\rangle_{\mathrm{all}}$$

results in an entangled atom-atom-pair in the same state. Therefore, the outcome of a BSM of the two photons will determine the entangled state of the two photons.

[3] we will adopted the simpler notation $|\psi_i\rangle \otimes |\psi_j\rangle \equiv |\psi_i\rangle|\psi_j\rangle$ from now on

The detection of the photons therefore heralds the entanglement between the atoms.

2.2.2b. Photonic Bell-State Measurement

One established method for a BSM on two photons employs a two-photon-interference on a beamsplitter [59, 61]. What happens when two indistinguishable single photons each impinging on a different input port of a beamsplitter is known as the Hong-Ou-Mandel effect [62]. This section will briefly introduce the effect with the additional polarisation degree of freedom taken into account, with a more detailed discussion in [45].

The experimental setup for the BSM (fig. 2.4) is simple. Two photons, each entangled with an atom in one of the traps, impinge onto two orthogonal input ports of a (perfect, polarization independent 50/50) beamsplitter (BS). Behind each output port sits polarization dependent beamsplitter (PBS) that separates V and H polarization. A single-photon detector is placed after each relevant PBS output.

The BSM only considers the photonic part of the entangled state $|\Psi\rangle_{\text{all}}$. It is easy to see that performing the partial trace over the atomic part results in mixed state for the two photons, with equal population in all four Bell states. Using the Fock state formalism (second quantization) [63] we can express the transformation of these input states as

$$|\Phi^{\pm}\rangle_{\text{in}} \rightarrow \quad \frac{1}{\sqrt{2}}\left(\hat{b}^{\dagger}_{1,H}\hat{b}^{\dagger}_{1,H} \pm \hat{b}^{\dagger}_{1,V}\hat{b}^{\dagger}_{1,V} - \hat{b}^{\dagger}_{2,H}\hat{b}^{\dagger}_{2,H} \mp \hat{b}^{\dagger}_{2,V}\hat{b}^{\dagger}_{2,V}\right)|0,0\rangle_{\text{out}}\,, \tag{2.23a}$$

$$|\Psi^{+}\rangle_{\text{in}} \rightarrow \quad \frac{1}{\sqrt{2}}\left(\hat{b}^{\dagger}_{1,H}\hat{b}^{\dagger}_{1,V} - \hat{b}^{\dagger}_{2,H}\hat{b}^{\dagger}_{2,V}\right)|0,0\rangle_{\text{out}}\,, \text{ and} \tag{2.23b}$$

$$|\Psi^{+}\rangle_{\text{in}} \rightarrow \quad \frac{1}{\sqrt{2}}\left(\hat{b}^{\dagger}_{1,H}\hat{b}^{\dagger}_{2,V} - \hat{b}^{\dagger}_{1,V}\hat{b}^{\dagger}_{2,H}\right)|0,0\rangle_{\text{out}}\,. \tag{2.23c}$$

$\hat{b}^{\dagger}_{i,P}$ is the creation operator for a photon at output port i^P in figure 2.4 and $|0,0\rangle_{\text{out}}$ is the vacuum state on the output side. Please not that here we refer to the *Fock state* or occupation number state with zero particles in each output port and not the qubit state "0" as in the previous and following sections.

A detailed description including possible imperfections can be found in [45, 53]. We will observe four different signatures with the single-photon detectors:

1. one photon in 1^V and 2^H or 1^H and 2^V: projection onto $|\Psi^{-}\rangle$
2. one photon in 1^V and 1^H or 2^H and 2^V: projection onto $|\Psi^{+}\rangle$
3. both photons on one detector: projection onto $|\Phi^{\pm}\rangle$ with no chance of distinguishing between the two

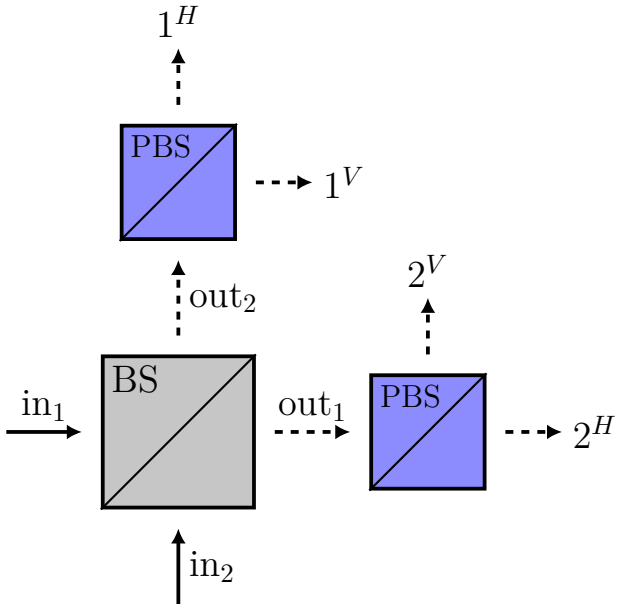


Figure 2.4.: Photonic BSM sheme: The two photons enter a beamsplitter (BS) through different input ports. Each output is split by into H and V by a polarizing beamsplitter (PBS), leading to four different outputs on single photon detectors ($1^H, 1^V, 2^H$, and 2^V).

4. one photon in 1^H and 2^H or 1^V and 2^V: only occurs due to experimental imperfections

Only signatures 1 and 2 herald the measurement of the photonic state onto $|\Psi^{\pm}\rangle_{\text{in}}$, therefore projecting $|\Psi\rangle_{\text{all}}$ onto $|\Psi^{\pm}\rangle_{\text{PP}}|\Psi^{\pm}\rangle_{\text{AA}}$ which we can express in terms of the polarization $P \in \{H, V\}$ and the atomic spin in x-direction $s_x \in \{\uparrow, \downarrow\}$ as

$$\left|\Psi^{\pm}\right\rangle_{\text{PP}} = \frac{1}{\sqrt{2}}\left(|HV\rangle \pm |VH\rangle\right) \text{ and } \left|\Psi^{\pm}\right\rangle_{\text{AA}} = \frac{1}{\sqrt{2}}\left(|{\downarrow\uparrow}\rangle_{\text{xx}} \pm |{\uparrow\downarrow}\rangle_{\text{xx}}\right). \tag{2.24}$$

For a real experimental implementation of such measurement there are various error sources that lead to a lower fidelity of the entangled atom-atom-state. First, imperfections of the beamsplitters, such as deviations from the $^{50}/_{50}$ splitting ratio or partial polarization dependence for the first beamsplitter, will lead to errors in the measurement of $|\Psi^{\pm}\rangle$ and an effective mixing of the Bell states as described in [60]. Second, the two photons might be not indistinguishable due to imperfect spectral, temporal, or spatial overlap. The former two have to be ensured by the excitation and emission process (sec. 2.4.2) of the respective photons, while the latter depends on the BSM setup.

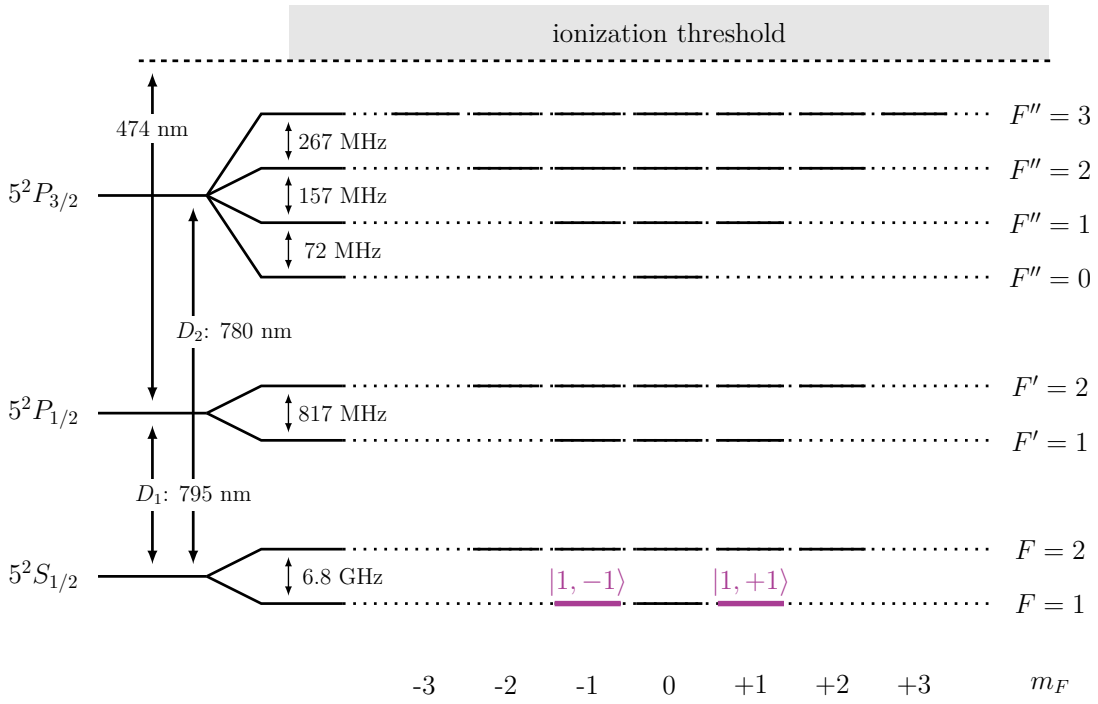


Figure 2.5.: Hyper-fine structure of ^{87}Rb as found in [64]. The ground state $5^2S_{1/2}$ is split into two ($F = 1$ and $F = 2$), the excited states $5^2P_{1/2}$ and $5^2P_{3/2}$ into two ($F' = 1, 2$) and four ($F'' = 0, 1, 2, 3$) hyper-fine levels, respectively. All hyper-fine states are further split into $2F + 1$ Zeeman substates labeled by m_F. Transition wavelengths between the ground and excited state are 795 nm and 780 nm for the D_2 ($5^2S_{1/2} \leftrightarrow 5^2P_{1/2}$) and D_1 ($5^2S_{1/2} \leftrightarrow 5^2P_{3/2}$) transition. The atomic qubit is encoded in the two substates $|F = 1, m_F = \pm 1\rangle$ of the atomic ground state $5^2S_{1/2}$, also denoted $|1, +1\rangle \equiv |\uparrow\rangle_z$ and $|1, +1\rangle \equiv |\downarrow\rangle_z$.

2.3. Trapping Single Atoms

After having introduced the theoretical steps to create atom-atom entanglement through a lambda scheme followed by entanglement swapping, we will turn our attention to the technical details on how to trap and control the single atoms.

2.3.1. Rubidium 87

In order to perform experiments generating heralded entanglement between distant particles, one needs a suitable quantum system that:

1. allows to distribute entanglement between the stationary qubits
2. has the possibility for a high fidelity readout
3. offers sufficiently long coherence times.

The candidate of choice for our experiment is a single atom of the alkali metal rubidium. Alkali metals are often used in quantum optics experiments as they hold the unique feature of having a single valence electron. Rubidium has two naturally occurring, long lived or stable isotopes, ^{85}Rb ($I = 5/2$; stable) and ^{87}Rb ($I = 3/2$; $\tau_{1/2} = 4.88 \times 10^{10}$ yr), of which the latter is chosen for our experiment as its hyper-fine structure (fig. 2.5) suit the experimental requirements, in particular the possibility for implementing a lambda level scheme. Generally, the physical constants about ^{87}Rb in this thesis will be identical to those found in [64], unless specified explicitly.

The qubit is encoded in the highlighted degenerate $m_F = \pm 1$ Zeeman substates of the $5^2S_{1/2}, F = 1$ ground state, with the lambda scheme (sec. 2.2.1) implemented between these states and the $5^2P_{3/2}, F'' = 0, m_F = 0$ excited state, entangling the Zeeman substate with the polarization of the emitted photon. An efficient readout is possible as the two Zeeman states can be addressed individually by appropriately polarized light (sec. 2.4.3). Unfortunately those states are particularly susceptible to magnetic fields, heavily affecting the coherence properties of the atomic state. Proposed solutions for this problem are a standing wave dipole trap [65] and a Raman state transfer of one superposition state the $5^2S_{1/2}, F = 2$ manifold [66].

Convention for Denoting Atomic States

Throughout the rest of this thesis we will use the following convention for the atomic states: $F = X$, $F' = X$, and $F'' = X$ will refer to states with hyperfine number X in the ground $5^2S_{1/2}$ level, and the excited $5^2P_{1/2}$ and $5^2P_{3/2}$ levels, respectively. The associated Zeeman substates (m_X) will be denoted $|(F =)X, (m_F =) = m_X\rangle$, $|F' = X, (m_F =)m_X\rangle$, and $|F'' = X, (m_F =)m_X\rangle$, where the information in might not be explicit. Unless stated otherwise, m_X is with respect to the quantization axis, that is defined by the microscope objective (figure 2.7).

2.3.2. Neutral Atom Trap

Each of the two stationary qubits, also called quantum memory, constitutes of a single ^{87}Rb atom, located in two separate laboratories in the basements of the faculty of physics in Schellingstraße 4 (lab 1) and faculty of economics in Schackstraße 4 (lab 2), which are 398 m apart and connected via 700 m of single mode optical fiber (fig. 2.6). The centerpiece of the experiment are the two traps located in the two laboratories. It should be noted that the setups for the traps in both laboratories differ slightly on a technical level but are conceptually identical. Thus, I will, unless specified explicitly, use the trap in lab 1 (fig. 2.7) as example to illustrate the working principles of the experiment. The trap in lab 2 is for example explained in [45, Appendix F].

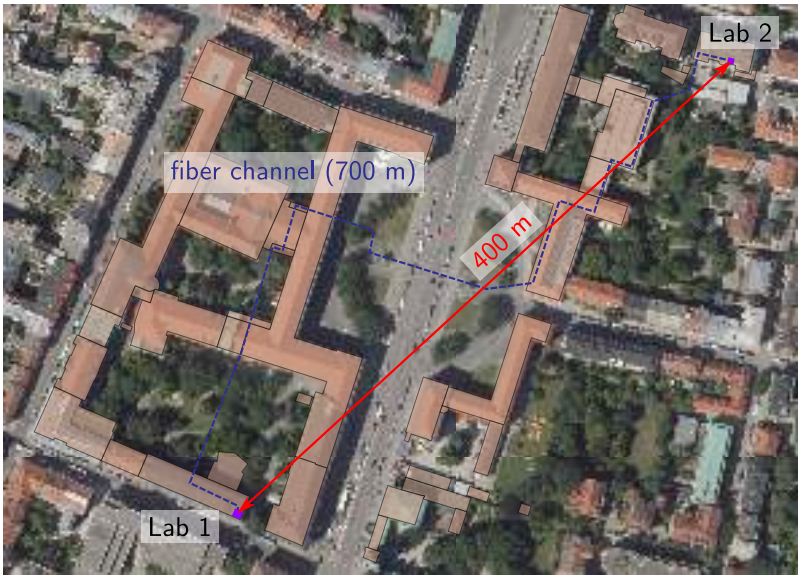


Figure 2.6.: Location of the two laboratories at the main campus of the Ludwig-Maximilians-Universität, Munich. Laboratory 1 (Lab 1) and Laboratory 2 (Lab 2) are located in the basement of the departments of the faculty of physics at Schellingstraße 4 and economics at Schackstraße 4. Optical fibers for communication and for guiding the fluorescence photons emitted by the atoms are connecting the two laboratories. Map data provided by the Bayerisches Landesamtfuer Digitalisierung, Breitband und Vermessung.

An optical dipole trap (ODT) is employed to store the atoms [67]. Before it is possible to transfer the atoms to the ODT, a cloud of cold atoms needs to be created inside a magneto-optical trap (MOT) [68] by means of Doppler cooling [69, 70].

2.3.2a. Vacuum, Magneto-Optical Trap, and Cooling

The atoms reside inside a vacuum chamber that can be evacuated to pressures below 10^{-9} mBar using an ion-getter-pump. For closing the locality loophole in the Bell test [18] the vacuum cell also houses two sets of channel electron multipliers (CEMs) that are currently not used. ^{87}Rb atoms are provided by a dispenser, which is kept off most of the time. The leakage of the vacuum cells is small, thus is sufficient to dissolve rubidium deposits that build up on the inner wall of the glass cell by using a UV LED. More details on the vacuum setup can be found in [71, 72].

MOTs are a well-established method for cooling and trapping neutral atoms. They use a combination of a magnetic quadrupole field created by coils in a anti-Helmholtz configuration and light to simultaneously cool and localized the atoms [68]. Three pairs of counter-propagating, circularly polarized beams, each containing a light from a cooling laser that is slightly red detuned from the closed D_2, $F = 2 \leftrightarrow F'' = 3$ transition and from a repump laser resonant to the D_2, $F = 1 \leftrightarrow F'' = 2$ transition,

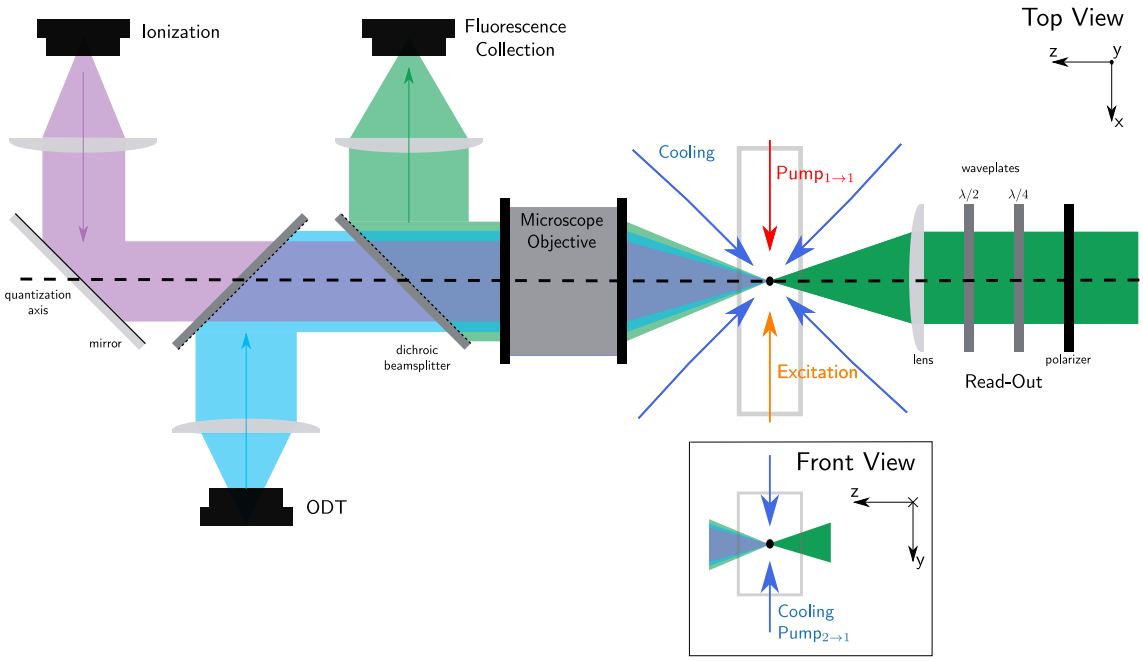


Figure 2.7.: Experimental setup. **ODT and Fluorescence Collection**: The the linearly polarized ODT beam (light blue) is focused by the microscope objective onto the position of the atomic cloud trapped by the MOT (dark blue arrows: cooling) inside the evacuated glass cell. Fluorescence collection from the atom (green, left) into a single-mode fiber is done through the same objective; its direction defines the quantization axis (z-axis). **Pumping and Excitation:** the pump$_{1\to1}$ (red) and excitation (orange) are π-polarized and orthogonal to the quantization axis (here $\pm x$). The pump$_{2\to1}$ is overlapped with the vertical cooling beams that propagate along $\pm y$. **Readout:** the readout laser (green, right) propagates in $-z$ direction. Its polarization χ_{ro} is set by a polarizer followed by a $\lambda/4$ and $\lambda/2$ waveplate. The ionization laser (purple) is overlapped with the ODT laser by a dichroic beamsplitter and focused with the objective on the atom.

are intersected inside the cell at the center of the coils where the magnetic fields vanish. The spatially varying induced Zeeman shift creates a force that pushes the atoms towards the trap center, slowing them down and thus cooling in the process. Typical parameters for the MOT in our setup are a cloud diameter $< 1\,\mathrm{mm}$, containing $> 10^4$ atoms [73] at a temperature well below the Doppler limit for ^{87}Rb of 146 µK. Actually, temperatures as low as 30 µK to 40 µK [74, 75] can be achieved through additional polarization gradient cooling [76].

2.3.2b. Optical Dipole Trap

Going from a trap with several thousand atoms to one containing only a single one is achieved by transferring one atom to an optical dipole trap (ODT) [67, 77] that uses the interaction between a laser field and the induced dipole moment of the atom. The laser field is far detuned from any transition such that excitation is negligible. It creates an AC Stark shift in the the atomic level structure depending on the intensity of the beam at the position of the atom. For red detuned lasers this creates an attractive three-dimensional potential dimple for the atom, centered at the focal point of the beam where the local intensity is maximal. With conventional lasers only small potential depths of few mK are achievable, making initial cooling of the atoms necessary. Assuming linearly polarized light the rotationally invariant potential is described by

$$U_{\mathrm{ODT}}(r,z) = \frac{\pi e^2 \Gamma}{2\omega_{D_1}} \left(\frac{2}{\Delta_{2,F}} + \frac{1}{\Delta_{1,F}} \right) I(r,z). \tag{2.25}$$

U_{ODT} depends on the decay rate of the excited state Γ, the transition frequency of the D_1-line ω_{D_1} and the detuning with respect to the D_1 and D_2 line, $\Delta_{1,F}$ and $\Delta_{2,F}$. The intensity $I(r,z)$ is assumed to be described by a Gaussian mode of the laser beam, taking the form

$$I(r,z) = I_0 \left(\frac{w_0}{w(z)} \right)^2 \exp\left(-\frac{2r^2}{w(z)^2} \right) \tag{2.26}$$

with the beam waist $w(z) = \sqrt{1 + (z/z_R)^2}$ at position z from the focus point of the laser ($z = 0$), the Rayleigh length $z_R = \pi w_0^2/\lambda$, the laser's wavelength λ, waist w_0, and intensity I_0 at the focus. We can use a harmonic approximation [78, 79] for the resulting potential, that has two different trap frequencies in transverse

$$\omega_T = \sqrt{\frac{4\, U_{\mathrm{ODT}}(0,0)}{m w_0^2}} \tag{2.27a}$$

and longitudinal

$$\omega_T = \sqrt{\frac{2\, U_{\mathrm{ODT}}(0,0)}{m z_R^2}} \tag{2.27b}$$

direction where m is the mass of ^{87}Rb.

Our setup employs a laser locked at a wavelength of 852 nm focused at the position of atomic cloud trapped by the MOT. The focal waist $w_0 \gtrsim 1.92\,\mu\mathrm{m}$ and Rayleigh length $z_R = 13.6\,\mu\mathrm{m}$ allow only for a single atom to be trapped due to collisional blockade effects [79, 80]. Choosing the optical power of the laser at

60 mW results in potential depth $U_{\mathrm{ODT}}(0,0) = k_B \cdot 3.2\,\mathrm{mK}$ and trap frequencies $\omega_T = 2\pi \cdot 92\,\mathrm{kHz}$ and $\omega_L = 2\pi \cdot 13\,\mathrm{kHz}$ [74].

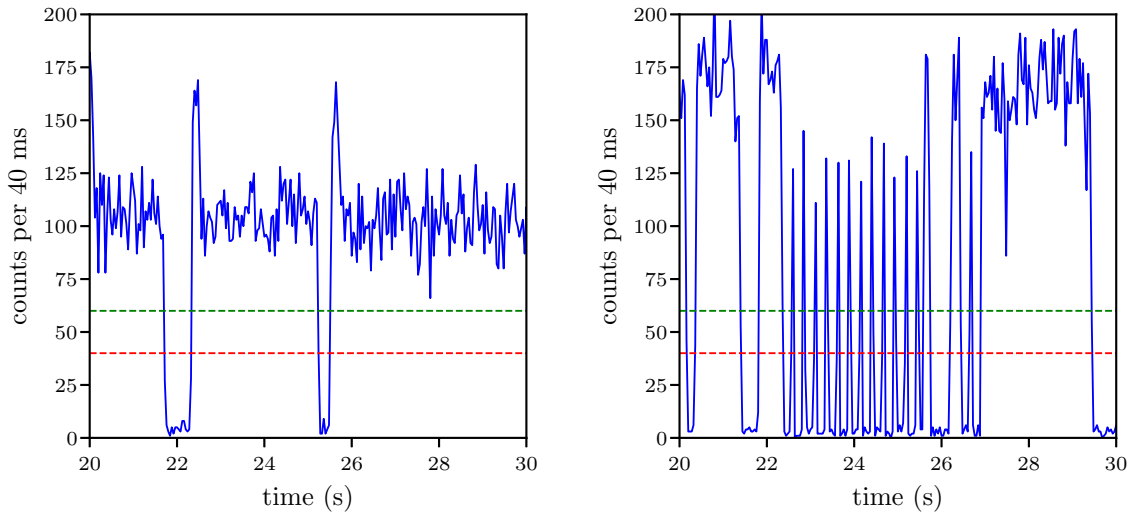
The ODT is created using a single high numerical aperture (NA) microscope (NA=0.5 [81]) that focuses the ODT beams to the desired waist. It also serves to collect fluorescence light scattered from the atom, allowing to decide whether or not an atom is present in the trap. Thus, a necessary requirement for the design has been the ability to focus 780 nm and 850 nm light at the point inside the glass cell. A dichroic beamsplitter makes sure only 780 nm light is guided to the single photon detection setup or the quantum frequency conversion setup that enables entanglement over even larger distances [82]. Despite the large NA, only a small fraction of the light, represented by the collection efficiency η_{col}, can be coupled into the optical fiber leading to the detection setup. The limiting factors are the the microscope objective itself, in particular the collection and aberration effects, on one hand. On the other, finite transmission losses in the optical fiber, the objective or other optical elements as well as the detector efficiencies reduce the achievable η_{col} [71].

The quantization axis of the systems is colinear with the direction of the fluorescence collection. Conventionally, this will be uses as the z-axis of the experiment's coordinate system as seen in figure 2.7.

2.3.2c. Loading Sequence and Fluorescence Collection

The loading or trapping of a single atom happens in a fully computer controlled and automated procedure. It starts by turning on the cooling beams and the current in the quadrupole coils , as well as the dipole trap laser. This creates a cloud of cold atoms out of which a single atom eventually is transferred to the ODT. The atom starts scattering the cooling light, increasing the photon detection rate rapidly. Once it surpasses a set threshold, the atom is assumed to be trapped and the quadrupole coils are switched off, discharging the other, untrapped atoms. The single atom in the ODT is now ready for experiments. Its presence in the trap is monitored and if the photon detection rate drops below a second, typically lower threshold, it is assumed to be lost and the procedure starts again.

This sequence was recently modified for the ongoing efforts to increase the spatial separation between the stationary nodes [82] using polarization-preserving quantum frequency conversion (QFC) [83] of the 780 nm photons to the telecom S-band at 1522 nm. Starting from just before the excitation try, all the light from the atom is guided towards the frequency converter, with no light reaching the silicon avalanche photodiode (APD) used for the fluorescence detection. The switching between those two channels is done using a microelectromechanical systems (MEMS) switch. The fluorescence count on the APDs subsequently drops to zero. After performing the readout, the MEMSs switch is switched back to verify the presence of the atom.



(a) Fluorescence trace without MEMS switch.

(b) Fluorescence trace with MEMS switch

Figure 2.8.: Fluorescence trace as the number of registered counts integrated for a fixed time integral (here 40 ms). Background counts for times when no atom is trapped are due to the cooling beams. The atom is considered trapped if the fluorescence counts rise above a value of 60 (green) and lost if fewer than 40 counts per 40 ms are registered. In (b), while the MEMS switch is guiding the photons towards the QFC, only background photons are recorded. The decision about the presence of the atom is only made after collecting fluorescence for 40 ms.

A small software delay ($20 - 40$ms) has to be added to allow enough counts to accumulate on the APDs for this procedure to work. The fluorescence traces for both sequences are shown in figure 2.8.

2.4. Entanglement Generation and Analysis

2.4.1. State Preparation

Before being able to start the atom-photon entanglement generation process, the atom needs to be initially prepared in $|1, 0\rangle$ state. From there the atom can be transferred into the excited $|F' = 1, m_F = 0\rangle$ state by applying a short laser pulse, from where atom-photon entanglement can be created via spontaneous decay (sec. 2.2.1).

2.4.1a. Optical Pumping

Preparing the atom in $|1, 0\rangle$ is done via a technique called optical pumping. After successfully loading and cooling, the atom has a high probability to be in the $F = 2$ manifold, as the dipole transition used for the cooling ($F = 2 \leftrightarrow F'' = 3$) is closed.

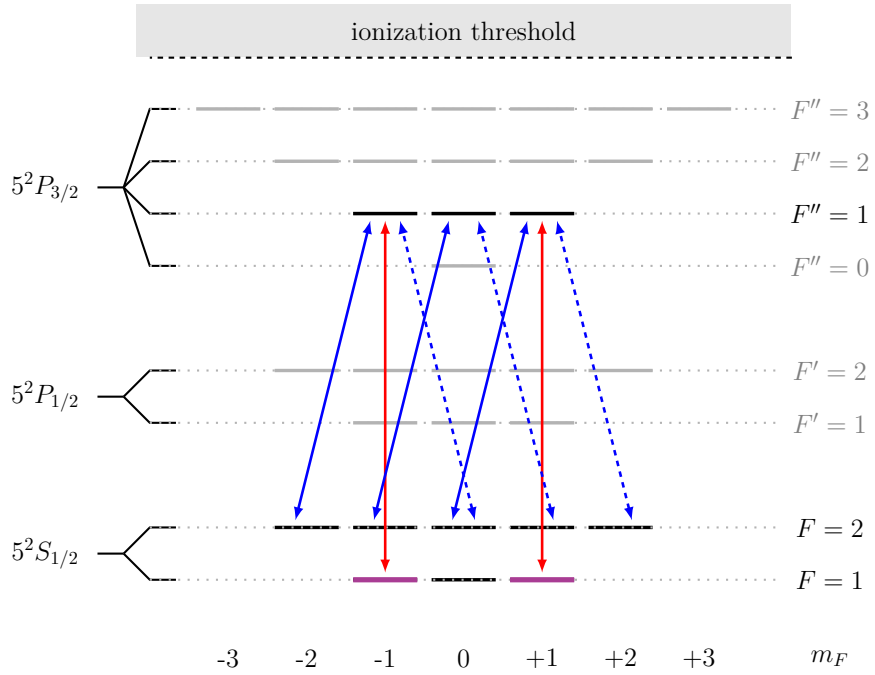


Figure 2.9.: Pumping scheme. The resonant $\text{pump}_{1\to1}$ ($\text{pump}_{2\to1}$) transition is drawn in red (blue). The two different polarizations of $\text{pump}_{2\to1}$ are indicated by dashed and undashed lines. Note that there is no transition that resonantly connects the dark state, $|1,0\rangle$ to any other level.

After several pumping-excitation cycles however, the atom will most likely be in either of the $F = 1$ substates as we will see briefly.

The pumping process is designed to excite all sub-levels of the ground state except for $|1,0\rangle$ which will be referred to as the dark state in the context of optical pumping. Therefore, two different laser fields are applied, referred to as $\text{pump}_{1\to1}$ and $\text{pump}_{2\to1}$. The former is resonant to the transition $F = 1 \leftrightarrow F'' = 1$, the latter to the transition $F = 2 \leftrightarrow F'' = 1$ (fig. 2.9). From the excited level, decays into all Zeeman substates of the ground manifold are possible as long as the dipole selection rules are respected. The decay amplitudes can be calculated from the linewidth of the excited level Γ_{D_1} and the corresponding Clebsch-Gordan coefficients. It should be noted, that instead of the excited $5^2P_{3/2}$ level, $5^2P_{1/2}$ could have been employed. However, spontaneous decays from $|F' = 1, \pm1\rangle$ to $|1,0\rangle$ only have a branching ratio of 1/6 compared to a branching ratio of 25/60 for decays from $|F'' = 1, \pm1\rangle$ to $|1,0\rangle$.

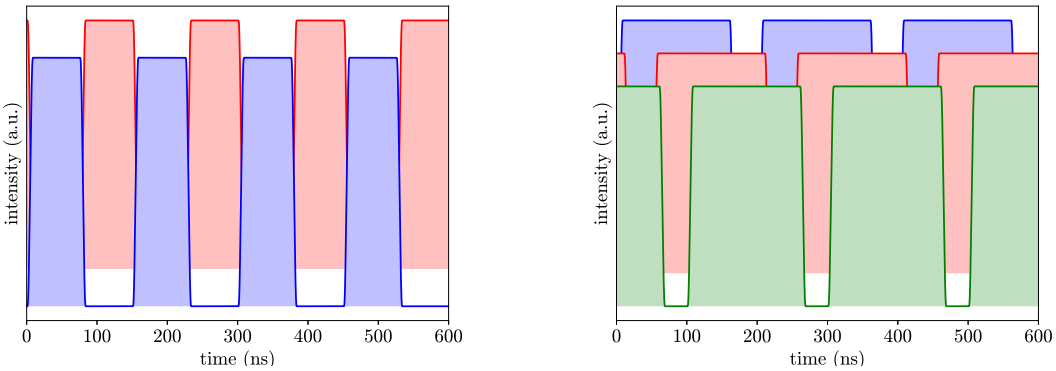
The $\text{pump}_{2\to1}$ laser aims at depopulating all $F = 2$ Zeeman sub-levels. Regardless of its polarization, two out of five sub-levels are unaffected by a single laser field. Therefore, the $\text{pump}_{2\to1}$ laser is circularly polarized and split into multiple beams,

that are guided at the atom from opposing directions. This is done differently in both laboratories:

lab 1: two independent, counter-propagating beams

lab 2: three orthogonal beams retro-reflected with piezo-mounted mirrors; $\text{pump}_{2\to1}$ is overlapped with the cooling light

Two coherent laser fields that have opposite circular polarization and identical field strength will recombine to form a linearly polarized light field. With π-polarized light it is not possible to excite the outermost Zeeman sub-levels ($m_F = \pm 2$), thus a pumping sequence is chosen, where the different polarizations are separated in time. In lab 1 this is realized by turning on and off the beam periodically before splitting it and than delaying one of the beams using additional optical fiber. The turning on and off is done by masking the amplitude of the RF-input of a AOM by a repeating logic signal of 6.6 MHz. The duty-cycle of the signal is 50%, thus requiring an additional fiber of length $75\,\text{ns} \cdot c/n_{\text{fiber}} = 15\,\text{m}$, assuming a refractive index $n_{\text{fiber}} = 1.5$ for the optical fiber (fig. 2.10a). Lab 2 uses a different time sequence, where the beam the RF-input modulation has a period of 200 ns, of which the AOM is switched on for 160 ns. Additional fibers of 10 m and 20 m are used to create a time difference of 50 ns and 100 ns (fig. 2.10b). Therefore, at any given time two of the three beams are switched on, with the combination alternating in time. The approach for atom 2 is more complex and it is hard to prove that it is advantageous compared to a case where the beams are continuously switched on. However, as experimentally the difference is in fact significant 2.11



(a) Lab 1 **(b)** Lab 2

Figure 2.10.: Illustration of the $\text{pump}_{2\to1}$ sequence in fig. 2.10a lab 1 and fig. 2.10b lab 2. Different colors indicate different beams. Note that in lab 1, the two different beams are counter-propagating whereas in lab 2, the beams are orthogonal and each beam is retro-reflected onto itself. Intensities are not to scale.

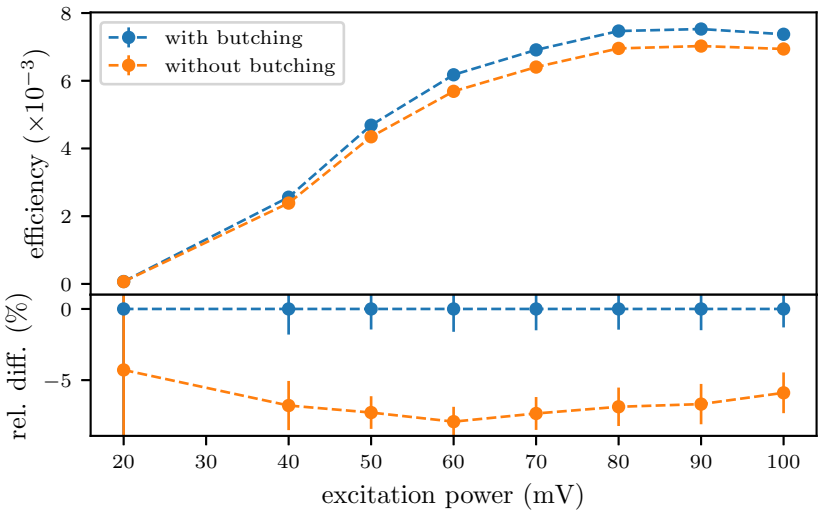


Figure 2.11.: Excitation-power-scan (error: 3σ) for optical pumping with (blue) and without (orange) sequenced $\text{pump}_{2\to1}$ laser in lab 2.

The $\text{pump}_{1\to1}$ laser is aimed to depopulate $|1, \pm1\rangle$. Due to photons necessarily carrying some non-vanishing angular momentum, a dipole transition $F = 1, m_F = 0 \leftrightarrow F'' = 1, m_F = 0$ is forbidden. Therefore, π-polarized light with a wavelength corresponding to the $F = 1 \leftrightarrow F'' = 1$ transition cannot resonantly interact with $|1, 0\rangle$ and only the Zeeman substates with non-zero m_F are depopulated, if $\text{pump}_{1\to1}$ is polarized parallel to the quantization axis. In lab 1, $\text{pump}_{1\to1}$ is counter-propagating to the excitation laser (fig. 2.7), while it is overlapped with the excitation in lab 2.

All involved levels and transitions required for the pumping are summarized in figure 2.9. In total, each pumping cycle takes around 2.3 µs and has a state preparation efficiency $\eta_{\text{prep}} \approx 80\%$. The pumping heats the atom, thus 350 µs of cooling are necessary after each 40th cycle, with the first pumping cycle after cooling being 2 µs longer [74]. In an earlier scheme [73] only one polarization for the $\text{pump}_{2\to1}$ was used and an effective mixing of the Zeeman substates in the $F = 2$ ground level was achieved by additionally switching on the cooling beams during pumping.

2.4.1b. Determining the State Preparation Efficiency

At this point, I want to address the definition and measurement of the state preparation efficiency. The current experimental configuration does not allow us to directly measure the state population of $|1, 0\rangle$. We therefore need multiple measurements to finally infer the desired quantity from the populations of the other Zeeman substates of $5^2S_{1/2}$. Population in any of the excited levels can be neglected due to the short lifetime already $\mathcal{O}(100\,\text{ns})$ after the pumping has ended. As we will see in section 2.4.3 we can perform a state selective ionization, of either the population

in $|F=2\rangle$ alone or the combined population of $|F=2\rangle$, $|1,0\rangle$ and an arbitrary superposition of $|1,+1\rangle$ and $|1,-1\rangle$. By choosing the superposition such that it corresponds to two orthogonal superposition states, e.g. $|\uparrow\rangle_x$ and $|\downarrow\rangle_x$, and realizing that

$$\Pr_{\text{ion}}(|F=2\rangle \cup |1,0\rangle \cup |\uparrow (\downarrow)\rangle_x) = 1 - \Pr(|\downarrow (\uparrow)\rangle_x) := \Pr_{\text{ion},\uparrow(\downarrow)}$$

it is possible to infer the state preparation efficiency $\eta_{\text{prep}} = \Pr(|1,0\rangle)$ as

$$\begin{aligned} \Pr(|1,0\rangle) &= 1 - \Big[\Pr(|F=2\rangle) + \Pr(|1,+1\rangle) + \Pr(|1,-1\rangle)\Big] \\ &= 1 - \Big[\Pr(|F=2\rangle) + (1 - \Pr_{\text{ion},\uparrow}) + (1 - \Pr_{\text{ion},\downarrow})\Big]. \end{aligned} \tag{2.28}$$

The equivalence follows from the fact that $|\uparrow\rangle_x$ and $|\downarrow\rangle_x$ can be obtained as linear combination of $|1,+1\rangle$ and $|1,-1\rangle$. The whole experimental sequence will be described in detail in chapter 4. After a successful ionization, the atom will be lost from the trap, making reloading necessary. Thus, measuring η_{prep} directly is very time intensive. For the purpose of optimizing the initial state preparation we will therefore rely on different methods.

2.4.1c. Effects Reducing the Initial State Preparation Efficiency

With the current pumping sequences, the initial state preparation yields $\eta_{\text{prep}} \approx 80\%$ after a total pumping time of $T_{\text{pump}} = 2.3\,\mu\text{s}$ ($4.3\,\mu\text{s}$ after cooling) in both labs. Increasing η_{prep} could help to increase the event rate as well as the the entanglement quality. This section will briefly discuss possible explanations for why η_{prep} is substantially below 100% after the pumping.

Generally, there are at least two possible scenarios. On one hand, it is possible that currently not enough time is allocated for the optical pumping and that increasing T_{pump} would be sufficient to reach higher η_{prep}. On the other hand, the evolution of the systems during the optical pumping might reach some steady-state distribution which corresponds to $\eta_{\text{prep}} \approx 80\%$. We performed a measurement of η_{prep} for different pumping times between $1.5\,\mu\text{s}$ and $39.5\,\mu\text{s}$ with the parameters of the pumping at the optimal values for $T_{\text{pump}} = 2.3\,\mu\text{s}$. It should be noted that the probabilities in figure 2.12 were measured directly after the cooling due to the nature of the measurement. This means that the atom is in $F = 2$ before the pumping, which however was not explicitly verified. It appears that the population of $|1,\pm1\rangle$, corresponding to the sum of $1 - \Pr_{\text{ion},\uparrow}$ and $1 - \Pr_{\text{ion},\downarrow}$, reaches a constant small, but non-vanishing value already after very short pumping times while the while the population of $F = 2$ only decreases slowly. It is possible that even shorter time steps could have resolved the dynamics of $|1,\pm1\rangle$ more accurately. During the experimental sequence, cooling only happens only after 40 pumping-excitation trials. Based on the measurement neither scenario can be fully excluded. We will now turn our attention to possible effects that might

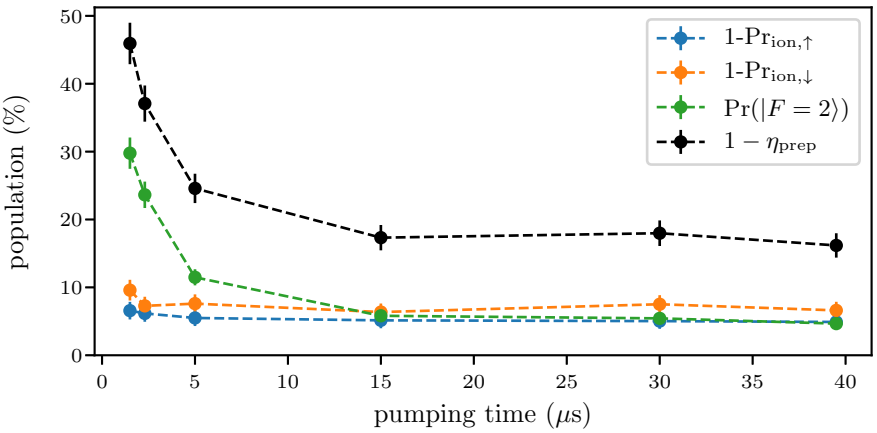


Figure 2.12.: Measured ionization probabilities $\mathrm{Pr}_{\mathrm{ion},\downarrow}$, $\mathrm{Pr}_{\mathrm{ion},\uparrow}$, and $\mathrm{Pr}(|F=2\rangle)$, as well as inferred initial state preparation efficiency η_{prep}. The pumping sequence was optimized for $T_{\mathrm{pump}} = 2.3\,\mu\mathrm{s}$. It should be noted however, that due to the nature of the measurement, the pumping is always preceded by cooling. In the experimental sequence, this is only the case for the first pumping sequence. Only points between 1.5 µs and 39.5 µs were recorded.

impair our ability to reach higher η_{prep}, and if applicable, how they could be reduced.

The first effect we considered is a possible misalignment of the $\mathrm{pump}_{1\to1}$ laser which can be decomposed into translations along and rotations around the three axes (fig. 2.7). Assuming that the laser itself is linearly polarized, it is easy to see that only rotations around the axis of propagation (x) result in effects that cannot be compensated by increasing the optical power in the beam. To only drive transitions from $|1,\pm1\rangle$, the polarization of $\mathrm{pump}_{1\to1}$ should be parallel to the optical axis ($\chi_{1\to1} = V$). Rotations by an angle α around x change this polarization to $\chi'_{1\to1} = \cos\alpha \cdot V + \sin\alpha \cdot H$, which will lead to effective $\sigma^{\pm}$ light fields, that would excite $|1,0\rangle$.

Second, we realize that transitions from any Zeeman substate to $|1,0\rangle$ are only possible by at least temporarily populating one of the excited states. While it is to be expected that the rate of change of the population of $|1,0\rangle$ would decrease over time as it becomes less likely to populate the $5^2P_{3/2}$ manifold, there are processes that could result in effectively leaving the excited states populated. As an example, stimulated Raman adiabatic passage (STIRAP) processes [54] could coherently transfer population between $|1,\pm1\rangle$ and $|F=2\rangle$ without ever populating the excited levels. STIRAP and similar techniques of coherent state transfer have wide applications in physics and chemistry [55, 84], however they are undesirable in our experimental setup.

As a third effect, off-resonant excitation, in particular of the target state $|1,0\rangle$,

could reduce η_{prep}. In ^{87}Rb the natural linewidth of the D_2 transition is $\Gamma_{D_2} = 2\pi \cdot 6.065(9)\,\text{MHz}$ as a result of the short lifetime of the excited $5^2P_{3/2}$ level [85]. On top, we have power-broadening [86]

$$\Gamma' = \Gamma_{D_2} \cdot \sqrt{1 + \frac{I_{1\to1}}{I_{\text{sat}}}}$$

by the $\text{pump}_{1\to1}$ laser. We can roughly estimate the intensity of the $\text{pump}_{1\to1}$ laser at the position of the atom by assuming the entire optical power $P_{1\to1} \approx 5.5\,\mu\text{W}$ is distributed over a disk of diameter $w_0 \approx 110\,\mu\text{m}$[4]. The resulting intensity $I_{1\to1} = 58\,\text{mW}\,\text{cm}^{-2}$ is well above the saturation intensity $I_{\text{sat}} \approx 3.58\,\text{mW}\,\text{cm}^{-2}$. Thus, off-resonant transition from $|1,0\rangle$ to $|F'' = (0,2)\rangle$, whose respective detunings are only 72 MHz and 156 MHz, are likely to contribute to the observed reduction of η_{prep}. This effect might be mitigated by modifying the intensity of $\text{pump}_{1\to1}$ as a function of time.

Another promising solution is to use the D_1 transition for depopulating $|1,\pm1\rangle$. As the energy difference of $h \cdot 816.656(30)\,\text{MHz}$ between $F' = 1$ and $F' = 2$ in $5^2P_{1/2}$ is significantly larger than the energy differences in $5^2P_{3/2}$ and than the natural linewidth $\Gamma_{D_1} = 2\pi \cdot 5.746(8)\,\text{MHz}$ of the transition, off-resonant excitation should be reduced. In addition, the possibility of coherent two-photon transitions between $F = 1$ and $F = 2$ is removed. However, we have to take into account that the branching ratio for decays from $|F' = 1, m_F = \pm1\rangle$ to $|1,0\rangle$ is only 1/6 compared 25/60 for decays from $|F'' = 1, m_F = \pm1\rangle$ to $|1,0\rangle$, thus reducing the effective speed of the pumping. Thus, allocating more time or the combining of pumping with 795 nm and 780 nm light for $\text{pump}_{1\to1}$ would probably be necessary to achieve an effective pumping scheme.

To the end of understanding how and if the different effects would influence the optical pumping, we designed numerical simulations based on the Lindblad equation 2.15. The simulations expand on the work in [73, 87], where the readout process was investigated. Major differences include a significant speed-up by taking into account the hermiticity of the density matrix and streamlining several software components, opening up the simulation to support different and more complex pulse shapes, and a way to interface the simulation with an external optimizer. The simulations where designed to model resonant excitations, off-resonant excitations to $F'' = 0, 2$, as well as polarization errors. For the simulation, we generated the $17 \times 17 = 289$ differential equations resulting from Lindblad master equations when considering all levels of $5^2S_{1/2}$ and the $F'' = 0, 1, 2$ levels of $5^2P_{3/2}$, of which we removed those equations that were simply the complex-conjugate of a second equation. We then employed Mathematica 11.0 to numerically integrated this system of complex-valued ordinary differential equations using a Runge-Kutta method.

[4] both those values are only crude approximations, based on measurements in lab 1

However, in preliminary investigations numerical errors numerical errors, that could have been neglected for shorter simulation times, accumulated and became more relevant, were found to be limiting. The simulated time period for the readout process ($\sim 200\,\mathrm{ns}$) is one order of magnitude smaller than the one for the optical pumping process ($2.3\,\mu\mathrm{s}$), thus increasing the working precision would have been necessary. Future simulations of the optical pumping process could benefit from solving real and imaginary part of each ODE separately and removing equations that can be shown to be trivial as done in [88]. Additionally, the speed of the simulations may be increased through the use of more efficient numerical solving methods and an implementation in a compiled language such as C++or Julia. Simulation that consider even larger systems of ODEs exist for atomic hydrogen [88, 89] and could potentially be adapted for ^{87}Rb. Despite not being able to fully describe the fully dynamics of the optical pumping process, the simulation helped to inform and motivate some of the pulse shapes we considered in the optimizations in chapter 4.

2.4.2. Entanglement Generation through Spontaneous Decay

After the pumping, the atom should be in the $|1,0\rangle$. In order to implement the lambda level scheme, a short laser pulse is used to drive a transition to $|F''=0,m_F=0\rangle$, which then decays with a short lifetime of $26.24\,\mathrm{ns}$ [64] back to the ground state, emitting a π-, σ^+-, or σ^--polarized photon with equal probability in the process. The emitted photons are collected along the quantization axis by a confocal microscope, thus filtering out linearly polarized photons. The remaining photons are guided towards a BSM, with the possibility of converting the photon to the telecom S-band [83].

The goal of the excitation process it to generate an entangled atom-photon pair described by equation 2.19, however not every excitation attempt results in the desired atom-photon state. Possible reasons excitation of $|1,\pm1\rangle$, both resonantly through residual circular polarization of the excitation pulse or off-resonantly to $|F''=1,m_F=\pm1\rangle$. $F''=0$ and $F''=1$ are only separated by $72.218(4)\,\mathrm{MHz}$ while the natural linewidth of the transition is only $\Gamma_0 = 6.065(9)\,\mathrm{MHz}$ [64]. The probability for off-resonant excitation is further increased as the short duration of the excitation pulse leads to significant spectral broadening ($\Delta\Gamma = 21.6\,\mathrm{MHz}$) of the laser [45]. Additionally, there is the chance for two photons to be emitted in one excitation process. These effects are strongly dependent on the properties of the excitation pulse, its polarization, temporal shape, and intensity. Their optimization requires careful attention to the photon collection and detection efficiencies, as well as detector dark counts. A thorough analysis of the excitation process has been done in [45, 53].

Choosing the correct excitation power P_{exc}, one has to consider two aspects: on

one hand, the excitation efficiency η_{exc}, the chance of correctly exciting $|1,0\rangle$ to $|F''=0,m_F=0\rangle$, rises as a function of P_{exc}. Thus, higher P_{exc} correspond to larger event rates, which becomes particularly important when performing measurements with atom-atom-entanglement where η_{exc} contributes quadratically. On the other hand, off-resonant excitation becomes more relevant as P_{exc} is increased, thus reducing the state fidelity. Taking all these different factors into consideration, the excitation pulse was chosen to be approximately Gaussian in shape with a full width at half maximum (FWHM) of 20.35 ns. Its intensity is set such that it yields around 80% of the maximum excitation efficiency and has to be adjusted on a regular basis [53]. The local photon detection probability, $\eta_1 = 6.7‰$ for lab 1 and $\eta_2 = 8.61‰$ for lab 2, is further decreased by additional time filtering, by fiber losses (700 m with losses of 4 dB/km) for photons from lab 2 and the by QFC [81]. Thus, a fast repetition rate of pumping and excitation is desirable for achieving reasonable event-rates, but at the same time it causes stronger heating of the atom. This increases not only the chance of loosing the atom from the trap but also the decoherences that result from stronger motion of the atom in the trap. A detailed discussion of the effects that lead to decoherence effects and ways to reduce them can be found in [45, 65].

2.4.3. Ionization-based Atomic State Readout

For the quantum-mechanical projection measurement needed to determine the state of the physical system in the experiment a state-selective ionization scheme is used. It is possible to transfer an arbitrary superposition of $|1,-1\rangle$ and $|1,+1\rangle$ to the excited $|F'=1,m_F=0\rangle$ from where it is ionized. The ion is subsequently lost from the trap which can be detected with high fidelity by a vanishing of the fluorescence, or for a faster readout by collecting the released charged particles with CEMs [45].

Excitation to $|F'=1,m_F=0\rangle$ is done with the readout-laser that is locked to the D_1 transition line and whose propagation direction is anti-parallel to the dipole laser. Its polarization χ_{ro} (eq. 2.29) dictates which superposition is transferred, defining the bright state $|\text{B}\rangle_{\text{ro}}$ (eq. 2.30a), which is orthogonal to the state that is not transferred, the dark state $|\text{D}\rangle_{\text{ro}}$ (eq. 2.30b). Any χ_{ro} can be chosen by appropriately rotating a quarter- and a half-wave plate in front of the glass cell. χ_{ro} can be decomposed into H and V polarizations

$$\chi_{\text{ro}} = \cos(\alpha)\cdot V + e^{-i\phi}\sin(\alpha)\cdot H \tag{2.29}$$

thus the dark and bright state take the form

$$|\mathrm{B}\rangle_{\mathrm{ro}} = \cos(\alpha)\frac{-1}{\sqrt{2}}(|1,-1\rangle - |1,+1\rangle) + \sin(\alpha)e^{+i\phi}\frac{i}{\sqrt{2}}(|1,-1\rangle + |1,+1\rangle), \text{ and} \tag{2.30a}$$

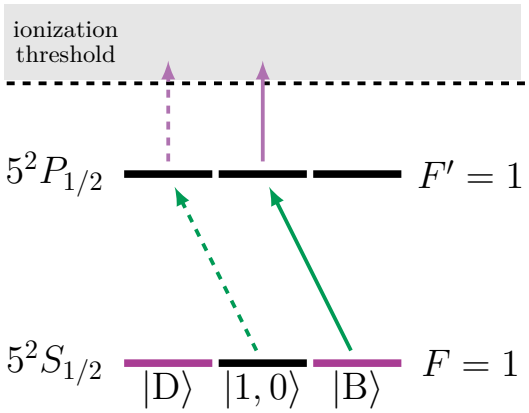
$$|\mathrm{D}\rangle_{\mathrm{ro}} = \sin(\alpha)\frac{+1}{\sqrt{2}}(|1,-1\rangle - |1,+1\rangle) + \cos(\alpha)e^{+i\phi}\frac{i}{\sqrt{2}}(|1,-1\rangle + |1,+1\rangle). \tag{2.30b}$$

The ionization threshold for the ground $F = 1$ and the excited $F' = 1$ state are 296.82 nm and 473.73 nm [90]. Therefore, in order to only ionize excited atoms, the ionization-laser has a wavelength below 473 nm and well above 300 nm. It is guided at the atom through the objective that is also used for fluorescence collection and ODT. Successful ionization immediately causes the atom to be lost from the trap, in turn reducing the fluorescence counts, making it possible to distinguish whether the atom was lost. Once again, additional waiting times might be necessary when using a MEMS switch as for the case of experiments with QFC. It should be noted that regardless of χ_{ro}, atoms that are still in $|1,0\rangle$ will always be ionized (fig. 2.13a), therefore the (theoretical) probability to find the atom in the trap after the readout-ionization cycle, $\mathrm{Pr_D} = \langle \mathrm{D}|\rho|\mathrm{D}\rangle$ for the given atomic state ρ, also called redetection probability, is

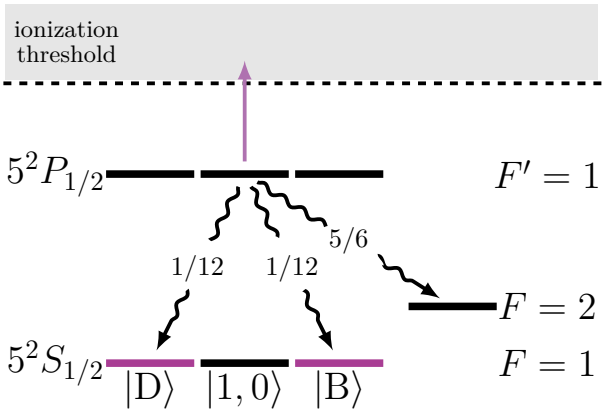
$$\mathrm{Pr_D} = \langle \mathrm{D}|\rho|\mathrm{D}\rangle = 1 - \mathrm{Pr_B} = 1 - \left(\langle \mathrm{B}|\rho|\mathrm{B}\rangle + \langle 1,0|\rho|1,0\rangle\right), \tag{2.31}$$

where we did not consider imperfections that could lead to the bright state remaining in the trap or the dark state being ionized.

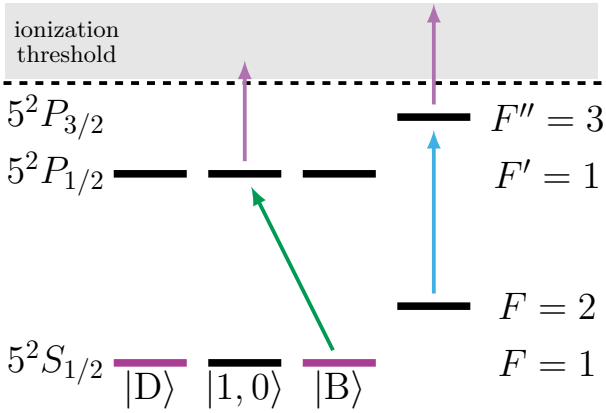
In the actual experimental setup, imperfections will play a role and will reduce the fidelity of this state readout scheme. The most prominent limiting factors are the life-time of the excited level as well as of resonant excitations of the dark state to $F' = 2$. The excited state has a short lifetime of 27.7 ns [64], and might decay spontaneously before being ionized (fig. 2.13b). This decay happens with a high probability of 5/6 to the $F = 2$ ground level. Applying laser resonant to the $F = 2 \leftrightarrow F'' = 3$, the cycling laser, opens up the possibility of a two-step ionization (fig. 2.13c). As from $F'' = 3$ only decays back to $F = 2$ are allowed, there arises no additional population, neither of the dark or the bright state via this channel as long as only resonant excitations are considered. Nevertheless, for the remaining times the state decays back to either the dark or bright state, with equal probability for each state. This evidently reduces the ionization probability of the bright state. Decays back to $|1,0\rangle$ are forbidden as no angular momentum would be exchanged. While the separation between the excited $F' = 1$ and $F' = 2$ manifolds of 816.7 MHz is large compared to the natural linewidth of only 5.746 MHz [64], the spectral broadening induced by the high-power ionization laser ($P_{\mathrm{ion}} = 200$ mW focused down to $w_0 = 1.0$ µm) leads to significant off-resonant excitations [71, 91].



(a) State-selective ionization: depending of the polarization χ_{ro} of the readout-laser (red) as specific superposition of the Zeeman states $|1,\pm1\rangle$, $|\mathrm{B}\rangle_{\mathrm{ro}}$, is excited to the $5^2P_{1/2}$ level. The orthogonal superposition, $|\mathrm{D}\rangle_{\mathrm{ro}}$, is not affected. The third state $|1,0\rangle$ is excited independent of the chosen polarization (red, dashed). From the excited level the atom can be ionized with the ionization laser (blue).



(b) Possible decay channels for the excited state. Due to the short life-time, the atom might decay before being ionized. Decays to the $F = 2$ manifold dominate (5/6), with smaller branching ratios for the dark and bright state (1/12). A decay to $|1,0\rangle$ is impossible as no angular momentum would be transferred.



(c) Second excitation after decay to $F = 2$. By applying the cycling laser (green), atoms in the $F = 2$ ground state can be transferred to the $F'' = 3$ level, from where ionization is possible. Decays from this level are only possible to $F = 2$.

Figure 2.13.: Ionization-based, Zeeman-state selective readout-scheme. It is both possible to record the drop in the fluorescence after the loss of the atom due to ionization or to detect the ionization fragments with CEMs for faster readout.

The performance of the readout-scheme, its ability to distinguish between dark and bright state, is characterized by the contrast

$$\mathrm{K}(\chi_{\mathrm{ro}}) = \mathrm{Pr}_{\mathrm{ionized}}\left(|\mathrm{B}\rangle_{\mathrm{ro}}\right) - \mathrm{Pr}_{\mathrm{ionized}}\left(|\mathrm{D}\rangle_{\mathrm{ro}}\right) \tag{2.32}$$

defined as the different between the probabilities of ionizing the bright and dark state. It depends on the characteristics of the read-out pulse. Simulation studies that tried to optimize the contrast under consideration of the experimental limitations [71, 87, 91] found an optimum for a pulse length of 140 ns and optical power of 1.24 µW. For these parameters, the simulations yielded ionization probabilities of $\mathrm{Pr}_{\mathrm{ionized}}\left(|\mathrm{B}\rangle_{\mathrm{ro}}\right) \simeq 0.98$ and $\mathrm{Pr}_{\mathrm{ionized}}\left(|\mathrm{D}\rangle_{\mathrm{ro}}\right) \simeq 0.04$, resulting in a contrast of $C_{\mathrm{ro}} \simeq 0.94$ [71]. These simulations were also examined and optimized with regard to their computational performance and expanded in order to examine if the flexibility to generate more complex pulse shapes could help to increase the contrast.

Determining the true contrast of a given set of parameter from the experiment is more challenging, as result of long measurement times, low event-rates and (mostly) statistical fluctuations.

All-together, the read-out itself takes less than 400 ns, which in combination with the fast charge detection in 570(3) ns at 0.965 fidelity with the CEMs [45, 71, 74] allowed to perform a Bell test that for the first time could close all loopholes for LHV theories [18]. For ionization detection via fluorescence, its trace needs to be integrated for several milliseconds. The total measurement time thus is larger than 30 ms and has a fidelity of 0.97.

2.5. Laser Pulse Sequence Generation

In previous sections we have introduced the different ingredients necessary to perform experiments with single atoms. It became clear that a high degree of control over the involved laser fields, its spatial, temporal and spectral shape, is one of the key components.

The spatial degree of freedom is, to large parts, fixed by the alignment and specialized design of optical components. The entire experiment has been optimized over years to yield high performance, most recently by the employment of a new microscope objective that was able to improve the local photon detection efficiencies by a factor of roughly 2.5 and 3.5 in lab 1 and and lab 2, respectively, which translates to an 6.5 fold increase in atom-atom-entanglement efficiency [81]. Overall, the spatial alignment is rather stable with respect to time, however, if necessary, re-alignment can proof to be tedious.

On the contrary, the spectral and temporal degrees of freedom are more volatile.

The lasers in the setup are all operated in continuous wave and single frequency mode. Their respective frequency is locked to either a reference cavity or atomic transition using Doppler-free saturation spectroscopy [92]. Those laser beams are distributed into several lines and frequency shifted to match their purpose in the experiment (fig. A.1). This is done by propagating the light through acousto-optic modulators (AOMs) whose efficiency can be varied to generate temporal pulse shapes. This section will briefly introduce the working principles behind said AOMs.

AOMs, also called acousto-optic deflector (AOD) or Bragg cell, make use of a specific type of photoelasticity in an optically transparent material, such a telirium oxide of germanium crystals. The required strain is induced through exciting an acoustic wave inside the medium, leading to a spatial and temporal modulation of its refractive index. This corresponds to a diffraction grating moving at the speed of sound c_s through the material. Similar to the physical Bragg grating etched into a waveguide, the condition for constructive interference of light at this grating can be expressed in terms of the wavelength $\Lambda_s = c_s/F_s$ of the acoustic wave

$$2\Lambda_s \sin(\theta_{B,m}) = m\lambda \tag{2.33}$$

where λ is the wavelength of the light and $\theta_{B,m}$ is the Bragg angle the integer order $m = 0, \pm 1, ...,$ measured perpendicular to the propagation direction of the sound wave. On a microscopic level, this can be understood as a photon-phonon-scattering process. Energy-and-momentum conservation dictate that the frequency will be shifted by an amount proportional to the frequency F_s of the sound wave

$$\nu \to \nu_m = \nu + m \cdot F_s \tag{2.34}$$

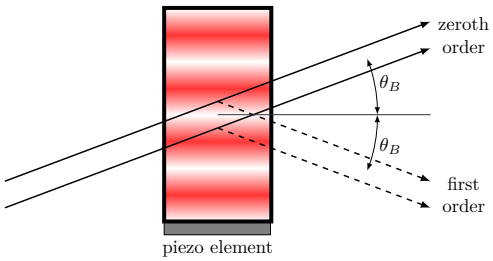
constituting a Doppler shift.

Normally, only the 0th and 1st order are considered for practical applications as they carry the largest amount of optical power. How the optical power is distributed over the different orders depends on the intensity of the sound wave. The efficiency given as the ratio between input and output intensity, I_{in} and I_{out} of the 1st order can be expressed as

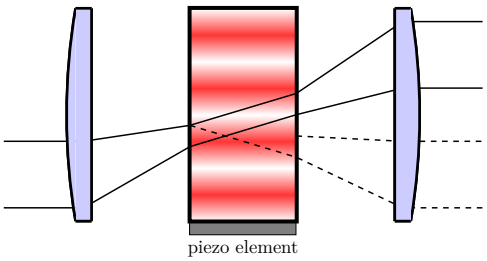
$$\eta_{\text{AOM}} = \frac{I_{\text{out}}}{I_{\text{in}}} = \sin^2\left(\frac{\pi\, L}{\lambda \sin\theta_B}\sqrt{\frac{M\, I_s}{2}}\right), \tag{2.35}$$

where M is a constant set by the material, L is the interaction length scale and I_s is the intensity of the sound wave.

The sound wave is generated by a vibrating piezo-electric transducer, that is driven by an oscillating electric signal, usually in the radio-frequency (RF) range. By changing its amplitude, the optical power in the 1st order can be varied between



(a) Bragg condition for an AOM. Here we assume a infinitely extended AOM as well as light field defined by a single wavevector $\vec{k}$. The condition for constructive interference is given in equation 2.33



(b) AOM pad using two identical plano-convex lenses. The AOM sits at the common focal point of both lenses. The spatial mode of the beam is not affected by the AOM. For a double pass configuration, the first order beam is deflected back with polarization rotated by 90°, e.g. by passing through a QWP twice.

Figure 2.14.: Operation principle of an AOM

0% and 80% of the input optical power for appropriate frequencies.

Realistically, the AOM itself, as well as the light beam and the acoustic wave are finite in size. Therefore, the Bragg condition (2.33) cannot be applied strictly and it is possible for multiple wave-vectors, such as in a focused Gaussian beam, to interfere constructively. It should be also noted that the response of the AOM to changes in the RF-input is not instantaneous. First, there will be a constant time delay as the density wave needs a finite time to reach the location of the beam, which can be used to fine-tune the timings of the generated laser pulses. Second, the wave front takes a finite time to cross the beam diameter, approximated by $\tau = 2w_0/c_s$. This results in a finite rise time $\tau_r = 0.85\tau$ of the AOM, defined as the time it takes for increasing the first order efficiency from 10% to 90% of its maximum. This effectively leads to a low-pass filtering of the input signal. The situation is illustrated in in figure 2.14b, where a AOM sits in the middle of two-plano convex lenses, whose focal points are overlapped inside the AOM. This configuration can be used in single- or double-pass configuration, corresponding to a frequency shift of F or $2 \cdot F$, with little disturbance of the beam's spatial mode after the AOM pad. In the latter, the outgoing first order beam is deflected back with its polarization rotated by 90°, which mitigates alignment problems [93].

So far, in the presented experiment AOMs were used in combination with fixed RF drivers, that could only generate square pulses as input for the AOM, whose amplitude and frequency were the only dynamic degree of freedom. This thesis explores the possibility to use a computer programmable arbitrary waveform generator for producing more complex pulse forms that might improve the overall performance of the experiment. To this end, machine learning techniques were employed to op-

timize these waveforms, both in an open-loop and closed-loop configuration, using either simulations or the experiment itself as feedback.

3. Optimization Algorithms

Up to this point, Machine Learning (ML) algorithms, apart from a simple gradient descent algorithm for the fiber polarization control, have not been used for the optimization of the trap setup introduced in the previous chapter. For this thesis, we have designed ML optimizers based on the publicly available *Machine-Learner Online Optimization Package* (M-LOOP) which has been specifically developed for the use in scientific applications [34, 94].

This chapter will give an introduction into two different ML algorithms used within M-LOOP, Gaussian process regression (GPR) and differential evolution (DE). Both algorithms fall into the category of supervised learning algorithms. Unlike techniques like gradient descent, which often do not find the global optimum in the presence of local minima, they are capable of global optimization. Finally, we will see how M-LOOP can be integrated into an experimental setup.

3.1. Gaussian Process Regression

The goal of this section is to introduce the reader briefly into the concept of Gaussian processes and their relevance in machine learning, concentrating specifically on the aspects relevant for their implementation within M-LOOP. For a more thorough discussion, we highly recommend the publicly available book "Gaussian Processes for Machine Learning" by Rasmussen and Williams [95] from which lots of the inspiration for this section is drawn.

GPR falls into the class of supervised learning algorithms. The purpose of supervised learning is to find (or *learn*) a mapping between input parameters ($\vec{x}$) and output ($f(\vec{x})$) based on a limited set of labelled input-output pairs, the training data. Outputs can either take the form of continuous or discrete labels, for regression or classification tasks, respectively. Supervised learning algorithms should eventually be able to accurately predict the label of some input that is not part of the training data set [95].

We will limit our discussion to regression tasks, as our output refers to the performance of the experiment, that is the average over many realization and thus in general continuous.

The process of finding the best approximate map or model $\tilde{f}$ for f is called fitting. Apart from requiring consistency with the training data, fitting often involves making some additional assumption about the form of f. Those assumptions restrict the class of functions that can be considered for f and define a parametric model $\tilde{f}(\vec{x}, \vec{\theta})$

for the output. Fitting in this context means finding $\vec{\theta}$ such that the similarity similarity between $\tilde{f}(\vec{x}, \vec{\theta})$ and the observed data is maximized. Although this approach can be useful to encode prior knowledge, it can result in either of two problems. On one hand, if the class of functions is chosen too wide, $\tilde{f}$ might be chosen too specific to the training data and performs poorly on unseen data. On the other hand, if it is chosen too narrow, $\tilde{f}$ might not be able to catch all the defining features of the underlying system.

3.1.1. Gaussian Process Regression

In contrast, Gaussian processes (GP) use observed similarity in the input space to infer similarity in the output space. Knowledge about how close two points $\vec{x}$ and $\vec{x'}$ are is used to make predictions about the relation between $f(\vec{x})$ and $f(\vec{x'})$ [96].

In essence, GPs generalize the concept of (multivariate) *Gaussian* distributions and stochastic *processes*, describing functions rather than scalar or vector variables. Modelling is done by assigning probabilities to functions according to their likelihood to describe the presented data [95]. This enables the GP

1. to be applied to a large variety of possible problems, even if those that have complex nonlinear or multi-modal underlying functions,
2. to make predictions about the regressed functions itself as well as the uncertainty of the regression, and
3. to extrapolate its prediction into unknown regions of the parameter space [97].

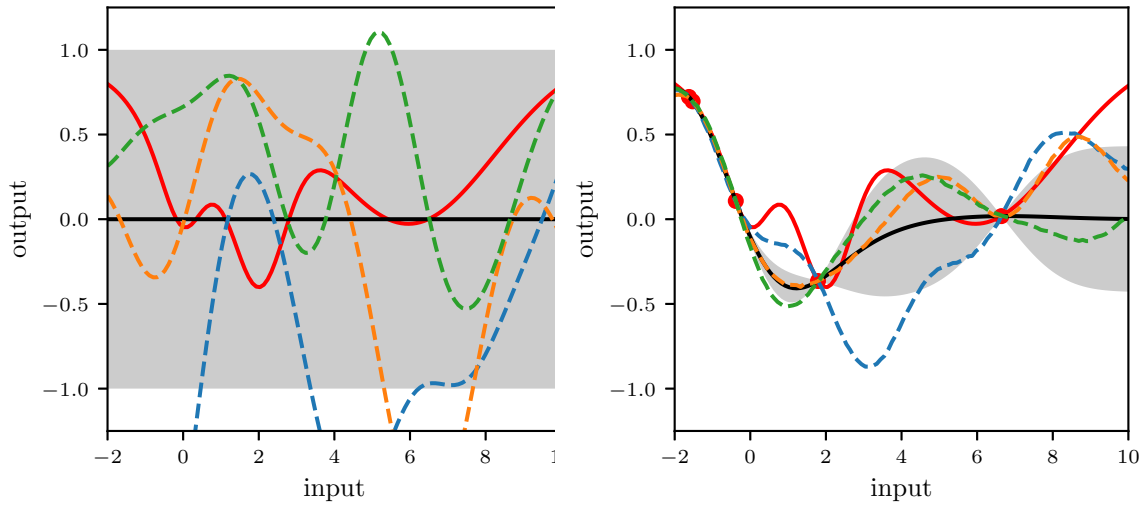
In the following we will examine the mathematical framework necessary to understand a GP as well as possible implementations of it, particularly those used in M-LOOP.

A GP describes a (infinite) collection of random variables (functions). Of this collection, any finite subset has a joint Gaussian distribution. A GP is fully specified by its mean $m(\vec{x})$ and covariance $k(\vec{x}, \vec{x'})$:

$$m(\vec{x}) = \mathbb{E}[f(\vec{x})] \tag{3.1a}$$

$$k(\vec{x}, \vec{x'}) = \mathbb{E}[(f(\vec{x}) - m(\vec{x})(f(\vec{x'}) - m(\vec{x'})] \tag{3.1b}$$

$k(\vec{x}, \vec{x'})$ is often referred to as the kernel of the GP. While in practise it might be hard or even impossible to know the exact form of m and k, any finite number of samples from the GP can be used to approximate them. M-LOOP does exactly this: by executing an experiment a limited number of times with different parameters, it develops a statistical model capable of predicting how the experiment performs for arbitrary parameters.



(a) Prior distribution **(b)** Posterior distribution

Figure 3.1.: Illustration of a GP learning a multi-modal function ($f(x) = 2 - e^{-(x-2)^2} - 0.1e^{-(x-6)^2} - (x^2+1)^{-1}$, red line) by drawing (noise-free) samples (red points). The prior (a) and posterior (b) prediction are shown by black lines, with the shaded area indicating the 1σ error bands, also called uncertainty. Dashed line represent random functions drawn from the prior (a) and posterior (b) kernel. The point-wise average and standard deviation of all functions that could be drawn from either kernel are precisely the mean and and uncertainty of the GP. Note how at the observation points the uncertainty shrinks to zero and the sampled functions cross.

Assuming a (training) set of N input-output pairs $\{(\vec{x}_i, f_i \equiv f(\vec{x}_i))\}_{i=1}^N := (\vec{X}, \vec{f})$, we can form a normal probability distribution function (PDF) $\mathcal{P}(\vec{f}|\vec{X})$. We will use the shorthand $\mathcal{P}(\vec{f}|\vec{X}) \sim \mathcal{N}\left(\vec{m}, K(\vec{X},\vec{X})\right)$ to denote

$$\mathcal{P}(\vec{f}|\vec{X}) = \frac{1}{\sqrt{(2\pi)^N \det\left[K(\vec{X},\vec{X})\right]}} \exp\left[\frac{-1}{2}(\vec{f}-\vec{m})^T K(\vec{X},\vec{X})^{-1}(\vec{f}-\vec{m})\right], \quad (3.2)$$

where $\vec{m}$ and $K(\vec{X},\vec{X})$ have components $m_i \equiv m(\vec{x}_i)$ and $K(\vec{X},\vec{X})_{ij} \equiv k(\vec{x}_i,\vec{x}_j)$, respectively. $K(\vec{X},\vec{X})$ is the Gram matrix of the covariance function k.

Selecting specific m and k is done based on the problem at hand. Often m is chosen to be constant vector. There exist different options for choosing the covariance (or kernel) function $k(\vec{x},\vec{x'})$, all of which at least have to be positive semi-definite [95, 97]. M-LOOP uses the squared exponential or radial bases function (RBF) kernel, which will be discussed in section 3.1.2.

The main purpose why we are interested in GPs is evidently their ability use observations $(\vec{X}, \vec{f})$ to make predictions about unseen inputs $\vec{X}_*$ and their corresponding outputs $\vec{f}_*$, also referred to as test set $(\vec{X}_*, \vec{f}_*)$. Together they form the joint prior PDF

$$\mathcal{P}(\vec{f}, \vec{f}_*|\vec{X}, \vec{X}_*) \sim \mathcal{N}\left(\begin{bmatrix}\vec{m} \\ \vec{m}_*\end{bmatrix}, \begin{bmatrix}K(\vec{X}, \vec{X}) & K(\vec{X}, \vec{X}_*) \\ K(\vec{X}_*, \vec{X}) & K(\vec{X}_*, \vec{X}_*)\end{bmatrix}\right) \tag{3.3}$$

where we assume a noise-free environment for the moment. This means that predicting the PDF for $\vec{f}_*$ corresponds to restricting the joint prior such that it only contains those function that exactly agree with the training data. Mathematically, this corresponds to conditioning $\mathcal{P}(\vec{f}, \vec{f}_*|\vec{X}, \vec{X}_*)$ on the observed output $\vec{f}$ (see [98, chapter 3] for details), again forming a multivariate normal PDF

$$\mathcal{P}(\vec{f}_*|\vec{f}, \vec{X}, \vec{X}_*) \sim \mathcal{N}\left(\vec{\mu_*}, \Sigma_*\right), \tag{3.4}$$

where the mean and covariance are given by

$$\vec{\mu_*} = \vec{m}_* + K(\vec{X}_*, \vec{X})K(\vec{X}, \vec{X})^{-1}(\vec{f} - \vec{m}) \tag{3.5a}$$

$$\text{and } \Sigma_* = K(\vec{X}_*, \vec{X}_*) - K(\vec{X}, \vec{X}_*)K(\vec{X}, \vec{X})^{-1}K(\vec{X}_*, \vec{X}), \tag{3.5b}$$

respectively. Figure 3.1 illustrates how adding more observations restrict the

For a more realistic treatment we will assume that we only have access to training output $\vec{y}$ which is our true function $\vec{f}$ subjected to noise that is independent and identically drawn from a normal distribution with standard deviation σ_n^2 (white noise). While this does not affect our mean function, it changes the covariance matrix of the training set

$$K(\vec{X}, \vec{X}) \to K(\vec{X}, \vec{X}) + \sigma_n^2 I, \tag{3.6}$$

and thus lead to the new joint probability distribution

$$\mathcal{P}(\vec{y}, \vec{f}_*|\vec{X}, \vec{X}_* \sim \mathcal{N}\left(\begin{bmatrix}\vec{m} \\ \vec{m}_*\end{bmatrix}, \begin{bmatrix}K(\vec{X}, \vec{X}) + \sigma_n^2 I & K(\vec{X}, \vec{X}_*) \\ K(\vec{X}_*, \vec{X}) & K(\vec{X}_*, \vec{X}_*)\end{bmatrix}\right). \tag{3.7}$$

Once again conditioning (3.7) yields the key predictive expressions for the GP

$$\mathcal{P}(\vec{f}_*|\vec{X}, \vec{y}, \vec{X}_*) \sim \mathcal{N}\left(\mathbb{E}[\vec{f}_*|, \vec{X}, \vec{y}, \vec{X}_*], \text{cov}(\vec{f}_*)\right), \tag{3.8}$$

where

$$\mathbb{E}[\vec{f}_*|, \vec{X}, \vec{y}, \vec{X}_*] = \vec{m}_* + K(\vec{X}_*, \vec{X}) \left[K(\vec{X}, \vec{X}) + \sigma_n^2 I\right]^{-1} (\vec{y} - \vec{m}) \tag{3.9a}$$

and

$$\text{cov}(\vec{f}_*) = K(\vec{X}_*, \vec{X}_*) - K(\vec{X}, \vec{X}_*) \left[K(\vec{X}, \vec{X}) + \sigma_n^2 I\right]^{-1} K(\vec{X}_*, \vec{X}). \tag{3.9b}$$

M-LOOP only considers white noise as presented here, but generally, GPs can also treat other noise models. that transform the covariance matrix (3.6) in non-trivial ways. Some of those are examined in [95, chapter 9].

3.1.2. Kernel Function and Hyperparameters

A *kernel* (also kernel function or covariance function) is a general function $k(\vec{x}, \vec{x'})$ mapping two inputs $\vec{x}, \vec{x'} \in \mathbb{X}$, where $\mathbb{X}$ is the set of all possible inputs, into $\mathbb{R}$. Without loss of generality we can assume that $\mathbb{X}$ is a region in $\mathbb{R}^D$ such that $\vec{x}$ and $\vec{x'}$ are D-dimensional real vectors. Not any k can serve a valid covariance function for our GP. First, k should be symmetric such that $k(\vec{x}, \vec{x'}) = k(\vec{x'}, \vec{x})$. Second, the kernel needs to be semi-positive, meaning that the inequality

$$\int k(\vec{x}, \vec{x'}) f(\vec{x}) f(\vec{x'}) d\mu(\vec{x'}) d\mu(\vec{x}) \leq 0 \tag{3.10}$$

holds for all $f \in L_2(\mathbb{X}, \mu)$, the set of square-integrable functions. This definition is equivalent to saying that all eigenvalues for a Gram matrix resulting for k for any finite subset $\{\vec{x_i} \in \mathbb{X}\}$ are non-negative [95]. Please note, that starting here, we will use the terms 'covariance function' and 'kernel' interchangeably.

Covariance functions fall into several sub-classes. If k is invariant under arbitrary translations in input space, it is called stationary. A popular stationary kernel is the squared exponential or radial basis function (RBF) kernel defined by

$$k_{\text{RBF}}^{(\text{iso})}(\vec{x}, \vec{x'}) = \exp\left(-\frac{\|\vec{x} - \vec{x'}\|^2}{2\ell^2}\right), \tag{3.11}$$

where the free parameter $\ell > 0$ defines the characteristic *length scale*. Equation 3.11 represents an isotropic kernel, as it only depends on the Euclidean distance between $\vec{x}$ and $\vec{x'}$. Informally speaking, ℓ tells us how far we can move in input space without affecting the output significantly [95].

Generally, when dealing with multi-dimensional input we would want to assign an individual length-scale to each dimension. This is also called automatic relevance determination (ARD) as the the length scale parameters $\ell_1, \ell_2, ..., \ell_D$ govern the relative importance of each dimension [99], under the assumption that all parameters

have the same or at least similar range. One possible anisotropic generalization of 3.11 is

$$k_{\mathrm{RBF}}(\vec{x},\vec{x'}) = \exp\left(-\frac{1}{2}\sum_{d=1}^{D}\frac{(x_d - x'_d)^2}{\ell_d^2}\right), \tag{3.12}$$

which is probably the most popular kernel used in GP applications [99]. From a practical point of view, they have relatively few parameters with clear interpretation. In addition, they are part of the group of universal kernels, that can be shown to be capable of approximating any arbitrary continuous function [100].The RBF kernel is particularly smooth as it indefinitely differentiable.

Taking noise into account can either be done explicitly as in (3.6) or by having an implicit, local kernel function representing the noise. A kernel describing white noise (noise level σ_n^2) can be achieved by

$$k_{\mathrm{WN}}(\vec{x},\vec{x'}) = \sigma_n^2 \cdot \delta^{(D)}(\vec{x} - \vec{x'}). \tag{3.13}$$

It is easy to see, that the sum or product of two kernels again form a kernel [95, chapter 4.2]. For example, the anisotropic RBF kernel can be obtained by multiplying one-dimensional RBF kernels. Moreover, we can combine different kernels to account for different features in our data. The kernel used in M-LOOP in general is the sum of a (an)isotropic RBF kernel and a white-noise kernel

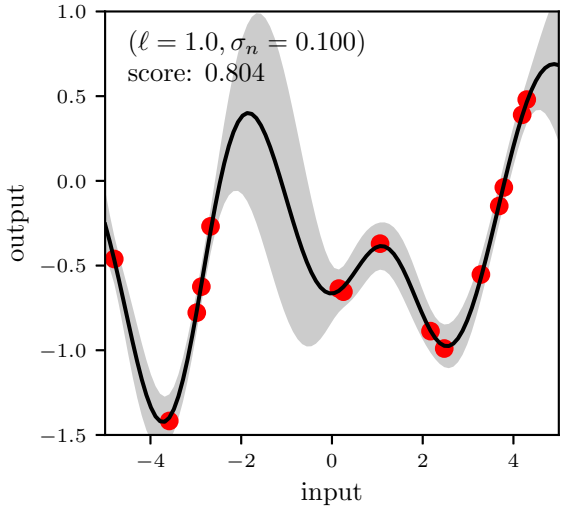
$$k_{\mathrm{M-LOOP}}(\vec{x},\vec{x'}) = k_{\mathrm{RBF}}(\vec{x},\vec{x'}) + \lambda \cdot k_{\mathrm{WN}}(\vec{x},\vec{x'}), \tag{3.14}$$

where $\lambda \in \{0, 1\}$ depending on whether we want to incorporate noise into our model.

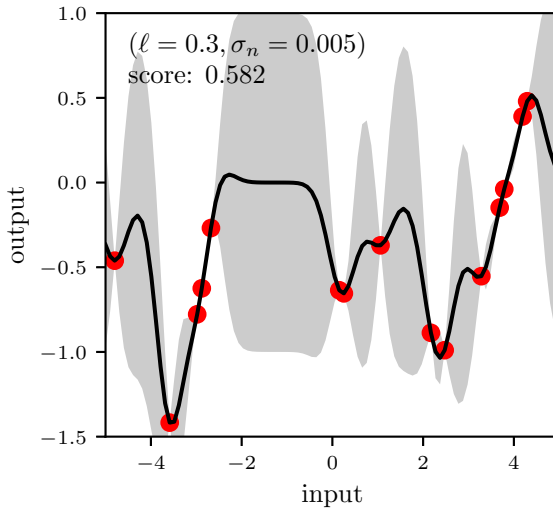
The collection of all parameters describing a GP's kernel are called the GP's *hyperparameters* ($\vec{\theta}$). Their values greatly affect how well the GP is able to make predictions for unseen data. This is illustrated in figure 3.2 where we generated data from f that was randomly drawn from a GP's prior. The GP had a kernel as described by (3.14) with length-scale $\ell = 1$ and noise variance $\sigma_n = 0.1$. An estimate for the quality the prediction $f_{\mathrm{pred}}(\vec{x})$ is the so called score $(1 - u/v)$ where

$$u = \sum \left(f(\vec{x}) - f_{\mathrm{pred}}(\vec{x})\right)^2 \quad \text{and} \quad v = \sum \left(f(\vec{x}) - \mathbb{E}[f(\vec{x})]\right)^2. \tag{3.15}$$

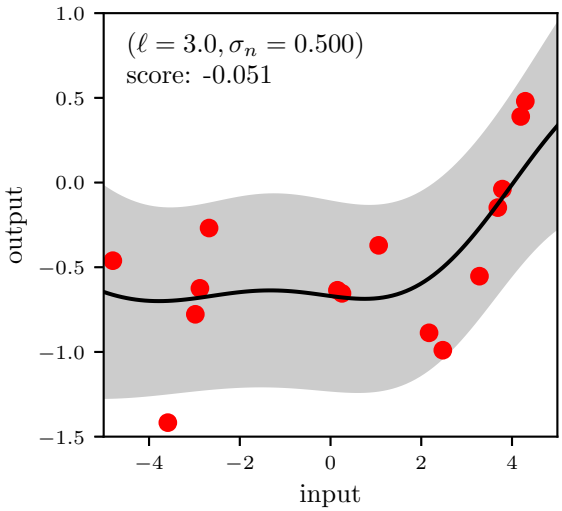
Ideally, $\vec{x}$ are chosen that are not part of the training set $\{\vec{x_i}\}$. A score close to the maximum value of 1 indicates that f_{pred} is a good approximation of f. Figure 3.2a uses the true hyperparameters and has the highest score. In figure 3.2b the length-scale was chosen too short. Fluctuations are fully explained by the 'signal' component, while the noise variance is small. We notice how the uncertainty grows rapidly as we move away from know data points. In contrast, the length scale in figure 3.2c is too large. A high noise level (σ_n^2) is necessary to incorporate the training data and the score for unseen data is very low.



(a) Correct hyperparameters.



(b) Lenght scale too small.



(c) Length scale too large.

Figure 3.2.: Effect of varying the hyperparameters. The data (red points) sampled from a function that is drawn from a GP prior with a combined RBF and white noise kernel as in (3.14). The true hyperparameters are $\ell = 1$ and $\sigma_n = 0.1$ as in (a). The predicted output (black line) and 1σ uncertainty interval (gray area) for a GP with these hyperparameters that is fitted to the data are superimposed. (b) and (c) again show the prediction, but for different hyperparameters.

Deciding what values for hyperparameters are appropriate is easy if we know the true f, which makes an GP unnecessary to begin with. For a real GP regression, we would be interested in adapting the hyperparameters based on our observations. To

this end, we will consider the marginal likelihood or evidence of observing outputs $\vec{y}$ given inputs $\vec{X}$

$$\mathcal{P}(\vec{y}|\vec{X},\vec{\theta}) = \int \mathcal{P}(\vec{y}|\vec{X},\vec{f}) \cdot \mathcal{P}(\vec{f}|\vec{X}) d\vec{f}, \tag{3.16}$$

where we omitted the explicit dependence on the right-hand-side. Thanks to the Gaussian nature, we can compute the integral exactly to yield the log-marginal-likelihood (LML)

$$\log\left(\mathcal{P}(\vec{y}|\vec{X},\vec{\theta})\right) = -\frac{1}{2}\vec{y}^T(K+\sigma_n^2 I)^{-1}\vec{y} - \frac{1}{2}\log|K| + \text{const.} \tag{3.17}$$

Numerically, we calculate the LML using Cholesky decomposition instead of direct matrix inversion, see algorithm 2.1 from [95, page 19] for details. By maximizing the LML w.r.t. $\vec{\theta}$ we find the best hyperparameters given the available data. For M-LOOP this optimization can be done using the Limited-memory BFGS (L-BFGS) algorithm, a method approximating the Broyden–Fletcher–Goldfarb–Shanno algorithm [101].

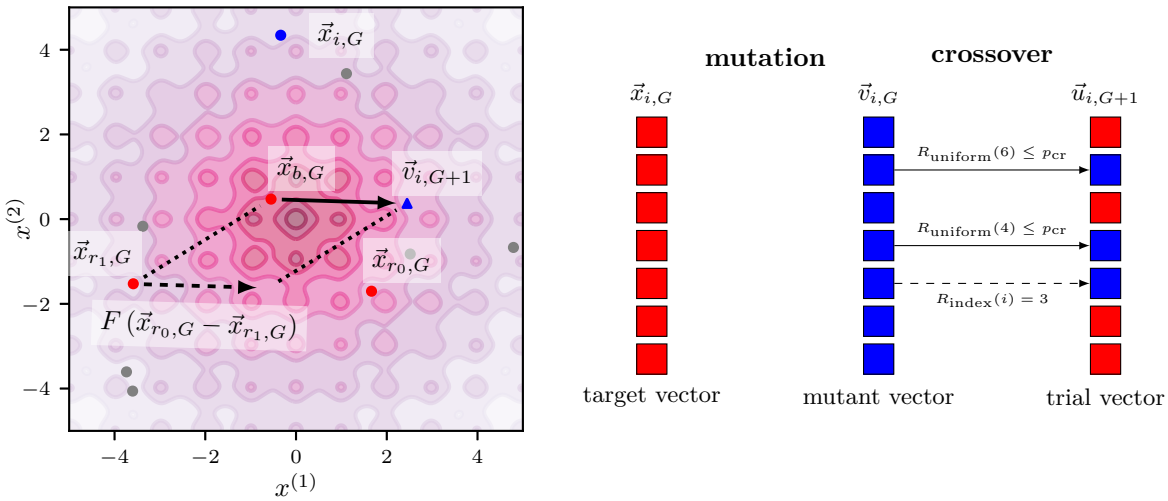
3.2. Differential Evolution

Differential Evolution (DE) is a method for solving global optimization problems based on stochastic population. It is particularly suited for non-linear and even non-differential problems. Unlike other direct search approaches such as the the Nelder-Mead algorithm [102], it does not solely rely on the greedy criterion. While slowing down it's convergence it helps to avoid getting trapped in local minima during the optimization [103]. As their name suggests, DE an other evolutionary algorithms are inspired by the natural selection process. There are many examples of their application in quantum optics [32, 104].

During the optimization process the algorithm evolves a set or population of D-dimensional parameter vectors, $\{\vec{x}_{i,G}\}_{i=0}^{N_P}$, in order to find the global optimum of some cost function $\mathcal{C}$. N_P denotes the size of the population while D is the number of parameters to be optimized. The subscript $G = 0, 1, ...$ is the generation number which is increasing as the optimization progresses. The initial set of vectors ($G = 0$) should in general be chosen randomly and distributed uniformly across the entire parameter space. The evolution from one generation to the next is performed in three steps: mutation, crossover, and selection which will be briefly described in the following.

Mutation

Mutation in DE describes the process of generating a set of vectors $\{\vec{v}_{i,G}\}_{i=0}^{N_P}$ from the the population of the generation G. In its simplest form this is performed by adding the weighted difference between two vectors to a third offset vector, all drawn



(a) Mutation with strategy `best1`: The current generation G is depicted as circles, with the target vector $\vec{x_{i,G}}$, the best vector of this generation $\vec{x_{b,G}}$, and two random vector highlighted in blue and red. The blue triangle denotes the mutant vector $\vec{x_{i,G}}$ after amplifying the differential variation by $F = 0.7$.

(b) Crossover: Generic mechanism of combining the target vector $\vec{x_{i,G}}$ and the mutant vector $\vec{v_{i,G}}$ to form the new trial vector $\vec{u_{i,G+1}}$. The random index is $R_{\text{index}} = 5$.

Figure 3.3.: Illustration of the mutation (a) and crossover (b) step of a differential evolution. In (a) the contour of the Ackley function is superimposed.

from the population of generation G [103] such that

$$\vec{v}_{i,G} = \vec{x}_{r_0,G} + F \times (\vec{x}_{r_1,G} - \vec{x}_{r_2,G})\,, \tag{3.18}$$

where $r_0, r_1, r_2 \in \{1, ..., N_P\}$ are chosen randomly and different from each other and the running index i. The amplification $F > 0$ is a real and constant factor that weights the differential variation $(\vec{x}_{r_1,G} - \vec{x}_{r_1,G})$. Other mutation strategies invole more vectors and therefore determine minimal population size. Table 3.1 gives an overview of the mutation strategies available in M-LOOP.

Crossover

Crossover is DE's strategy of parameter mixing, a procedure introduced to increase the diversity of the parameter vectors in the next generation $G+1$. From the target vector $\vec{x}_{i,G}$ and the mutant vector $\vec{v}_{i,G+1}$ the trial vector $\vec{u}_{i,G+1} = (u^1_{i,G+1}, ..., u^D_{i,G+1})$

strategy	offset vector	differential variation	minimal population size
`rand1`	$\vec{x}_{r_0,G}$	$\vec{x}_{r_1,G} - \vec{x}_{r_2,G}$	4
`best1`	$\vec{x}_{b,G}$	$\vec{x}_{r_1,G} - \vec{x}_{r_2,G}$	4
`rand2`	$\vec{x}_{r_0,G}$	$\vec{x}_{r_1,G} - \vec{x}_{r_2,G} + \vec{x}_{r_3,G} - \vec{x}_{r_4,G}$	6
`best2`	$\vec{x}_{b,G}$	$\vec{x}_{r_1,G} - \vec{x}_{r_2,G} + \vec{x}_{r_3,G} - \vec{x}_{r_4,G}$	6

Table 3.1.: Different mutation strategies for calculating the mutated vector $\vec{v}_{i,G+1}$ available in M-LOOP. $r_0, r_1, r_2, r_3, r_4 \in \{1, ..., N_P\}$ are chosen randomly and different from each other and from i. $\vec{x}_{b,G}$ is the best parameter vector for the current generation.

is formed, where the components $u^j_{i,G+1}$ $(j = 1, ..., D)$ satisfy

$$u^j_{i,G+1} = \begin{cases} v^j_{i,G+1} & \text{if } R_{\text{uniform}}(j) \leq p_{\text{cr}} \text{ or } j = R_{\text{index}}(i) \\ x^j_{i,G} & \text{if } R_{\text{uniform}}(j) > p_{\text{cr}} \text{ and } j \neq R_{\text{index}}(i) \end{cases} \tag{3.19}$$

where $R_{\text{uniform}}(j) \in [0, 1]$ is the jth evaluation of a uniform random number generator, $p_{\text{cr}} \in [0, 1]$ the predefined crossover probability and $R_{\text{index}} \in \{1, ..., D\}$ is randomly chosen index, ensuring that at least one parameter from the the mutated vector is used.

Selection

The trial vector is now compared to the target vector with respect to the cost function. If the trial vector has a smaller cost than the target vector, it is accepted and becomes the new target vector $\vec{x}_{i,G+1}$ in the next generation; otherwise, $\vec{x}_{i,G}$ is moved to the next generation.

Figure 3.5 shows how DE can used for finding the global minimum of the Ackley function

$$f_{\text{ackley}}(\vec{x}) = -A \cdot \exp\left(-B\sqrt{\frac{1}{2}|\vec{x}|^2}\right) - \exp\left(\frac{1}{2}\left(\cos\big(C \cdot x^{(1)}\big) + \cos\big(C \cdot x^{(2)}\big)\right)\right) + A + e,$$

whose minimum occurs at $\vec{x} = (0, 0)^T$. This function can be used for bench-marking optimizers and illustrates the advantages of global optimizers. While local optimizers would likely get trapped in local minima, DE will find the global minimum within ~ 11 generations.

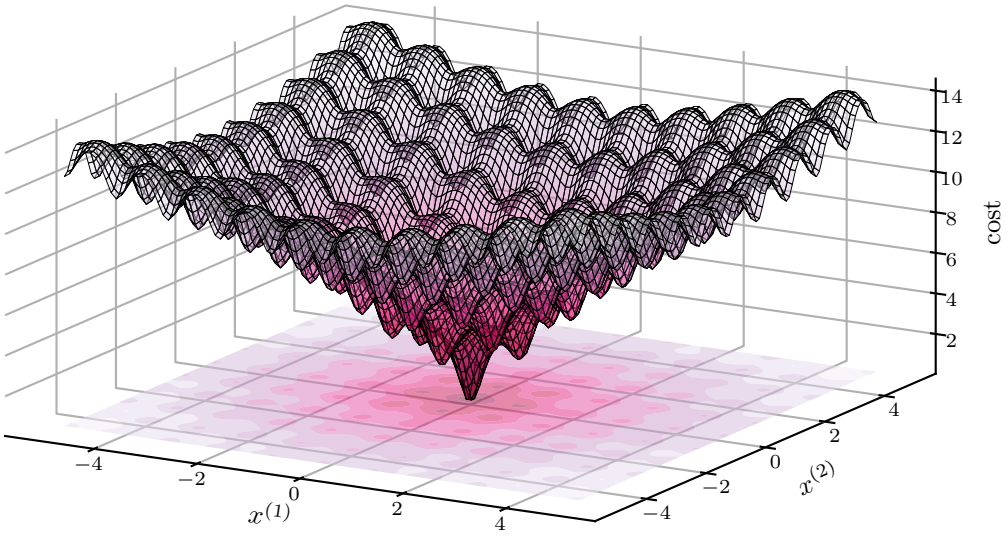
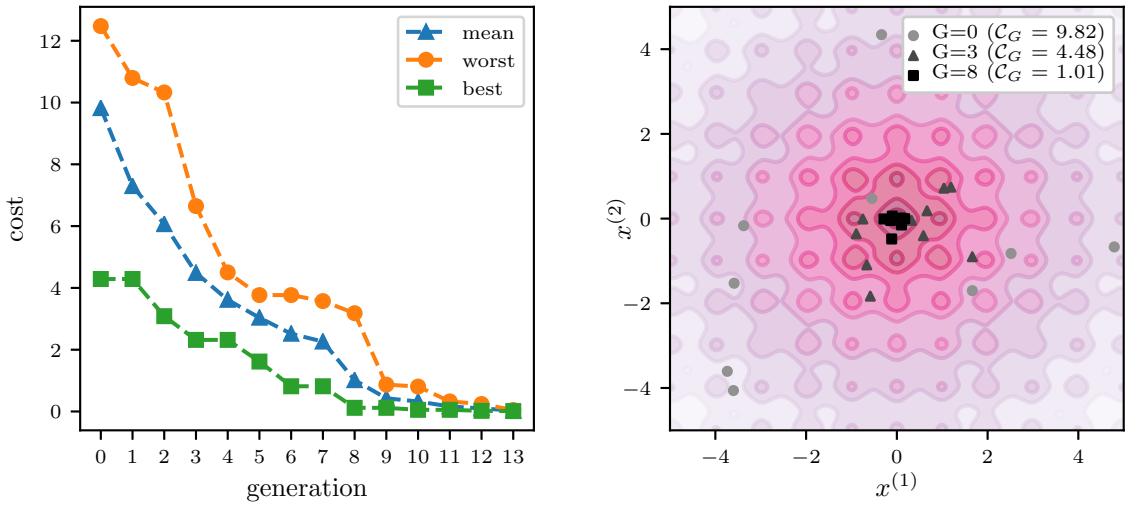


Figure 3.4.: 3D representation of the $f_{\text{ackley}}(\vec{x})$. Finding the global minimum using gradient descent would be very challenging.



(a) Evolution of mean, worst and best cost of the population.

(b) Evolution of the population at $G = 0$ (initial random population), $G = 3$ and $G = 8$.

Figure 3.5.: Finding the minimum of the Ackley function using DE.

3.3. M-LOOP Optimization Routine

After introducing the key concepts of GPR and DE, we now describe how they are employed in an optimization routine. Optimization routine refers to the process of iteratively generating new trial parameter vectors in order to achieve some optimal performance.

Let us first clarify a few terms related to the API. In M-LOOP, a *controller* handles the different threads of the program and determines which strategy to use for generating new trial parameters *learner* and *trainer*. Available strategies are

- for the learner: random[1], Neural Net (NN)[2], Gaussian Process, Nelder-Mead[2], and Differential Evolution
- for the trainer: random[1], Nelder-Mead[2], and Differential Evolution.

The controller is connected to the physical experiment via the *interface*.

Generally one has to differentiate between two cases. In the first case, no machine learner is used, making the program acts as regular optimizer only using one of the three available training strategies.

If a NN or GP learner is being used, the optimization routine primarily draws new points from the predictions of the machine learner. Nevertheless an initial training is necessary to seed the first iteration of the learner cycle. After this, the controller can be configured only use the machine learner to generate new parameters. Here one should note, that over time the effective training data for our ML models increases as the experiment is run over and over again. Therefore, the time needed to fit the model does as well. It is possible that this time exceeds the duration of one experiment. M-LOOP therefore offers the option to perform training runs (using one of the available strategies) while the learner is fitting the model. Figure 3.6 illustrates a typical optimization routine, where one sees that the number of training runs increases towards the end of the optimization.

Using M-LOOP over the course of this thesis, we have identified several problems related to the design of the software that partially have been discussed and solved together with developers of M-LOOP. Two other possible improvements of the API have been identified and will be brought to the attention of the lead developers. First, the treatment of ”bad” iterations might be improved. Bad here refers to runs where no appropriate cost could be returned. At the moment, this solely attributed to a wrong parameter choice, and bad runs are included in the Gaussian fits and contribute to the parameters selection. As we will see in the following chapter, this

[1] generate random parameters for each run
[2] this learner/trainer was not considered for this thesis

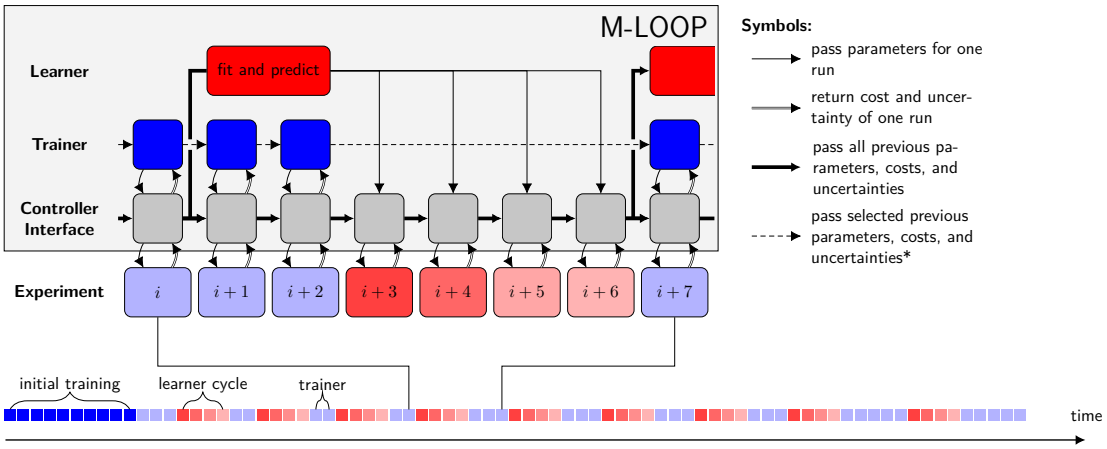


Figure 3.6.: Graphical depiction of a typical optimization routine.

can greatly affect the optimizers ability to develop an appropriate statistical model for the problem. An option to exclude all or certain bad runs from the Gaussian fit might improve the overall performance of the optimizer.

Second, M-LOOP currently only considers uncorrelated, white noise on top of the correlated changes to the cost arising from changing the parameters. Long-term drifts in the cost, that would exhibit some correlation in time cannot be taken into account accurately. Using a GPR with an additional kernel that could describe correlated changes in the cost that cannot be attributed to changes in the parameters might help to improve M-LOOP's performance in an experimental setting.

4. Online Optimization in a Single-Atom Trap Experiment

The problem of finding the best way to use resources to steer a system is known as optimal control, where we call the measure for how good the control over the system is as cost ($\mathcal{C}$). Generally we can differentiate between two categories of optimizations: open-loop and closed-loop. For open-loop optimizations the inputs or parameters are determined beforehand by minimizing some model of the expected cost $\mathcal{C}_{\text{mod}}$. In contrast, closed-loop optimizations use the output of the system itself as feedback for the optimization. Both schemes are illustrated in figure 4.1.

In our concrete case the "system" to be optimized can be either a initial state preparation or a state readout process. One major aspect of the control of these systems, as described in sections 2.4.1a and 2.4.3, are the applied laser fields. Those, in turn, are controlled by modulating the RF input of several acousto-optical modulators (AOM).

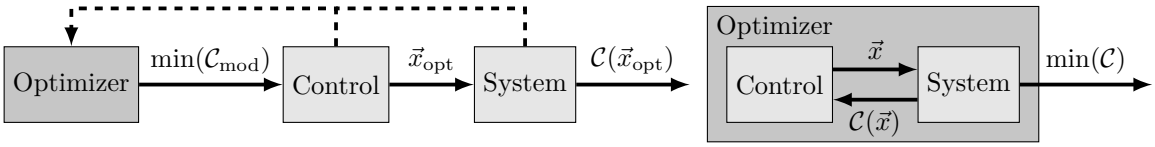
For this thesis we have developed several optimizers based on the open source M-LOOP API(sec. 3.3), implementing both (unsuccessful) open-loop optimization on simulation data and closed-loop optimization with feedback from the experiment itself.

This chapter is structured as follows: First, we briefly introduce the arbitrary waveform generator (AWG), which provides the RF inputs to the AOMs. Using an AWG allows for a more flexible temporal and spectral form the of the applied laser pulses. Second, we will give a short overview of how the optimizer was interfaced with the experiment and the DAQ. Lastly, the remaining two sections present the experimental results of our optimizations for the state preparation and readout processes.

Convention for Denoting Error Tolerance and Notation

Error tolerances will be denoted as follows, unless explicitly stated otherwise: *measured* data in section 4.2 is displayed with errorbars corresponding to three standard deviations (3σ), while in section 4.3 only one standard deviation (1σ) is displayed. Data for *fits* in both section has either no error tolerance (e.g. for the hyperparameters returned by M-LOOP) or a 1σ error-band (e.g. for the cross-sections).

In addition, for plots in section 4.2 the simplified notation $\text{pump}_{i\to j} \equiv \text{pump}ij$ was used. To prevent ambiguity, references to the "best" or "optimal" parameters



(a) Open-loop optimization. **(b)** Closed-loop optimization.

Figure 4.1.: Difference between open-loop (a) and closed-loop (b) optimization. The goal is to optimize the control of some system such that the cost $\mathcal{C}$ is minimized. In (a) the optimizer optimizes the cost of some model $\mathcal{C}_{\mathrm{mod}}$. It uses some information from the control and the system, indicated by dashed arrows. It predicts some optimimal parameters $\vec{x}_{\mathrm{opt}} = \mathrm{argmin}(\mathcal{C}_{\mathrm{mod}})$ which then are passed through the control into the system, yielding some $\mathcal{C}(\vec{x}_{\mathrm{opt}})$. In (b) the control and system are part of the optimization process. They exchange parameters $\vec{x}$ and the resulting cost $\mathcal{C}(\vec{x})$ until some optimal set of parameters argmin(C) are found.

mean the best *predicted* parameters of the optimization. In general, those are close to the actually best *observed* parameters.

4.1. Technical Implementation of a Closed-Loop Optimization

For running the closed-loop configuration and implementing the results from the open-loop optimizations, several ingredients are necessary:

1. **Output from the Experiment**
   automatized evaluation of the performance of the state preparation (see sec. 2.4.1a) and readout process (see sec. 2.4.3)

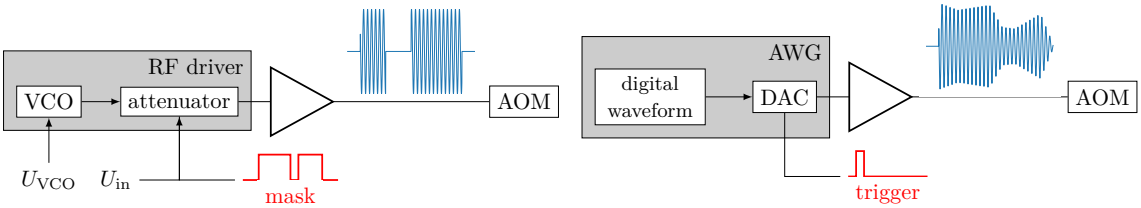
2. **Inputs to the Experiment**
   control of the laser fields via the RF input of the AOMs (see sec. 2.5)

3. **Optimization Routine**
   developed based on M-LOOP (see sec. 3.3)

4. **Interfacing**
   replacing the human generate input by machine (see sec. 4.1.2)

4.1.1. Arbitrary Waveform Generator

The AOMs used in our experiments RF input in the range of roughly 60 MHz $\sim$ 400 MHz. Generally, the RF input is generated by a voltage-controlled oscillator (VCO) whose output is attenuated to a given value U_{in}. Both, the input voltage of the VCO (U_{VCO}) and U_{in}, can be set internally or externally. In our setup



(a) Operation of an AOM using a RF driver. A VCO generates a sinusoidal signal with a frequency depending on the input voltage U_{VCO}. This signal is attenuated to U_{in}, which is masked square pulses. The resulting waveform is fed into the AOM through a amplifier.

(b) Operation of an AOM using a AWG. A digital waveform is generated and loaded to the buffer of the AWG. Upon receiving a trigger the waveform is fed directly into the amplifier for amplification before being passed to the AOM.

Figure 4.2.: Operating an AOM with a RF driver (a) and an AWG (b). With the AWG a greater variety of waveforms can be fed into the AOM. Both the mask in (a) and the trigger in (b) are provided by the same pattern generator.

the input of U_{in} is an external logic signal, producing a sequence of square pulses with fixed amplitude that are multiples of 20 ns or 2 µs,. Finally, the attenuated VCO signal is amplified before being fed into the AOM (fig. 4.2a).

To allow for more freedom in the applied RF pulses and hence in the laser pulses, the decision was made to acquire two arbitrary waveform generators (AWGs)[1]. AWGs are electronic devices capable of generating analog voltage outputs from digital waveforms. Unlike traditional function generators, whose output is generally restricted to certain types of waveforms, AWGs are capable to generate, in principle, any shape of output, hence their name. In practice, their operating range is mainly limited by the maximum output voltage, the output amplitude resolution, the sampling rate and bandwidth, and, depending on the operation mode, the memory depth.

In the experiment the trigger will be provided by the same logic pattern generator that is used for masking the input voltage of the RF drivers. The output of the AWG can easily be shifted relative to the time the trigger was received by analog or digital means. A schematic block diagram contrasting the operation of an AOM using an AWG or the RF driver can be found in figure 4.2b.

[1]Spectrum Instrumentation (short: Spectrum) M4i.6622 (4 channels, bandwidth 200 MHz) and M4i.6631 (2 channels, bandwidth 400 MHz)

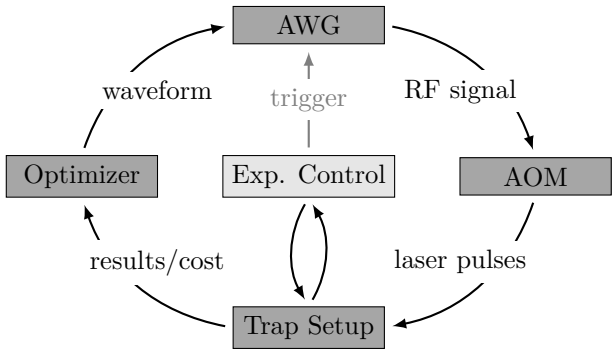


Figure 4.3.: Simplified block diagram of the optimization loop. The optimizer generates waveforms that are loaded to the AWG and replayed upon receiving a trigger from the pattern generator. The RF signal is fed into an AOM which determines the resulting laser pulse sequence that is steering the trap setup. Finally, the output of one iteration is used by the optimizer to generate new waveforms. The experimental control is responsible for various settings and starts and stops a measurement; thus communication between the optimizer and the experimental control is necessary.

4.1.2. Interfacing M-LOOP and the Experimental Infrastructure

In both labs the experiments are controlled from a single computer (*atom* or *atom2* in lab 1 and 2, respectively), whereby a master-slave configuration is used for atom-atom experiments. Additional computers that are needed for tasks like the regular calibration of the magnetometers or the polarization in the long optical fibers are also present in the labs.

A specifically designed program (countersuite) allows to control everything that needs to be set up to ms timescales[2], including the voltages of a digital output board[3] that are fed into the RF drivers of the AOMs as U_{in}. While simple, discrete scans that are used for optimizations are possible using the current software framework, more complex optimizations using external programs made some changes necessary. In principle the optimization is implemented as shown in figure 4.3. The optimizer generates a waveform which is loaded to the AWG, upon which the measurement is started by countersuite. The AWG is repeatedly triggered through the pattern generator, to output the waveform through the DAC. The resulting laser pulse sequence steers the experiment in the trap setup. Eventually, the measurement is stopped by countersuite and the output is evaluated and used to generate new waveforms.

The technical implementation called for setting up an additional computer, *jupiter*,

[2]For settings that change at even faster timescales, down to 20 ns a logic pattern generator is used.
[3]Data Translation DT330

in lab 2, both due to the need for an additional PCIe slot for the AWG and the potential risk of slowing down *atom2* with the intensive calculations needed for fitting the GP. Moreover, an extension to countersuite was developed (scheduler). It enables *automatized* execution of "task", sets of instructions about which settings to use for and when to start and stop a measurement. The execution of the task and the end of the run are communicated through a signal file on a hard drive accessible to both computers. Additionally, scheduler can be used to "schedule" tasks at specific times which can be used to for example switch between measurements without human input.

4.2. Optimizations in the Pumping Process

One area of the experiment where the possibility for improvements was identified was the optical pumping process (sec. 2.4.1c). At the moment, only in about $\eta_{\text{prep}} \lesssim 80\%$ of the cases the desired state $|1,0\rangle$ can be prepared within the $\sim 2.3\,\mu\text{s}$ of optical pumping, which has a significant effect on the event rate and entanglement fidelity.

4.2.1. Model for Determining the Pumping Efficiency

As mentioned before, optimization on the state preparation efficiency (η_{prep}) itself, which involves three (destructive) measurements, is very time-consuming and hence slows down the feedback loop. The Event rate for an ionization-based measurement scheme such as the one used here (see sec. 2.4.1b) are limited by the time that is needed to trap an atom ($\mathcal{O}(1\,\text{Hz})$). Instead, we will optimize a closely related quantity that describes the probability of detecting a single photon in a given excitation try, η_{det}. Hereby, the event rate is limited by the repetition rate times the success rate of the single photon collection (typically $\sim 50\,\text{kHz}$ and $\sim 5‰$ resulting in $\mathcal{O}(250\,\text{Hz})$). Simulations based on a quantum jump model [53], performed in [45] and discussed in more detail below, show that the detection of a photon after an excitation try is >28 times more probable when the atom is in the desired state relative to the other states, which makes the photon detection efficiency an ideal quantity to optimize on. We will briefly show how we arrive at this figure in the following.

After the optical pumping the atom is in a mixed state of the Zeeman-substates of $5^2S_{1/2}$, described by the density matrix

$$\begin{aligned}\rho_{\text{prep}} = {} & \eta_{\text{prep}}\,|1,0\rangle\langle 1,0| + (1-\eta_{\text{prep}})\times \\ & \times\left(A\,|1,+1\rangle\langle 1,+1| + B\,|1,-1\rangle\langle 1,-1| + \sum_m C_m\,|2,m\rangle\langle 2,m|\right),\end{aligned} \tag{4.1}$$

where A, B, and $C_m\,(m = \{-2,...,+2\})$ describe the relative populations in the other Zeeman-substates and trivially sum to 1.

$F'' = 0$ and $F'' = 1$ are only separated by 72 MHz while the separation between $F = 1$ and $F = 2$ is 6.8 GHz. Thus, applying the excitation pulse that is intended to be π-polarized and resonant to the $F = 1 \leftrightarrow F'' = 0$ transition only affects the populations in $F = 1$. We can identify three different effects, each described by an individual excitation efficiency:

$\eta_{\text{exc}}^{(r)}$ resonant excitation from $|1, 0\rangle$ to $|F'' = 0, m_F = 0\rangle$

$\eta_{\text{exc}}^{(\sigma)}$ resonant excitation from $|1, \pm 1\rangle$ to $|F'' = 0, m_F = 0\rangle$ due to residual circular polarization of the excitation laser

$\eta_{\text{exc}}^{(o)}$ off-resonant excitations with π-polarized light from $|1, \pm 1\rangle$ to $|F'' = 1, m_F = \pm 1\rangle$[4]

Furthermore, there is a finite number of events where two photons are emitted for one excitation try. Of these two-photon events only those where the first excitation was resonant have a significant branching ratio. In a two-photon event either the first or the second photon can be detected with non-vanishing probability. Thus, there are in total eight relevant excitation processes with different effect on the final atom-photon state [45]. However, for the purpose of optimizing the pumping efficiency, only the state of the atom before the excitation is of interest. We can write the probability to emit a $\sigma^{\pm}$-polarized photon for a given excitation try

$$\Pr_{\text{det}} = \left[\eta_{\text{prep}} \cdot \left(\Pr_{\text{det}}^{(r)} + \sum_{i \in \{r,\sigma,o\}} \Pr_{\text{det}}^{(r,i)} \right) + (1 - \eta_{\text{prep}}) \cdot (A + B) \cdot \left(\Pr_{\text{det}}^{(o)} + \Pr_{\text{det}}^{(\sigma)} \right) \right], \tag{4.2}$$

where $\Pr_{\text{det}}^{(\ldots)}$ is the excitation efficiency of the corresponding process, conditioned our ability to detect it. We assumed that unwanted processes where equally likely for atoms initially in either $m_F = \pm 1$ state. $\Pr_{\text{det}}^{(r,i)}$ describes two-photon events the second excitation was resonant (r, r), off-resonant (r, o), or due to σ-polarized light (r, σ). If both excitations were resonant, the first photon was π-polarized and therefore not detectable.

The misalignment of the polarization $\chi_{\text{exc}} = \cos\alpha \cdot V + \sin\alpha \cdot H$ of the excitation laser was set to $\alpha = 5°$. The temporal shape of the excitation pulse determines the prevalence of off-resonant excitations. Its finite duration (FWHM=20.35 ns) leads to a significant spectral broadening ($\Delta\Gamma = 21.6$ MHz) of the otherwise narrow laser frequency. As higher excitations powers P_{exc} also increase the probability for off-resonant excitations, P_{exc} is set at around 80% of the power yielding the largest number of emitted photons ($\max_{\eta_{\text{det}}}(P_{\text{exc}})$. A simulation considering $\eta_{\text{prep}} = 80\%$ and all remaining population split amongst the other two $F = 1$ substates ($A + B = 100\%$) is shown in figure 4.4a.

[4] Off-resonant excitation $|1, 0\rangle$ to $|F'' = 1, m_F = 0\rangle$ are dipole forbidden.

The simulations show that emitting a detectable photon is about 28.3 times more likely if the atom is initially in $|F=1, m_F=0\rangle \equiv |\Psi_0\rangle$ compared to events where the atom was in either $|F=1, m_F=\pm 1\rangle$. We can also infer the conditional probability that the pumping was successful given a detectable photon was emitted

$$\Pr(\Psi_0|D) = \frac{\Pr(\Psi_0)\cdot\Pr(D|\Psi_0)}{\Pr(D)} = \frac{\eta_{\text{prep}}\cdot\int\left(\Pr_{\text{det}}^{(r)} + \sum_{i\in\{r,\sigma,o\}}\Pr_{\text{det}}^{(r,i)}\right)dt}{\int\Pr_{\text{det}}dt} \tag{4.3}$$

as a function of the state preparation efficiency $\eta_{\text{prep}} = \Pr(\Psi_0)$ for different values of $A+B$ (figures 4.4b). We see that increasing η_{prep} will increase the number of detected photons and conversely, that an increased number of detected photons can almost exclusively be attributed to an increased η_{prep}. Finally, figure 4.4c shows the ratio of $\Pr_{\text{det}}$ between those events where the atom is prepared in $|1,0\rangle$ and those where it was in any other Zeeman substate as a function of η_{prep} and different $A+B$. We can directly read off the value for

$$\frac{\Pr(D|\Psi_0)}{\Pr(D|\overline{\Psi_0})} = \frac{\eta_{\text{prep}}}{1-\eta_{\text{prep}}}\cdot\frac{\int\Pr_{\text{det}}(\Psi_0)dt}{\int\Pr_{\text{det}}(\overline{\Psi_0})dt} = \frac{\eta_{\text{prep}}}{1-\eta_{\text{prep}}}\cdot 28.3\cdot(A+B)^{-1} \tag{4.4}$$

from figure 4.4c at $\eta_{\text{prep}} = 1-\eta_{\text{prep}} = 0.5$. We should, however, note that changing the excitation power would significantly change the above calculations as further increasing P_{exc} will lead to more excitations from $|1,0\rangle$ as well as from $|1,\pm 1\rangle$, thus lowering this factor.

Only a fraction η_{col} of all emitted (correctly polarized) photons can be actually detected on the APDs. η_{col} describes the photon collection efficiency that is much smaller than 1 due to finite angular acceptance, imperfect detectors, or losses in the fiber leading to the APDs. Therefore we can express the number of photons we detect on the APDs, n_{ph}, as

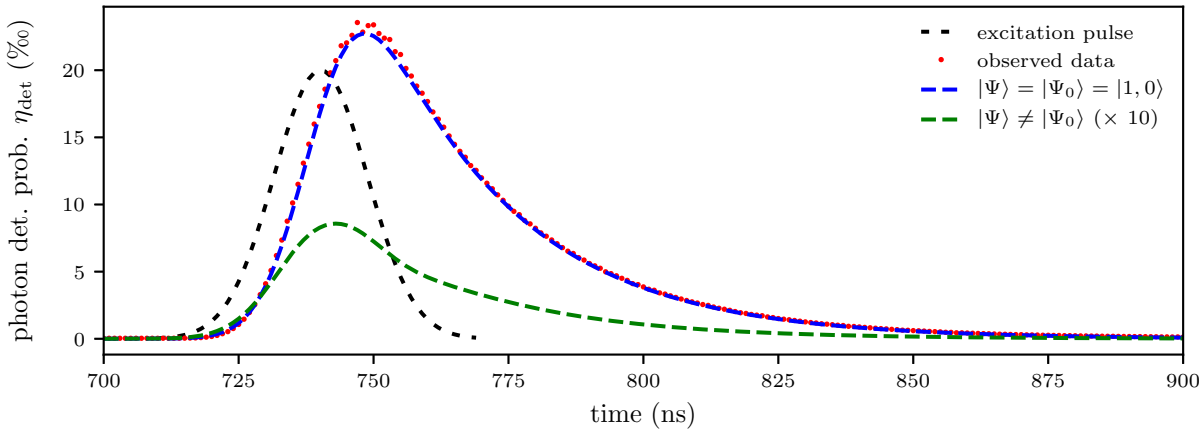
$$n_{\text{ph}} = \eta_{\text{col}}\cdot\Pr_{\text{det}}\cdot n_{\text{tries}} + n_{\text{dark}}. \tag{4.5}$$

where n_{tries} is the (known) number of excitation attempts and n_{dark} is the number of dark counts of the APDs, which can largely be neglected as it is only $\mathcal{O}(10\,\text{Hz})$.

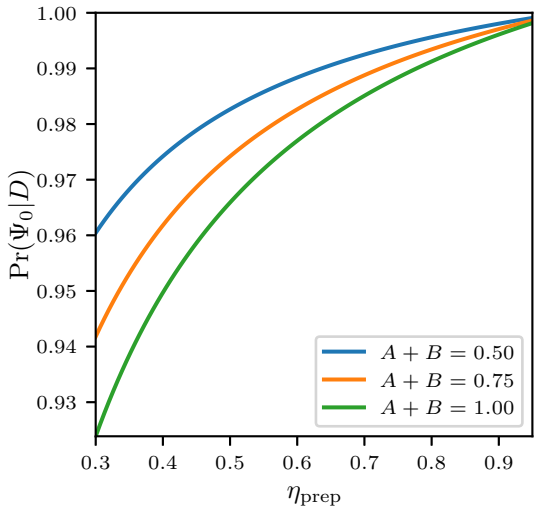
The probability to detect a photon for a given excitation try is identical to and independent from all other excitation tries, giving us a constant mean photon detection rate $n_{\text{ph}}/n_{\text{tries}}$. This allows us to model the process by a Poisson distribution yielding the probability to observe exactly k photons

$$\Pr(k) = \frac{n_{\text{ph}}^k\cdot e^{-k}}{k!}. \tag{4.6}$$

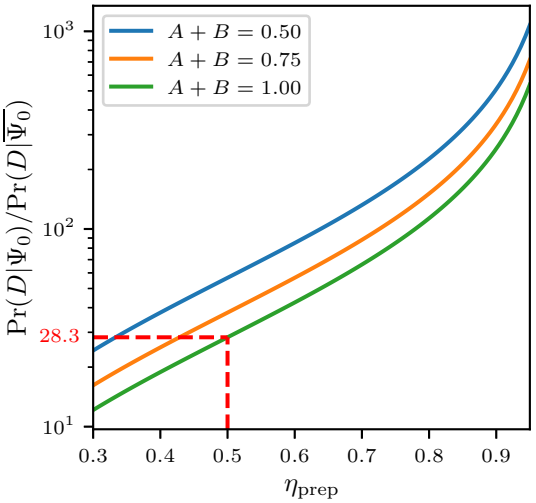
The mean as well as the variance of this PDF are given by n_{ph}. Therefore we will



(a) Measured photon detection histogram (red dots) and fitted photon detection probabilities for all processes where the atom was initially in $|\Psi_0\rangle = |F = 1, m_F = 0\rangle$ (blue) or in any other Zeeman substate (green). The excitation pulse (black, a.u.) is superimposed to visualize the relative timing. Data is identical to what is presented in [45]. It is displayed only for illustration, thus without any associated error.



(b) Conditional probability that the atom was initially in $|1,0\rangle$, given that a photon was detected within the detection window (a), $\Pr(\Psi_0|\mathrm{det})$, as function of the state preparation efficiency η_{prep}. $\Pr(F = 2|\overline{\Psi_0})$ indicates the population in $F = 2$ as a fraction of the population that is not in $|1,0\rangle$, $1 - \eta_{\mathrm{prep}}$.



(c) Relative chance that a photon is detected in case the atom was in $|1,0\rangle$ compared to in case it was not for different values of $A + B$. The aforementioned factor of 28.3 can be read of at $0.5 = \eta_{\mathrm{prep}} = 1 - \eta_{\mathrm{prep}}$.

Figure 4.4.: Simulated photon detection possibilities (a) from [45]. From them it is possible to calculate the conditional probability that the atom was initially in $|F = 1, m_F = 0\rangle$ after the pumping given we detect a photon.

define cost and uncertainty for the pumping optimizer as

$$\mathcal{C}_{\text{pump}} = 1 - \alpha \cdot \frac{n_{\text{ph}}}{n_{\text{tries}}} \tag{4.7a}$$

$$\text{and } \mathcal{U}_{\text{pump}} = \alpha \cdot \frac{\sqrt{n_{\text{ph}}}}{n_{\text{tries}}}. \tag{4.7b}$$

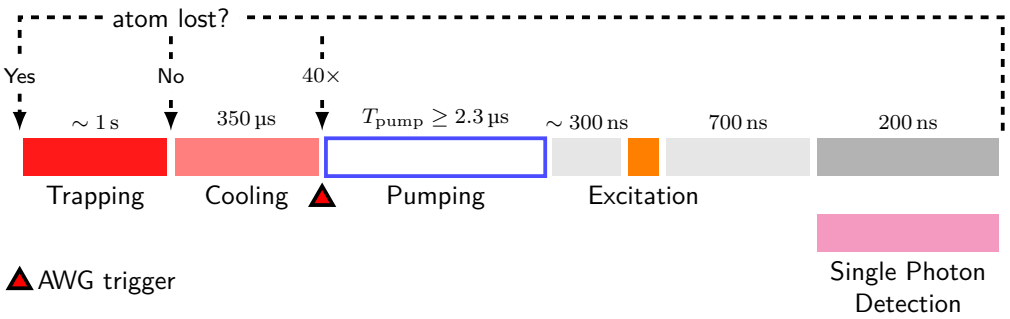
where α is just an arbitrary scaling parameter, typically 100, which is just used for computational reasons.

4.2.2. Experimental Setup for Optimizing the Initial State Preparation

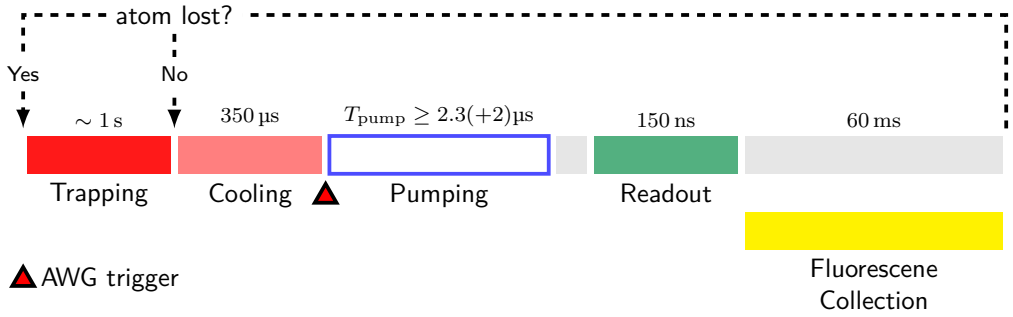
To optimize the initial state preparation, two output channels of the AWG are used to manipulate the $\text{pump}_{1\to1}$ and $\text{pump}_{2\to1}$ laser fields via the respective AOMs which are operated at around 200 MHz and 75 MHz. After finding the optimal settings by maximizing η_{det} at fixed excitation power, these may be tested by performing an excitation-power-scan and by determining η_{prep} according to section 2.4.1b. Optimizing based on η_{det} enables fast iterations of the optimizer, about one minute per parameter set. By scanning the excitation power it is possible to eliminate fluctuations in η_{det} that arise from changes in in the excitation power occurring over time. Lastly, we want to circle back and motivate the performance physically.

The experimental sequence for the optimization and the excitation-power-scan is depicted in figure 4.5a. It is started once a waveform has been loaded to the buffer of the AWG. The pattern generator triggers the AWG such that the loaded waveform is executed in the pumping window of $\geq 2.3\,\mu\text{s}$. To verify the timings the light pulses are monitored with fast photodiodes. Roughly 300 ns after the end of the pumping window, a short excitation pulse is applied. This might lead to the emission of a photon which can be registered by either single-photon APD within the 200 ns photon detection window. This window is delayed by 700 ns relative to the excitation pulse. The excitation process is repeated 40 times before the atom needs to be cooled again. The fluorescence trace (fig. 2.8) is continuously recorded and the control program switches back to trapping once the atom is lost. The experimental sequence is repeated for 60 seconds before the output is evaluated. This output will be used to generate new waveforms to load to the AWG during the optimization.

For determining η_{prep} the experimental sequence will be modified as shown in figure 4.5b. The state preparation efficiency is determined as described in section 2.4.1, inferring the population in $|F = 1, m_F = 0\rangle$ as described in 2.28. After the pumping window, instead of the excitation pulse as before, the cycling, ionization, and possibly readout lasers are applied. In the following 60 ms the presence of the atom in the trap is checked, returning the sequence to the cooling or trapping depending on



(a) Determining η_{det}: After the atom has been trapped and cooled (red), the AWG is triggered (red triangle) such that the waveform is replayed during the pumping window (blue). Waiting times (gray) before and after the excitation (orange) are necessary such that the photons can be detected during the acceptance window (magenta). After 40 preparation-excitation attempts, the atom is additionally cooled (red) to assure a low atom temperature is maintained. Loosing the atom results in starting the sequence from at the trapping.



(b) Determining η_{prep}: The pumping window is followed by the atomic state readout (green) as described in the text. By collecting the fluorescence for 60 ms (yellow) we determine if the atom was ionized. Depending on the outcome of this test, the sequence is restarted at the trapping or cooling.

Figure 4.5.: Timing schemes the experimental sequences used for determine η_{det} (top) and η_{prep} (bottom).

the outcome. If the readout laser was not used, the probability of ionizing the atom is equal to the remaining population in $F = 2$. In case the readout laser is used with polarization $\chi_{\mathrm{ro}} = V$ ($\chi_{\mathrm{ro}} = H$) we ionize the atom if it is in either $|2, m_F\rangle$, $|1, 0\rangle$ or $\frac{i}{\sqrt{2}}(|1, +1\rangle + |1, -1\rangle)$ ($\frac{i}{\sqrt{2}}(|1, +1\rangle - |1, -1\rangle)$). Therefore, we can determine η_{prep} as

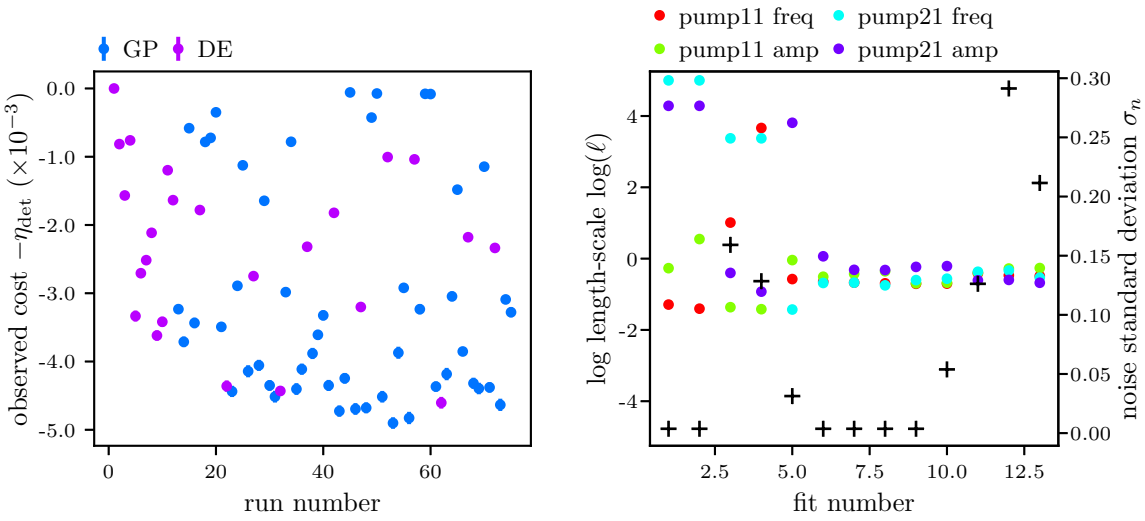
$$\eta_{\mathrm{prep}} = 1 - \Big[\mathrm{Pr}_{\mathrm{red}}(\chi_{\mathrm{ro}} = V) + \mathrm{Pr}_{\mathrm{red}}(\chi_{\mathrm{ro}} = H) + \big[1 - \mathrm{Pr}_{\mathrm{red}}(P_{\mathrm{ro}} = 0)\big]\Big]. \tag{4.8}$$

4.2.3. Measurements for Optimizing the Initial State Preparation

The following section will summarize some of the optimizations performed for the pumping, along with an analysis of the results and a meta-analysis of the optimization itself. Several problems, mostly concerning parameter choice, have been identified. Unfortunately, due to technical difficulties with the experiment, we were not able to repeat the optimizations after fixing those issues. However, with the infrastructure put in place for this thesis, doing so in the future should not pose major problems.

4.2.3a. Verification of the Optimization Process

In the current experimental setup, the pumping is optimized by a series of independent scans over the frequency and amplitude of both $\text{pump}_{1\to1}$ and $\text{pump}_{2\to1}$. Within one scan, one parameter singled out and varied in discrete steps while



(a) Obtained cost for each run.

(b) Obtained hyper-parameters for each fit.

Figure 4.6.: Overview for the optimization laid out in sec. 4.2.3a where the four parameters are the frequencies and amplitudes of both pumping lasers. Note that in (a) the uncertainties are smaller than the markers and the different colors correspond to the optimizer that provided the parameters for the corresponding run, either the GP or the DE. (b) show how the hyper-parameters that were fitted during the optimization evolve as more data becomes available. For convenience the logarithm of the length-scale ($\log_{10}(\ell)$) has been used. The square root of the noise level for each fit is indicated on the secondary axis.

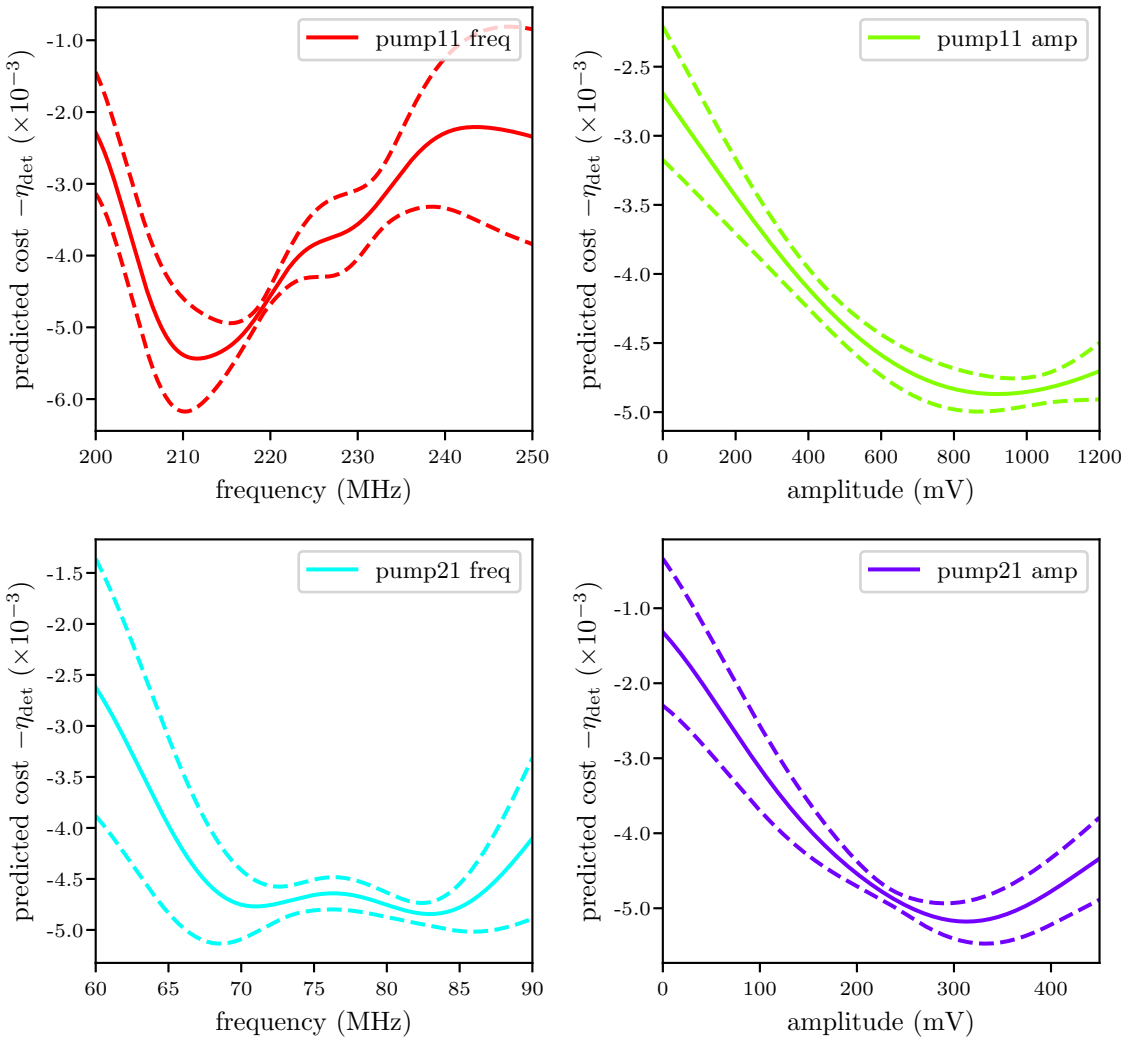


Figure 4.7.: One dimensional cross-sections centered around the best predicted parameters for the pumping optimization with four free parameters.

keeping all other parameters fixed. Rather than exploring the entire four dimensional parameter scan, we concentrate on several points along one dimensional cross-sections that, together with experience and knowledge of the experiment, let us estimate a set of optimal parameters. Using a ML optimizer, in particular on capable of global optimization such as M-LOOP, can relax this requirement for prior knowledge while also offering a more quantitative approach for estimating the optimal parameters.

To verify the optimization process we started by replicating the current pumping sequence. The RF pulse going into the $\mathrm{pump}_{1\to1}$ AOM is rectangular in shape and has fixed frequency and amplitude. The corresponding waveform for $\mathrm{pump}_{2\to1}$ is a series of identical square pulses (width $\sim$ 200 ns, fixed amplitude and frequency)

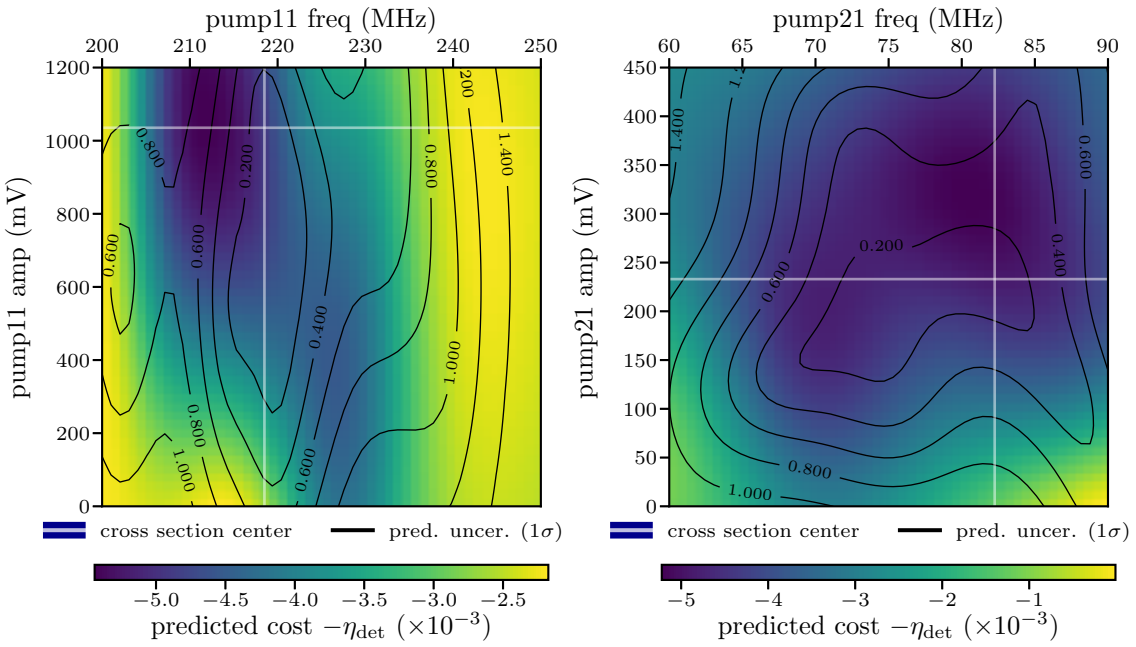


Figure 4.8.: Two dimensional cross-sections centered at the best predicted parameters for the pumping optimization using four free parameters. The predicted cost as a function of two parameters is plotted, with the cross-section-lines corresponding to the best predicted parameters (fig. 4.7) indicated by white lines.

with a duty cycle of 80%. We had several goals with this optimization. First, we wanted to test the hardware, software and their interplay. Second, the advantages of using a Gaussian process for the optimization should be demonstrated. Finally, the obtained optimal waveform was supposed to stand in as benchmark facilitating the comparison of different optimizations.

With only four parameters the optimizer was able to find an optimum of $\mathcal{C}_{\text{pump}}$ within a few runs (fig. 4.6a), considering the large parameter space. The evolution of the hyper-parameters (fig. 4.6b), in particular the length scales, indicate that M-LOOP is able to successfully learn the underlying landscape. For each parameter, the length scale seems to stabilize after the sixth fit. Post-processing of the obtained data found comparable values. The length scale for each parameter is $\mathcal{O}(1)$, indicating that all four parameters have similar relevance in this parameterization. The noise variance σ_n (fig. 4.6b) has not settled but is fluctuating at a level that is does not greatly affect the overall uncertainty of the Gaussian fit.

The *predicted* best parameters lie well within the parameter boundaries, and correspond to a predicted cost of $\mathcal{C}_{\text{pump}} = 4.9(4)\,‰$, which is compatible with the best *observed* cost of 5.1. The one dimensional predicted cost and uncertainty landscapes, centered at the predicted best parameters, for each parameter are on

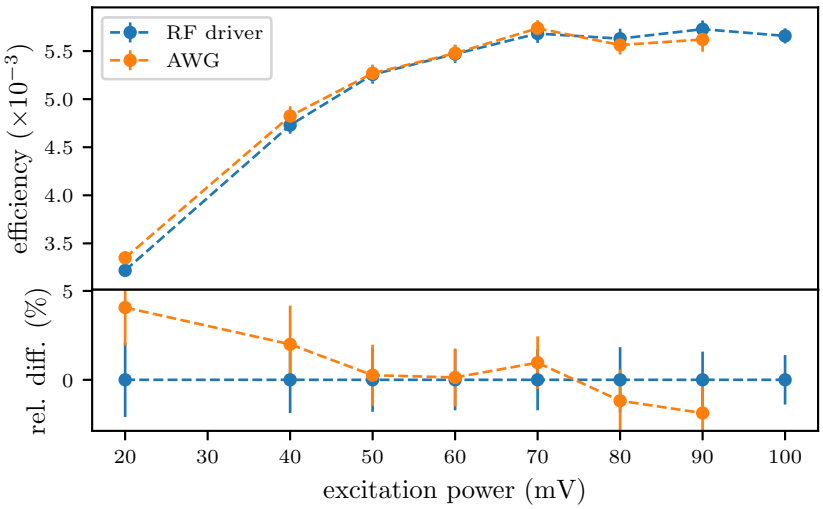


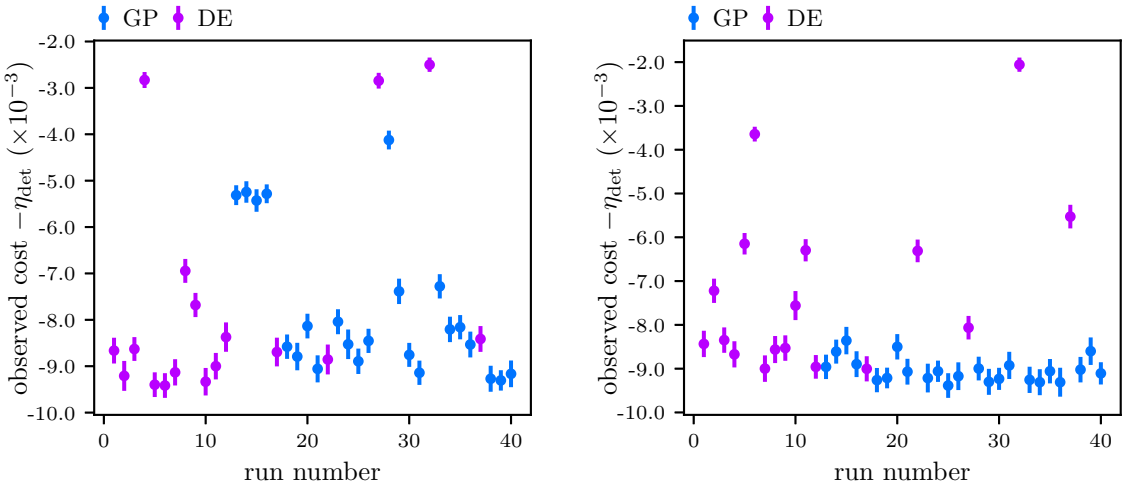
Figure 4.9.: Scanning the excitation power with either the RF drivers or the AWG. For better visibility, errorbars indicate three standard deviations. The bottom panel shows the relative different between the two curves. We find that the performance of the pumping when using the RF drivers or the AWG is indeed comparable. It is thus possible to use the waveform generated by the AWG as benchmark.

display in figure 4.7. Even greater insight can be obtained by projecting the cost landscape onto a two dimensional surface spanned by two parameters of choice as in figure 4.8, where the center was once again the best predicted parameters. The aforementioned feature of having more than one optimal value for the $\text{pump}_{1\to1}$ frequency could not be inferred by M-LOOP. It should be noted, that due to finite sampling and the inherent inability to properly visualize a four dimensional parameter space more features such as local or even global optima could remain hidden.

After having established that by using M-LOOP it is possible to find some local or global optimum for the excitation efficiency, we next set out to investigate how this optimum compares to the performance we can achieve when using the RF driver instead of the AWG. The figure or merit that determines the cost used for optimization is the single photon efficiency, η_{det}, which in turn is dependent on the excitation pulse. Therefore, we performed an excitation-power-scan, measuring η_{det} as a function of the excitation power[5], for both cases. The results in figure 4.9 imply that the AWG performs similarly well as the RF driver, motivating the use of the optimal waveforms obtained here for bench-marking later optimizations.

Please note that the optimal frequencies and amplitudes change over time and are dependent on the alignment of various optical elements. References to the optimal waveform in this section that will occur in the following sections always imply that

[5]the measured maximum of the excitation pulse on an oscilloscope



(a) $T_{\mathrm{pump}} \in [1.5\,\mu\mathrm{s}, 14.5\,\mu\mathrm{s}]$ **(b)** $T_{\mathrm{pump}} \in [14.5\,\mu\mathrm{s}, 39.5\,\mu\mathrm{s}]$

Figure 4.10.: Optimizations for varying the total pumping time.

the parameters have been re-optimized around the same time as the measurement they are compared to.

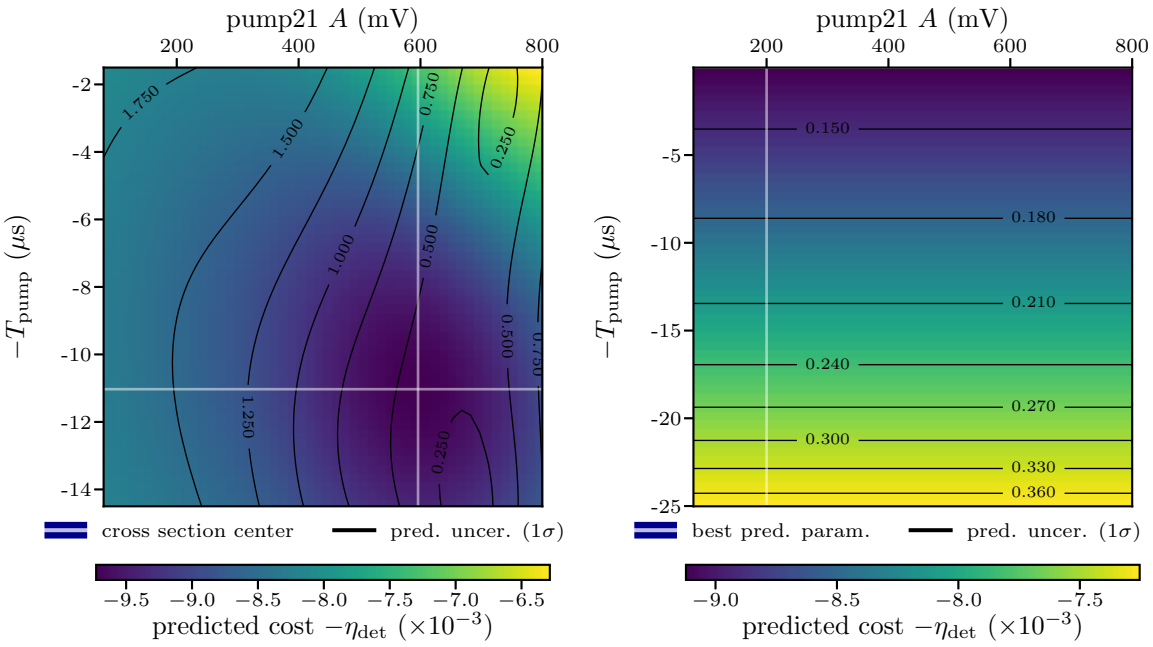
4.2.3b. Varying the Pumping Time

By varying the total time allocated to the pumping, T_{pump}, we wanted to investigate if η_{prep} could be increased by simply pumping longer and potentially with less optical power. We performed two optimizations (fig. 4.10) varying T_{pump} from 1.5 µs to 14.5 µs and from 14.5 µs to 39.5 µs, respectively. We fixed the frequencies of both pumping beams at the frequencies we had found for 2.3 µs and used a single square pulse for $\mathrm{pump}_{1\to1}$ and sequence of square pulses with $(\Delta T, \beta) = (200\,\mathrm{ns}, 80\%)$ for $\mathrm{pump}_{2\to1}$. For these measurements, we extended the period in which pumping could take place from 2.325 µs to 39.5 µs and increased the cooling time to 600 µs every 40 pumping-excitation tries. The resulting repetition rate, that can be estimated as

$$\frac{40 \text{ tries}}{40 \cdot T_{\mathrm{pumping}} + T_{\mathrm{cooling}}} = 18\,\mathrm{kHz} \tag{4.9}$$

was smaller than the repetition rate for the other pumping optimizations by almost a factor 10, yet still significantly higher than the rate for a direct measurement of η_{prep}.

Both optimizations found an optimum, yet the optimization with the smaller values of T_{pump} (fig. 4.10a) spent more time exploring other regions of parameter space. While the optimal amplitude for $\mathrm{pump}_{1\to1}$ lays around the value that had been found in the optimization with fixed pumping time, the landscapes for the amplitude of $\mathrm{pump}_{2\to1}$ and T_{pump} differ notably. The landscape for shorter available



(a) $T_{\text{pump}} \in [1.5\,\mu\text{s}, 14.5\,\mu\text{s}]$ **(b)** $T_{\text{pump}} \in [14.5\,\mu\text{s}, 39.5\,\mu\text{s}]$

Figure 4.11.: Two dimensional cross-section for the predicted cost as a function of total pumping time T_{pump} and amplitude of $\text{pump}_{2\to1}$. T_{pump} is measured from the back of the pumping window, hence the negative values. The landscape in (a) appears to be richer than the landscape in (b). Hence, as both optimization had the same number of iterations, the larger predicted uncertainty in (a) is to be expected.

pumping times (fig. 4.11a) has clear optimum at the value of A that corresponds to the maximum efficiency of the AOM and thus the maximum available power in the $\text{pump}_{2\to1}$ beam path. For longer available T_{pump} it appears as if the amplitude for $\text{pump}_{2\to1}$ is less relevant but T_{pump} should be chosen as short as possible ($T_{\text{pump}} = 14.5\,\mu\text{s}$) to yield the optimal cost.

Performing an excitation power scan (fig. 4.12) found that indeed this predicted minima corresponded to higher η_{det} than for the case of fixed pumping time. The relative excess (fig. 4.12, bottom panel) grows as the excitation power is increased, indicating that both the population of the target state and $|1, \pm1\rangle$ are increased. For $P_{\text{exc}} \approx 80\% \cdot \max_{\eta_{\text{det}}}(P_{\text{exc}})$ the largest excess is observed for the optimum found by the optimization with shorter available pumping times.

In fact, when we measure η_{prep} directly (fig. 4.13) we see that the best state preparation efficiency can be reached for $T_{\text{pump}} = 11.0\,\mu\text{s}$ as for even longer T_{pump} the population $|1, \pm1\rangle$ is further increased. However, it should be noted that the large errors on the population of $|1, \pm1\rangle$ make a definite decision challenging. Each data-point for $\sum \Pr(|1, \pm1\rangle)$ corresponds 60 min of data acquisition, longer

measurements times were not feasible. Figure 4.13 also illustrates the importance of adding 2 µs to T_{pump} after cooling.

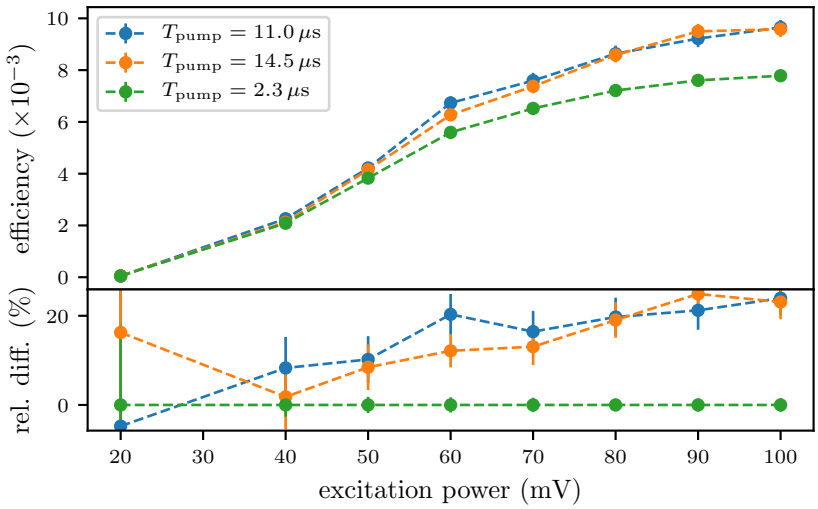


Figure 4.12.: Excitation power scan for varying the pumping time. Each data-set corresponds to the predicted best parameters of the optimization where T_{pump} was fixed at 2.325 µs (sec. 4.2.3a, green) or could be varied from 1.5 µs to 14.5 µs (blue) or from 14.5 µs to 39.5 µs (orange), respectively.

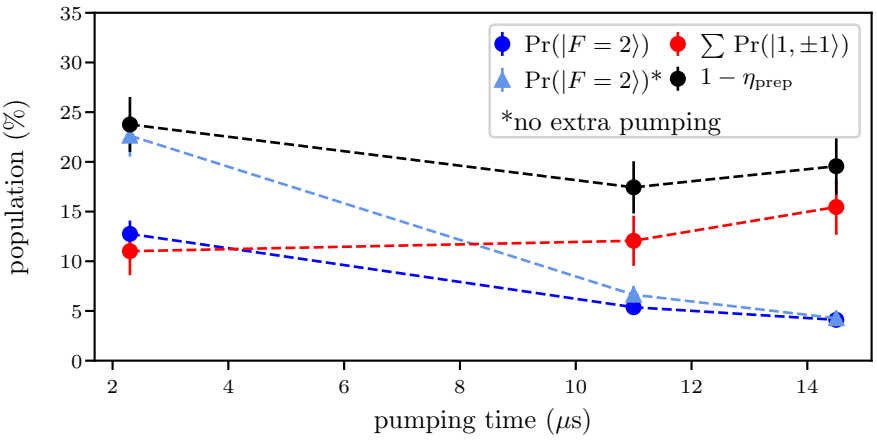
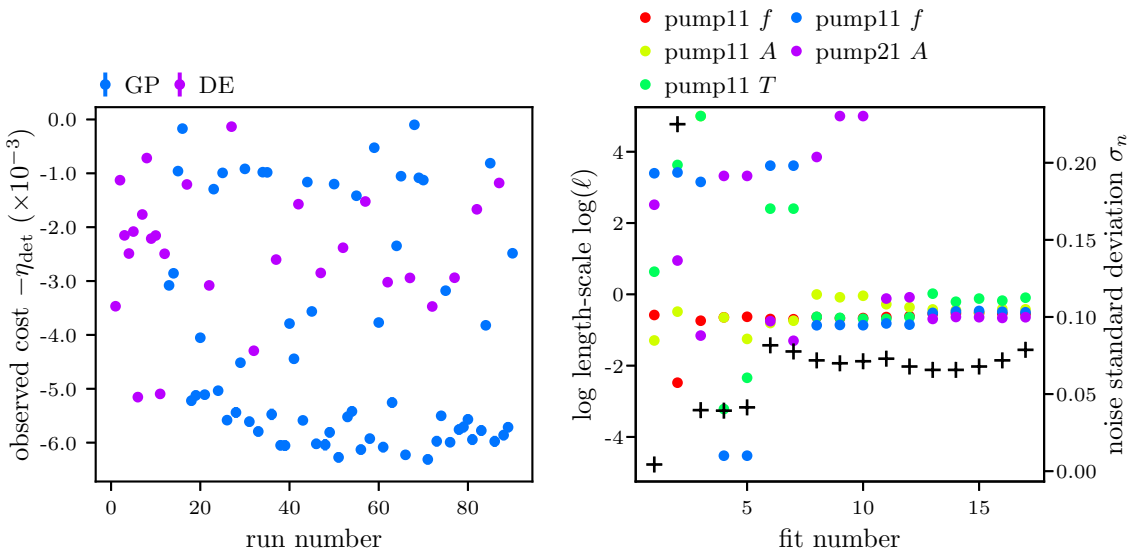
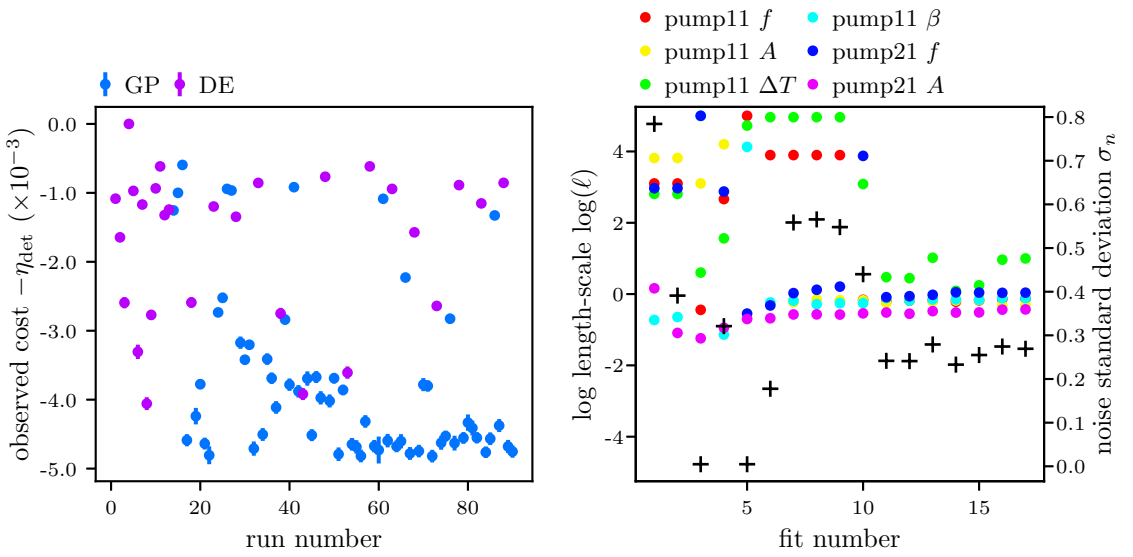


Figure 4.13.: Population of $|2,0\rangle$ (blue), the sum of $|1,\pm1\rangle$ (red) and $|1,0\rangle \equiv \eta_{\text{prep}}$ (black) with optimized parameters for $T_{\text{pump}} = \{2.3, 11.0, 14.5\}$µs. The optimal state preparation efficiency seems to be achieved for $T_{\text{pump}} = 11.0$ µs, however the large uncertainties make a definite decision challenging. Triangles denote the data for $F = 2$ where no extra puming time of 2 µs was added.



(a) Rectangular pulse with variable length for the $\mathrm{pump}_{1\to1}$.



(b) Sequence of square pulses with variable period T and duty cycle β for the $\mathrm{pump}_{1\to1}$.

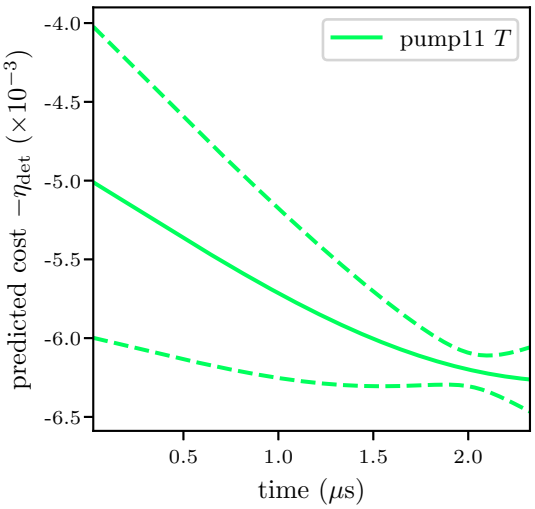
Figure 4.14.: Obtained cost per run (left) and hyperparameters per fit (right) for the pumping schemes laid out in section 4.2.3c. In (a) the cost seems to jump around run 75 which impairs the quality of the fit which manifests in a sudden jump of the length scale of the $\mathrm{pump}_{1\to1}$ amplitude. Solutions to this problem are offered in the text.

4.2.3c. Temporarily Turing Off pump$_{1\to1}$

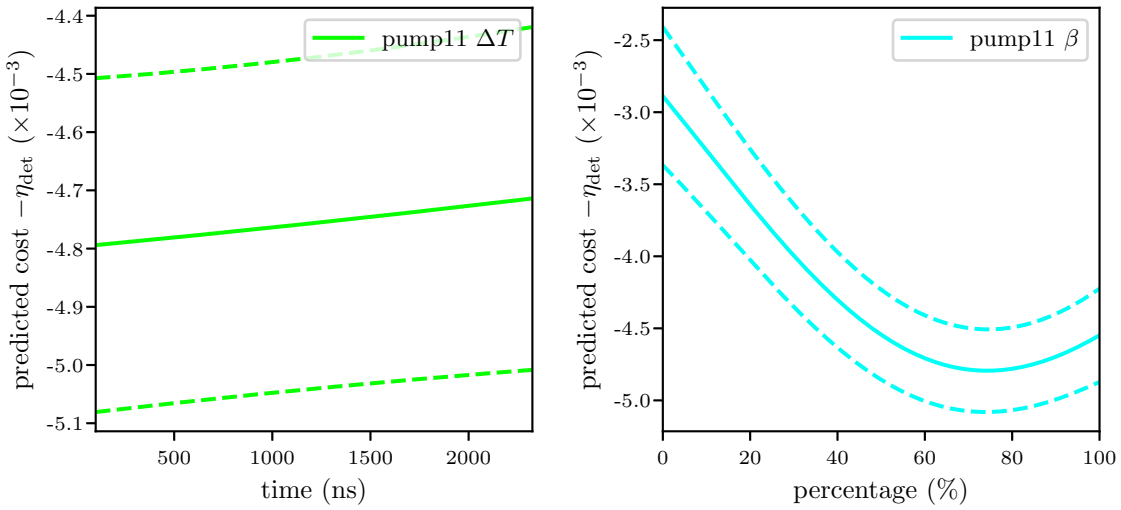
As we have seen the atom ends up in either $F = 2$ substate in a not insignificant number of pumping attempts. Redistributing occupation probabilities of these substates amongst the $F = 1$ substates should improve η_{prep}. To this end we investigated two pumping schemes: one where the pump$_{1\to1}$ would be turned off prior to the pump$_{2\to1}$ and one where the pump$_{1\to1}$ would consist of a sequence of square pulses, similar to the pump$_{2\to1}$. Both of these pumping schemes emerged as particularly useful in the simulations that only considered resonant excitations. In the simulations, the former scheme would allow moving the remaining population in $F = 2$ to $F = 1$ without the chance of subsequent re-excitation while the latter would almost step-like lower the population in $F = 2$.

In both schemes, amplitude and frequency of both pumping beams would be the same for the entire waveform, with one or two additional parameters describing the envelop of the pump$_{1\to1}$. In one case (fig. 4.14a) the additional parameter is the length of the pump$_{1\to1}$ pulse which can be varied up to $T_{\text{pump}} = 2.325\,\mu\text{s}$. For the other case the pump$_{1\to1}$ waveform consists of a series of square pulses, characterized by a period length T and a duty-cycle β.

Optimization with both schemes were possible, with all parameters being relevant to the problem. The measurements suggest that neither reducing the length of the square pulse nor sequencing the pump$_{1\to1}$ was favourable. While the former optimization predicted the best pulse length T to be $T = T_{\text{pump}}$ (fig. 4.15a), the latter predicted the optimum at $(\Delta T, \beta) \approx (25\,\text{ns}, 75\%)$ (fig. 4.15b). Taking the technical limitations set by the AOM into account the finite switching time of the AOM, this also points towards an optimum that corresponds to pump$_{1\to1}$ always being present.



(a) $\mathrm{pump}_{1\to1}$ square pulse with variable length. M-LOOP predicts that the optimal pumping efficiency can be achieved if the $\mathrm{pump}_{1\to1}$ beam is present during the entire pumping sequence.



(b) Sequenced $\mathrm{pump}_{1\to1}$ with varying period length (left) and duty-cycle (right). While the predicted optimum is achieved for a duty-cycle $\beta \sim 75\%$, having the $\mathrm{pump}_{1\to1}$ always present (duty-cycle 100%) is well within the 1σ error bounds, in particular if one also takes the sequence length and the technical limitation set by the AOM into account.

Figure 4.15.: Cross-sections for selected parameters of the two optimizations in section 4.2.3c.

4.2.3d. Varying the Spectral Profile for $\mathrm{pump}_{2\to1}$

The initial tests (sec. 4.2.3a) seemed to indicate that there existed multiple local optima or at least a wide range of suitable frequencies for the $\mathrm{pump}_{2\to1}$. Using the AWG, it is easy to vary the frequency throughout the pumping or even use multiple

frequencies at once. In particular, we looked at three different schemes implementing a wider spectral profile for $\text{pump}_{2\to1}$:

scheme 1: frequency modulation (FM) with single-tone modulation

scheme 2: feeding multiple frequencies at once into the AOM

Frequency modulation is a technique most notably known from FM radio broadcasting. Here we will restrict ourselves to single-tone modulation,

$$A(t) = A_c \cdot \cos\left[2\pi f_c t + h \cdot \sin(2\pi f_m t)\right], \tag{4.10}$$

where the carrier amplitude A_c and frequency f_c, and the modulation index h and frequency f_m fully characterize the time dependence of the signal. h and f_m are related to the peak deviation $\Delta f := h\dot{f}_m$, the maximum difference between the instantaneous frequency and f_c. To understand the spectral properties of such signals we first express (4.10) in terms of the Bessel functions of the first kind J_k as

$$A(t) = \sum_{k\in\mathbb{Z}} J_k(h) \cdot \cos\left[2\pi(f_c + k f_m)t\right].$$

We see that increasing h raises the number of side-bands that carry a non-negligible amount of power. Finite size effects further spread out the power, also over the intermediate regions as shown in figure 4.16.

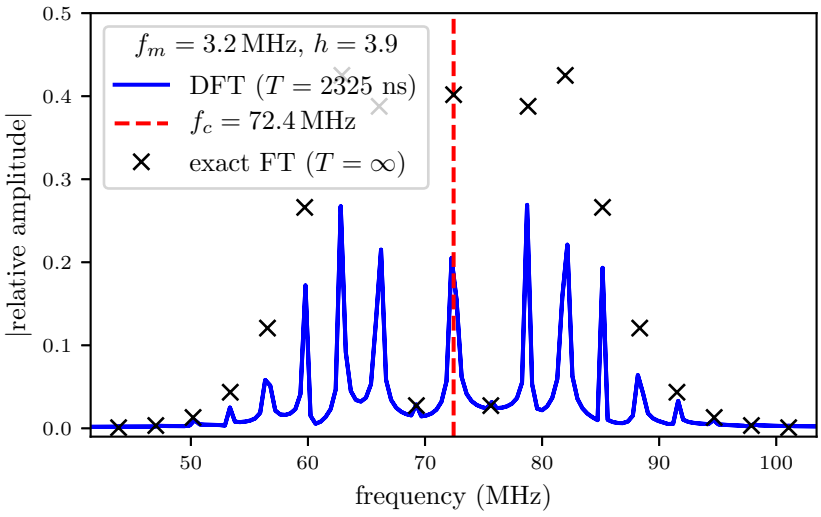


Figure 4.16.: Spectrum of $\text{pump}_{2\to1}$ for the best predicted parameters. The discrete Fourier transform (DFT, blue) is obtained finite length pulse at a sampling rate of 1.25 GHz while the components of the exact Fourier transform (FT, black crosses) are found by evaluating the Bessel functions of first kind J_k at the modulation index h (k here is $(f_c - f)/f_m$).

The results of the optimization found similar values for the frequency and amplitude of $\text{pump}_{1\to1}$ the corresponding optimization from sec. 4.2.3a. Likewise, the optimal carrier frequency and amplitude of $\text{pump}_{2\to1}$ do not differ significantly from the values for f and A in sec. 4.2.3a. The predicted minimum for the cost was

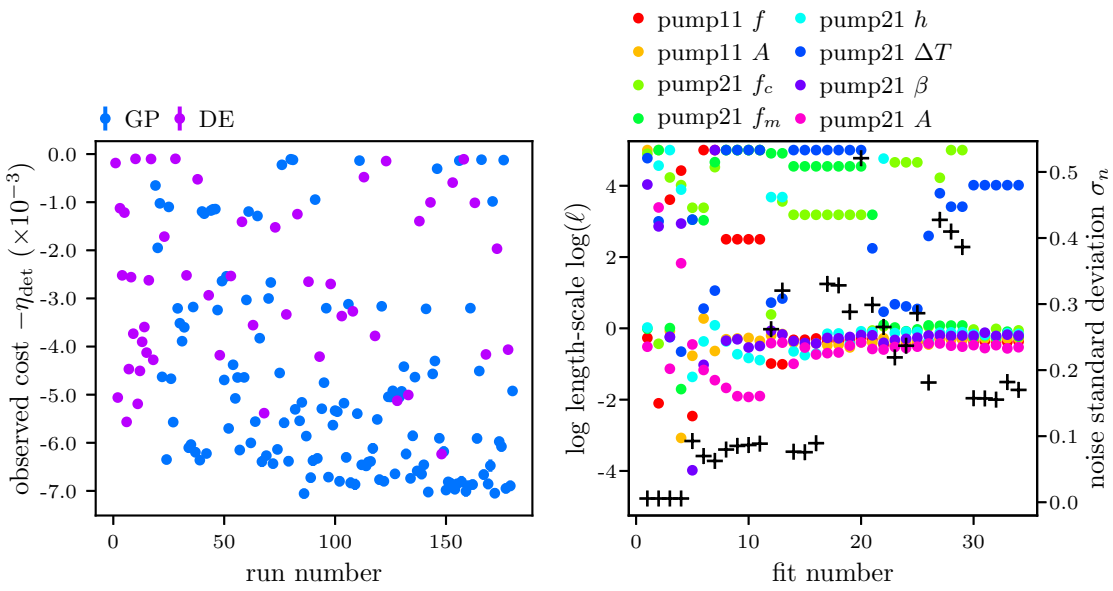


Figure 4.17.: Optimization with single-tone frequency modulated $\text{pump}_{2\to1}$.

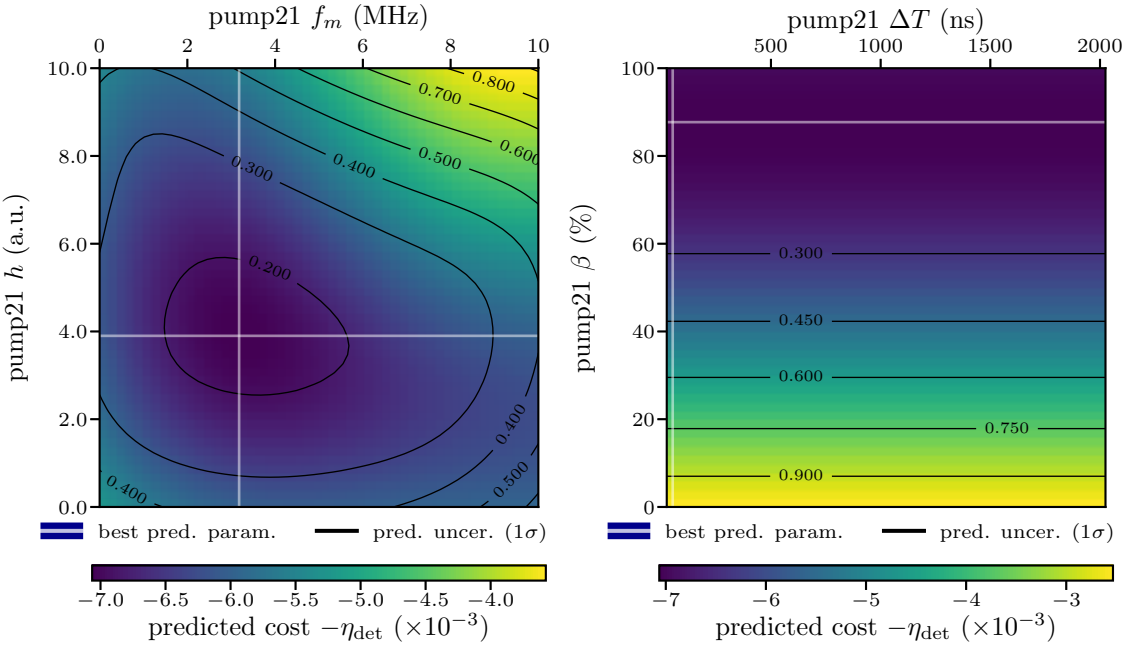


Figure 4.18.: Two-dimensional cost landscapes of the f_m versus h (left) and ΔT versus β (right).

found for modulation index $h = 3.9$ and frequency $f_m = 3.9\,\mathrm{MHz}$ which leads to a broad spectrum. In particular, M-LOOP predicts that this optimum is superior to the quasi-monochromatic extremum[6] at $h = 0.0$ as evident in figure 4.18. We should however take into account that thought the short pulses (160 ns in lab 2) the spectrum of $\mathrm{pump}_{2\to1}$ is already broadened.

Using frequency modulation would in theory also be possible with the RF drivers by modulating U_{VCO} (fig. 4.2a) as the required signal still has a single instantaneous frequency for any given time. On the contrary, the AWG can also produce a signal that consists of multiple frequencies components according to

$$A(t) = \frac{1}{N} \cdot \sum_{k=1}^{N} A_k \cdot \cos(2\pi f_k t + \phi_k), \tag{4.11}$$

where $A_k > 0$ and ϕ_k are the amplitude and phase associated with each frequency f_k. In a first optimization using this scheme we considered two distinct frequency components and additionally sequenced the square pulses (sequence length ΔT and duty-cycle β). We thus fixed $\phi_1 = 0$ and $\phi_2 = \phi_1 + \Delta\phi$. Finally, we let the optimizer also control frequency and amplitude for $\mathrm{pump}_{1\to1}$.

Similar as before, the optimizer found the same values for $\mathrm{pump}_{1\to1}$ as in sec. 4.2.3a. Unfortunately, bad runs during the optimization prevented accurate predictions during the optimization. The post-processing of the data (fig. 4.20)

[6] evidently finite sampling and pulse width broaden the spectrum

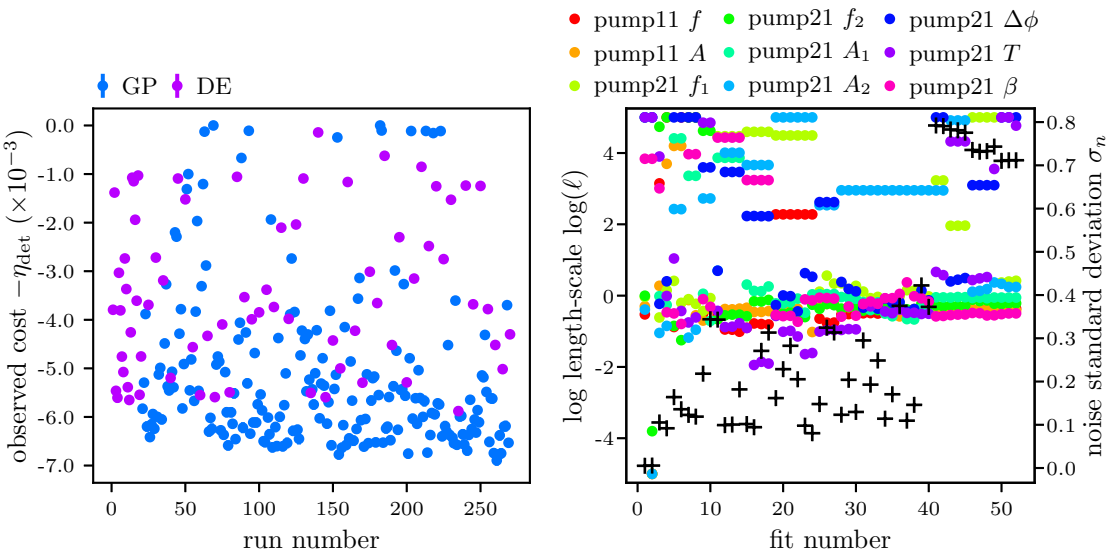


Figure 4.19.: Optimization using two overlaid frequencies for the $\mathrm{pump}_{2\to1}$.

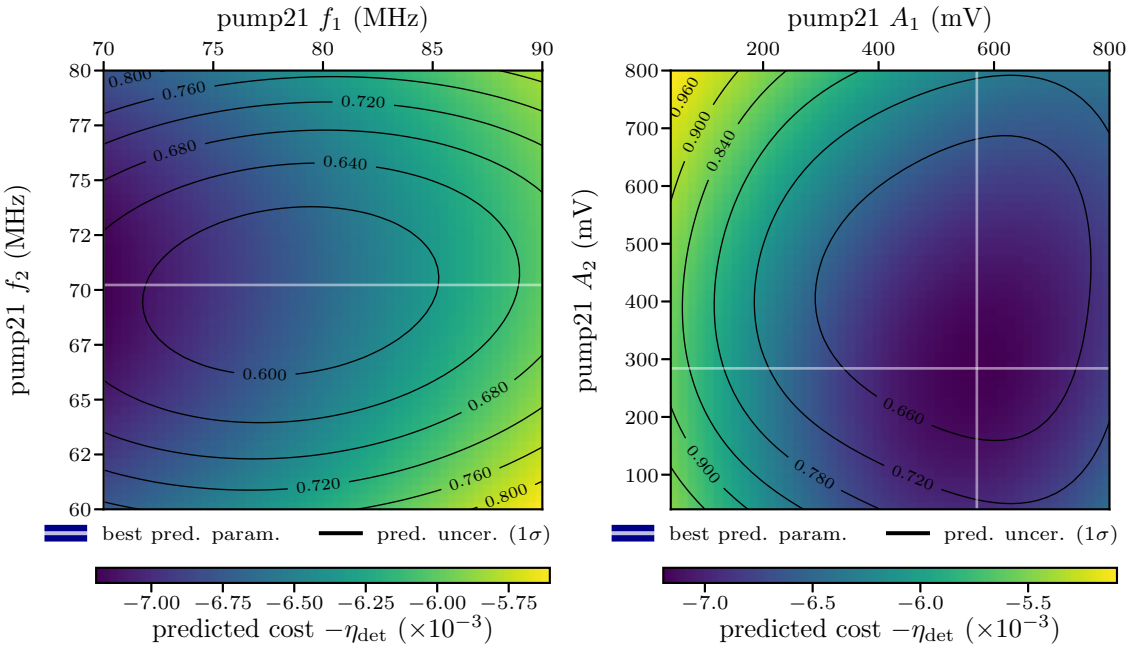


Figure 4.20.: Results for the optimization with two ovelaid frequencies for $\mathrm{pump}_{2\to1}$: Two-dimensional cost landscapes of the two frequencies (left) and amplitudes (right) of $\mathrm{pump}_{1\to1}$.

reveals that the optimizer preferred a waveform with a single frequency over the overlaying two frequencies.

Performing an excitation power scan with the two schemes discussed in this section (fig. 4.21) indeed show a small excess in η_{prep} for the pumping scheme where $\mathrm{pump}_{2\to1}$ is modulated. While a direct measurement of the state preparation was possible due to technical problems and time constraints, we were able to verify that the frequency modulation helped to reduce population of $F = 2$ (Table 4.1). However, the measurement did not take extra pumping time into account.

	reference	with overlaid frequency	frequency modulation
$\Pr(\lvert F = 2\rangle)$	22.24(96) %	21.72(92) %	15.41(71) %

Table 4.1.: Remaining population in $F = 2$ after 2.325 µs of pumping with pumping schemes where the spectral profile of $\mathrm{pump}_{2\to1}$ was modified (sec. 4.2.3d). The reference was the pumping sequence from sec. 4.2.3a. The best value can be achieved for single tone frequency modulation (right).

.

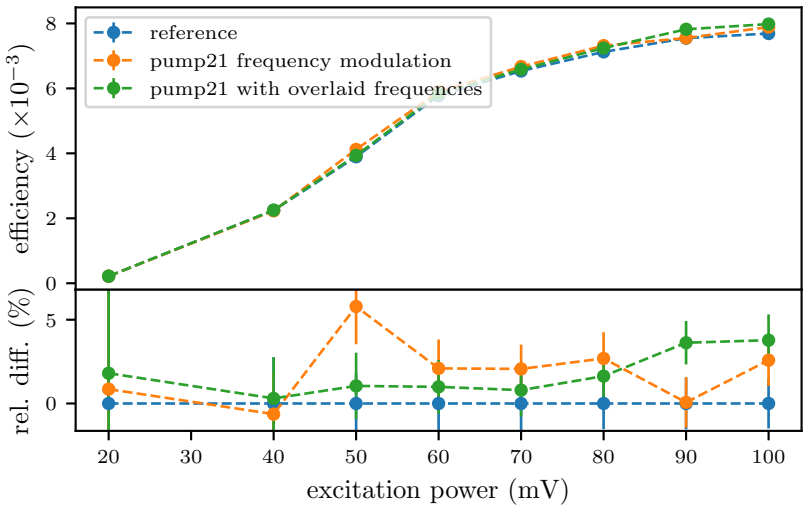


Figure 4.21.: Excitation power scan for the optimizations with a more complex spectral profile for $\text{pump}_{2\to1}$ as described in sec. 4.2.3d.

4.2.3e. Modulating the $\text{pump}_{1\to1}$ amplitude

As we have seen in section 2.4.1c off-resonant Rabi oscillations can be a factor that substantially reduces the pumping efficiency. While the complete description of the time evolution is complicated we can generally observe that lower optical powers decrease the amplitude of the off-resonant Rabi oscillations. Since always using low powers would drastically increase the required pumping time, we wanted to investigate if a non-uniform Rabi frequency would be able to improve η_{prep}. In practise, we implemented two different schemes of modifying the amplitude by interpolating between a finite number of points, either linearly or by step-wise changing the amplitude.

Our minimal approach used a linear interpolation between two points $n \cdot T$ and $n \cdot (T + \beta \cdot \Delta T)$ ($n \in \mathbb{N}$), where the sequence-length ΔT and the duty-cycle β can be varied by M-LOOP. Additional parameters are the frequencies of $\text{pump}_{1\to1}$ and $\text{pump}_{2\to1}$ as well as the amount of time T_{end} by which the waveform should be offset from the end of the pumping sequence. Post-processing indicate that 90 runs might have just been sufficient for optimization.

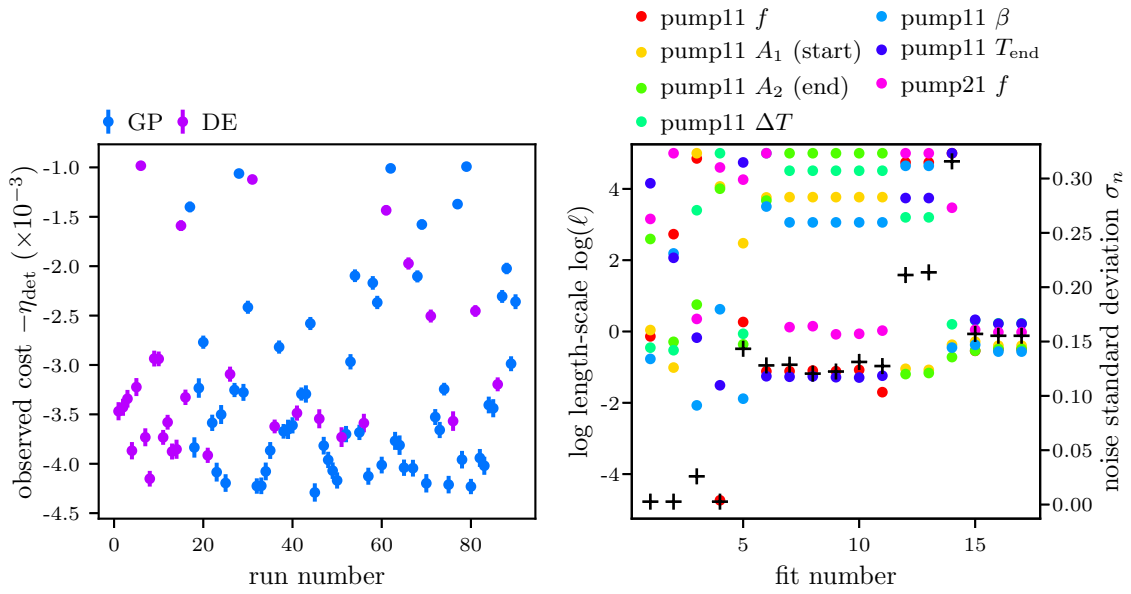


Figure 4.22.: Optimization of the amplitude modulation of $\mathrm{pump}_{1\to1}$ with linear interpolation, two variable time points T and $T + \beta \cdot \Delta T$ that repeat after the sequence-length ΔT. Additionally, the waveform is offset from the end by δt (post-time). The remaining parameters are the frequencies of $\mathrm{pump}_{1\to1}$ and $\mathrm{pump}_{2\to1}$.

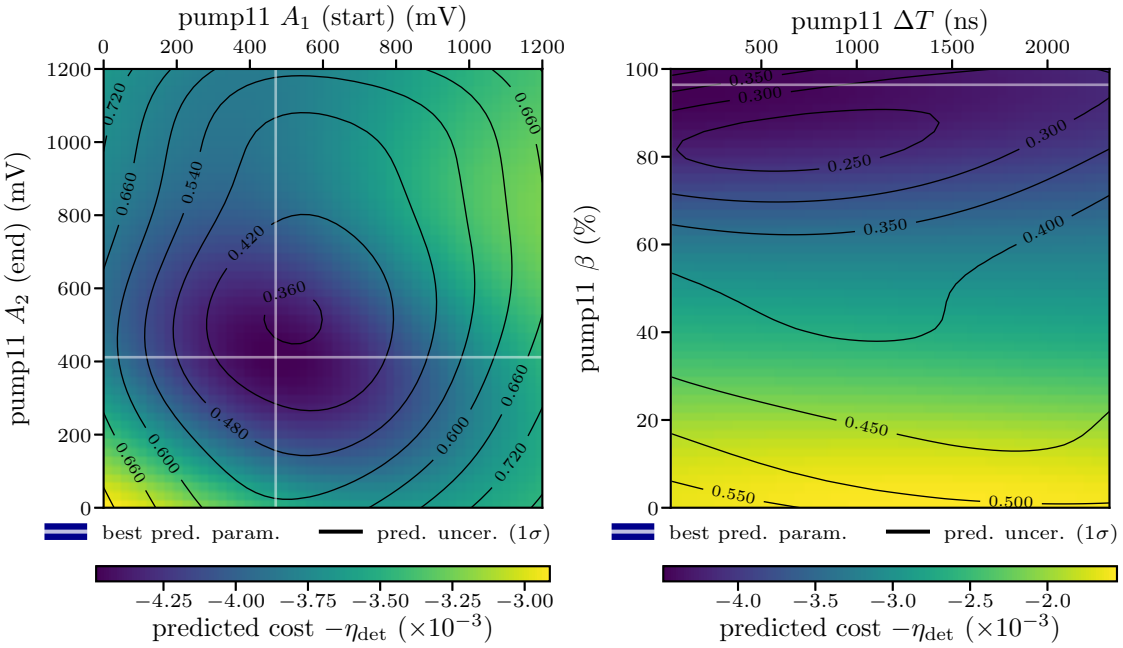


Figure 4.23.: Optimization of the amplitude modulation of $\mathrm{pump}_{1\to1}$ with linear interpolation, two variable time points T and $T + \beta \cdot \Delta T$ that repeats after the sequence-length ΔT. Additionally, the waveform is offset from the end by δt (post-time). The remaining parameters are the frequencies of $\mathrm{pump}_{1\to1}$ and $\mathrm{pump}_{2\to1}$.

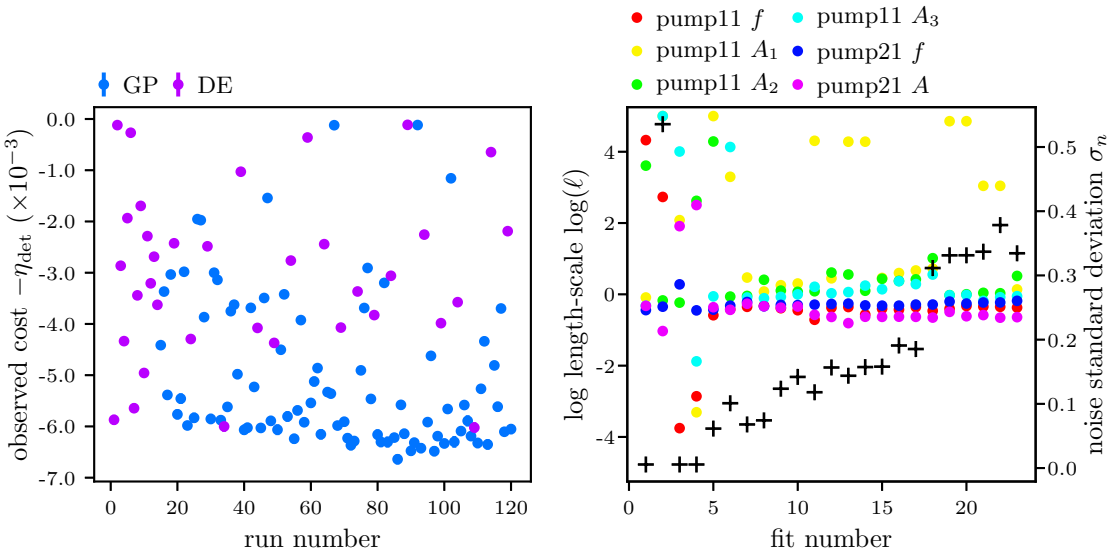


Figure 4.24.: Optimization of the step-wise amplitude modulation of $\mathrm{pump}_{1\to1}$, with three equidistant steps at $T_k = \{0.33, 0.67, 1.00\} \cdot T_{\mathrm{pump}}$. The three A_k are the parameters with the largest length scale in this optimization.

The short sequence length of 25 ns together with the large duty-cycle of 96.4% mean $\mathrm{pump}_{1\to1}$ is effectively always on. The fact that the optimum for the two amplitudes are almost identical ($A_1, A_2 = 392\,\mathrm{mV}, 343\,\mathrm{mV}$) indicates that indeed the preferred pulse shape is a simple square pulse. Interestingly, the landscape for T_{end} has a preference for small values but is overall rather shallow which raises the question if the overall pumping time could be reduced.

We further investigated the step-wise amplitude modulation

$$A(t) = \left(\sum_{k=1}^{N} \chi(t; T_{k-1}, T_k) \cdot A_k \right) \cdot \cos(2\pi f) \tag{4.12}$$

with three equidistant time steps between $T_0 = 0$ and $T_3 = T_{\mathrm{pump}}$, also varying the frequency of $\mathrm{pump}_{1\to1}$, and amplitude and frequency of $\mathrm{pump}_{2\to1}$. The three amplitudes for $\mathrm{pump}_{1\to1}$ had the largest length-scales. Notably, the landscape for the amplitude of the middle segment (A_2) was the flattest (fig. 4.25) pointing at the fact that the largest change in the population of $|1, \pm1\rangle$ occurs at the beginning of the pumping sequence.

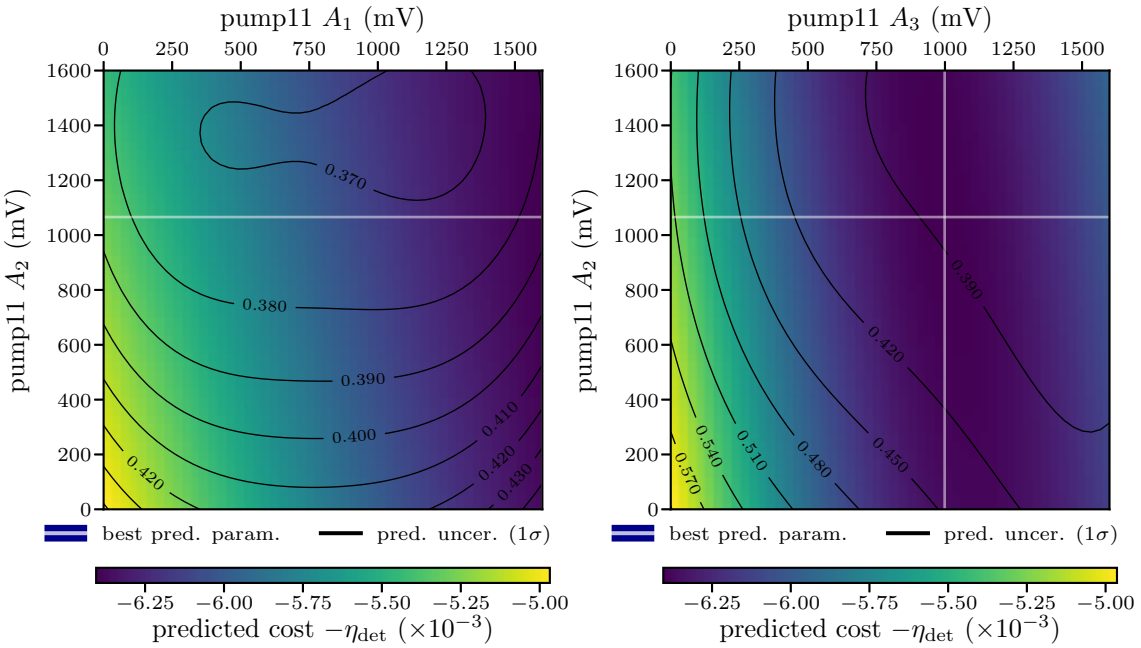
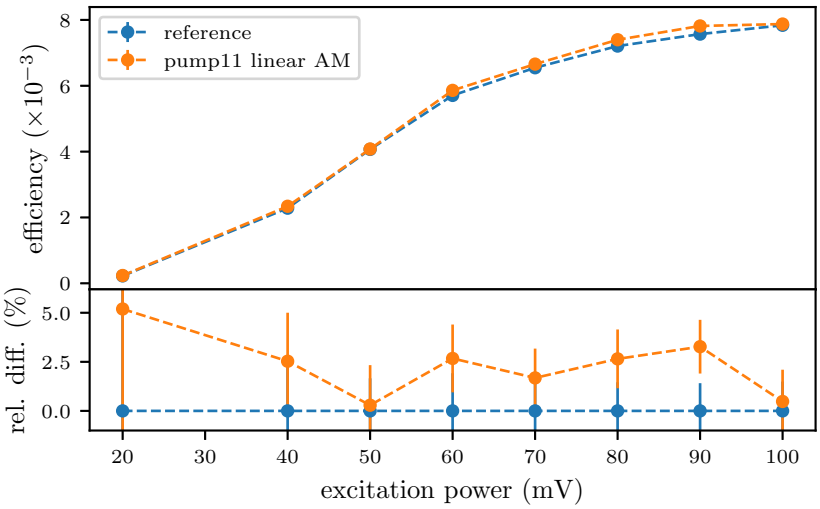
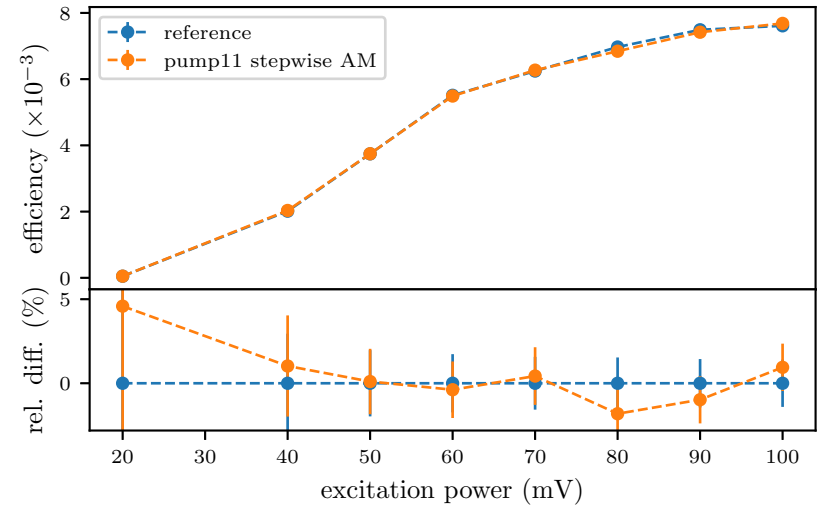


Figure 4.25.: Optimization of the step-wise amplitude modulation of $\mathrm{pump}_{1\to1}$: The amplitude of the middle segment (A_2, y-axis) is less relevant than the amplitude at the beginning (A_1) and end (A_3).

Excitation power scans reveal that neither scheme resulted in a pulse-form that could significantly increase η_{prep}. Using more points and thus more complex pulse-shapes might be able to produce better η_{prep}, yet due to time constraints this was not attempted.



(a) linear interpolation between two amplitudes



(b) step-wise interpolation between three amplitudes

Figure 4.26.: Excitation scans comparing the best predicted parameters for amplitude modulation of $\text{pump}_{1\to1}$ with linear (a) 4.22 or step-wise interpolation (b) 4.24.

4.2.3f. Varying the Spectral Profile for $\text{pump}_{1\to1}$

Similar to subsection 4.2.3d we implemented different schemes of varying the frequency, now for $\text{pump}_{1\to1}$. First, we investigated single-tone frequency modulation. During the optimization (fig. 4.27) in 9 out of 200 runs trapping an atom was not possible. The M-LOOP API unfortunately does not offer a straight-forward way of excluding those runs during the optimization, therefore the quality of the hyper-parameter fits is seriously impaired. Post-processing revealed lower noise-level and length-scales, reducing the best predicted cost and uncertainty from $4.97 \pm 0.68‰$ to $4.67 \pm 0.28‰$.

The optimization revealed similar landscaped as before for amplitude and frequency of $\text{pump}_{2\to1}$. The predicted best modulation frequency for $\text{pump}_{1\to1}$ was $f_m = 0.0\,\text{MHz}$, showing a clear preference for a monochromatic approach to the problem.

The optimization where we overlaid frequencies for $\text{pump}_{1\to1}$ according to 4.11 (fig. 4.29), once again using two distinct frequency components, found an optimum where most optical power was put into the second frequency component (fig. 4.30), indicating again that using more than frequency for $\text{pump}_{1\to1}$ was not beneficial.

Lastly, we tried a waveform where the $\text{pump}_{1\to1}$ consisted of a sequence of square pulses with alternating frequency,

$$A(t) = \chi(t; 0, \beta_1 T_1) \cdot A_1 \cos(2\pi f_1 t) + \chi(t; T_1, T_1 + \beta_2 T_2) \cdot A_2 \cos(2\pi f_2 t) + ... \quad (4.13)$$

Here $\chi(t; T, T')$ is the characteristic function which is one for the time interval $[T, T']$ and zero otherwise. The signal is periodic with period length $\sum T_k$. A_k, f_k, and β_k are the respective amplitude, frequency, and duty-cycle as before. Using two different frequencies we once again found that waveforms with a single frequency and amplitude could correspond to the optimum (fig. 4.32).

The idea of using a richer spectrum for $\text{pump}_{1\to1}$ originated from observing multiple optima for the $\text{pump}_{1\to1}$ frequency when scanning the frequency for optimizations that had been done prior to this writing. However, using the AWG, we were not able to reproduce those observations. The optimizations presented in this section all point to the fact that indeed using multiple frequencies for the $\text{pump}_{1\to1}$, at least within the range of the AOMs, does not significantly improve η_{det}. A direct comparison due to a sudden technical failure that prevented us from producing an appropriate reference excitation power scan.

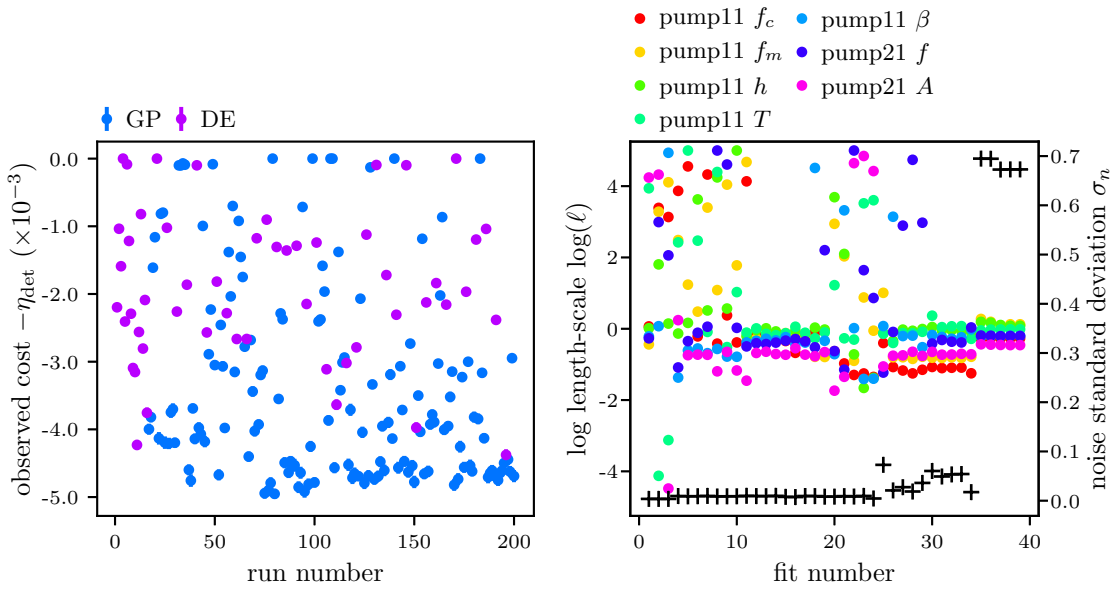


Figure 4.27.: Optimization using single-tone frequency modulation for the $\mathrm{pump}_{1\to1}$, with variable sequence length and duty-cycle. Amplitude and frequency of $\mathrm{pump}_{2\to1}$ completed the seven parameters.

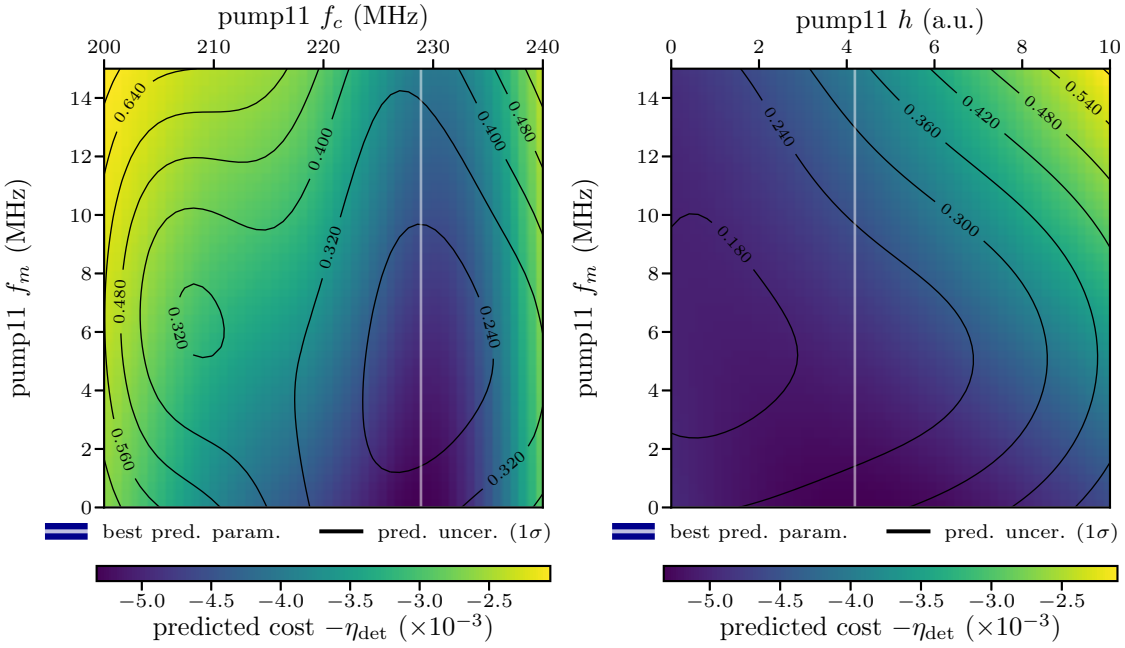


Figure 4.28.: 2D landscapes depicting the relation between f_m and f_c (left) and h (right). Please note that the best predicted f_m is 0.0 MHz. We see that the cost is also low for small h, regardless of f_m, further pointing towards the mono-chromatic extremum for $\mathrm{pump}_{1\to1}$.

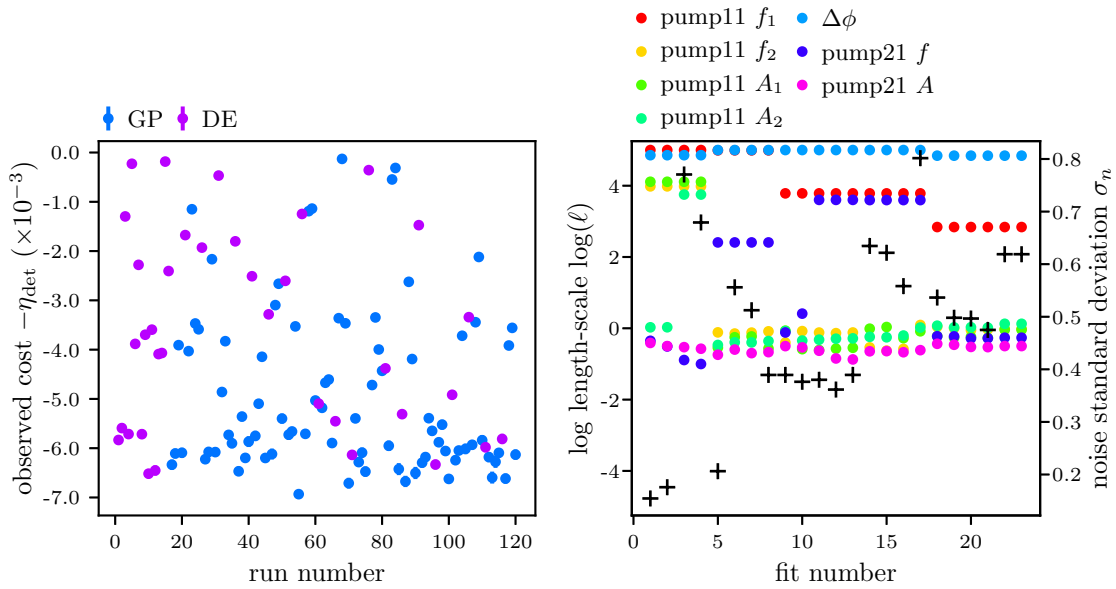


Figure 4.29.: Optimization with overlaying two distinct frequency components for $\mathrm{pump}_{1\to1}$.

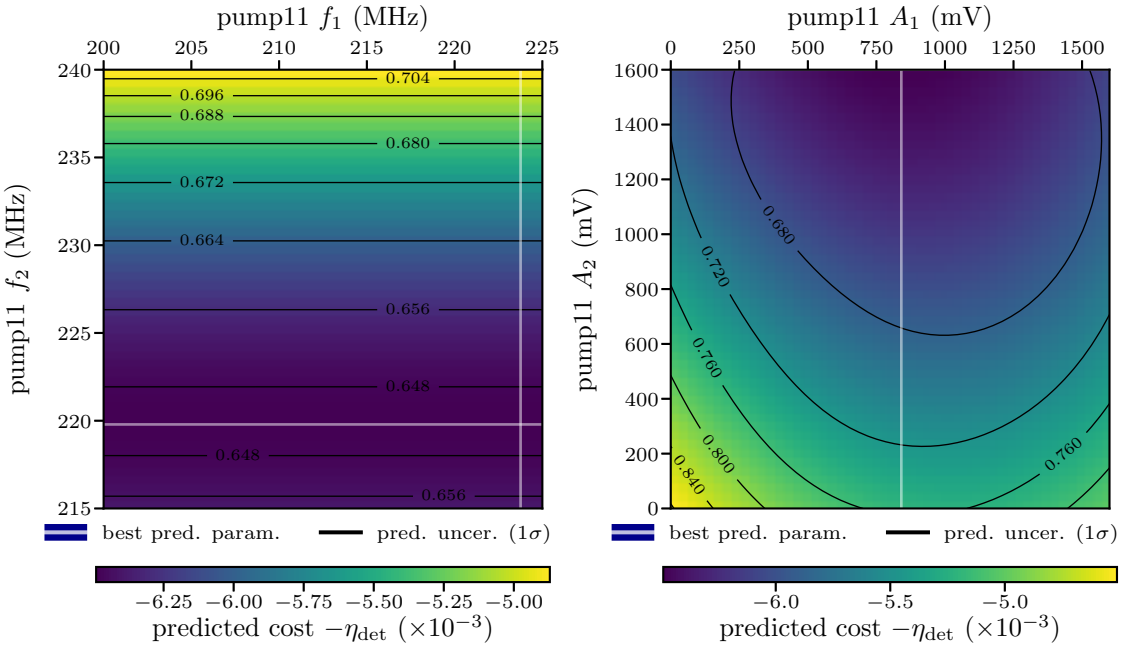


Figure 4.30.: Two dimensional cross-sections for the two frequencies (left) and amplitudes (right). Most optical power should be carried by one frequency (f_2) that is approximately equal to the frequency found in sec. 4.2.3a.

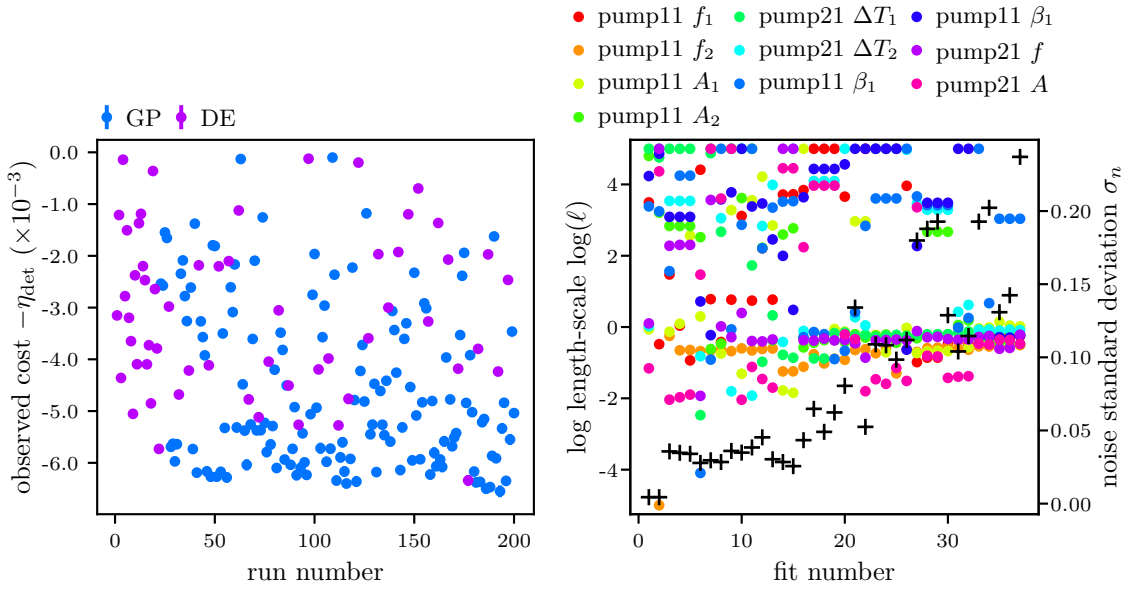


Figure 4.31.: Optimization with alternating frequencies for $\mathrm{pump}_{1\to1}$: cost per run (left) and hyper-parameters per fit(right).

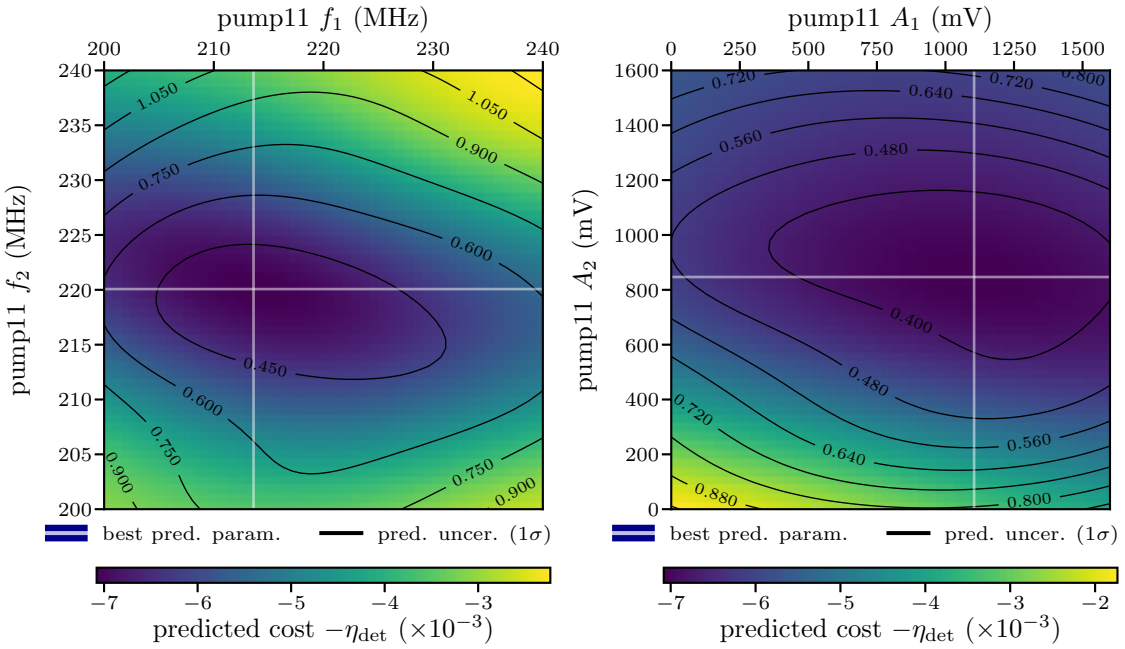


Figure 4.32.: Optimization with alternating frequencies for $\mathrm{pump}_{1\to1}$: Two dimensional cross-sections for the two frequencies (left) and amplitudes (right). Taking the uncertainty into account it appears possible that a waveform with a single frequency and amplitude corresponds to same cost as the optimum.

4.3. Optimization in the Readout Process

The current readout scheme via cycling-ionization has been subject to modifications and optimizations [73, 87]. In this section it will be investigating if and to what extend the greater flexibility offered by the AWG could make improvements of this aspect of the experiment possible.

4.3.1. Model for Determining the Contrast

After a successful excitation, the atom will spontaneously decay back to the ground level, generating the entangled state from 2.17. Collecting the photon along the quantization axis leaves the maximally entangled state (2.19)

$$|\Psi\rangle_{\mathrm{AP}} = \frac{1}{\sqrt{2}}\left(\left|\sigma^{+}\right\rangle_{\mathrm{P}}|1,-1\rangle_{\mathrm{A}} + \left|\sigma^{-}\right\rangle_{\mathrm{P}}|1,+1\rangle_{\mathrm{A}}\right),$$

where the Zeeman substate of the atom is entangled with the circular polarization of the photon. After appropriate unitary transformations and a polarization measurement on the photon, the atom is projected onto a superposition of $|1,\pm1\rangle$. A readout laser with polarization χ_{ro} will only interact with a certain superposition defining the dark and bright states $|\mathrm{D}\rangle_{\mathrm{ro}}$ (2.30b) and $|\mathrm{B}\rangle_{\mathrm{ro}}$ (2.30a). The corresponding contrast $\mathrm{K}(\chi_{\mathrm{ro}})$ then is the difference between the probabilities of successfully ionizing $|\mathrm{B}\rangle_{\mathrm{ro}}$ and $|\mathrm{D}\rangle_{\mathrm{ro}}$.

Experimentally we have access to four quantities that allow us to measure the contrast. First, we know the total number of events that herald the projection onto the dark or bright state, $n_{\mathrm{tries}}(|\mathrm{D/B}\rangle_{\mathrm{ro}})$. Second, we can observe the number of tries after which the atom was still present in the trap after the readout for each of the two cases, $n_{\mathrm{red}}(|\mathrm{D/B}\rangle_{\mathrm{ro}}) = n_{\mathrm{tries}}(|\mathrm{D/B}\rangle_{\mathrm{ro}}) - n_{\mathrm{ion}}(|\mathrm{D/B}\rangle_{\mathrm{ro}})$. Technically the atom can also be lost due to other reasons, but the fraction of these events is both small and constant, thus we will ignore it here. This allow to calculate the redetection probabilities

$$\mathrm{Pr}_{\mathrm{red}}(|\mathrm{D/B}\rangle_{\mathrm{ro}}) = 1 - \mathrm{Pr}_{\mathrm{ion}}(|\mathrm{D/B}\rangle_{\mathrm{ro}}) = \frac{n_{\mathrm{red}}(|\mathrm{D/B}\rangle_{\mathrm{ro}})}{n_{\mathrm{tries}}(|\mathrm{D/B}\rangle_{\mathrm{ro}})}. \tag{4.14}$$

Statistically, this will be treated as a sequence of n_{tries} Bernoulli trials, as each readout attempt is independent of all other attempts and yields a binary result.. The resulting PDF for the number of redetection events is a binomial distribution. The probability to have exactly k redetection events in n_{tries} is

$$\mathrm{Pr}_{\mathrm{binomial}}(X=k) = \binom{n_{\mathrm{tries}}}{k}(\mathrm{Pr}_{\mathrm{red}})^{k}(1-\mathrm{Pr}_{\mathrm{red}})^{n_{\mathrm{tries}}-k}, \tag{4.15}$$

The mean and variance of 4.15 are given by $\mathbb{E}[X] = n_{\mathrm{red}}$ and $\mathrm{Var[X]} = n_{\mathrm{red}}\cdot$

$n_{\mathrm{ion}}/n_{\mathrm{tries}}$. We can now define the figure of merit for the optimization of the readout process as

$$\mathcal{C}_{\mathrm{ro}} = 1 - |\mathrm{Pr}_{\mathrm{red}}(|\mathrm{B}\rangle_{\mathrm{ro}}) - \mathrm{Pr}_{\mathrm{red}}(|\mathrm{D}\rangle_{\mathrm{ro}})| = 1 - |\mathrm{K}(\chi_{\mathrm{ro}})| \tag{4.16a}$$

that has an uncertainty of

$$\mathcal{U}_{\mathrm{ro}} = \sqrt{\frac{\mathrm{Pr}_{\mathrm{red}}(|\mathrm{B}\rangle_{\mathrm{ro}}) \cdot \mathrm{Pr}_{\mathrm{ion}}(|\mathrm{B}\rangle_{\mathrm{ro}})}{n_{\mathrm{tries}}(|\mathrm{B}\rangle_{\mathrm{ro}})} + \frac{\mathrm{Pr}_{\mathrm{red}}(|\mathrm{D}\rangle_{\mathrm{ro}}) \cdot \mathrm{Pr}_{\mathrm{ion}}(|\mathrm{D}\rangle_{\mathrm{ro}})}{n_{\mathrm{tries}}(|\mathrm{D}\rangle_{\mathrm{ro}})}}. \tag{4.16b}$$

4.3.2. Experimental Setup for Evaluating the Readout Process

One output channel of the AWG will be connected the amplifier of the AOM that controls the path of the readout laser. Originally, this AOM operates at a frequency of $\sim$ 403 MHz which is slightly above the bandwidth of the M4i.6631. While earlier tests on the M4i.6622 showed that decent operation at frequencies above the nominal bandwidth were still possible to a lesser degree, it was desirable to modify all AOMs in the path to allow operation at roughly 390 MHz.

For this optimization the entire experimental sequence from sections 2.3 and 2.2.2 is necessary (fig. 4.33). After trapping and cooling the atom, the optical pumping (using the RF drivers) is initiated. Subsequently, we try to excite the atom using a short laser pulse. We collect the photon along the quantization axis and guide them

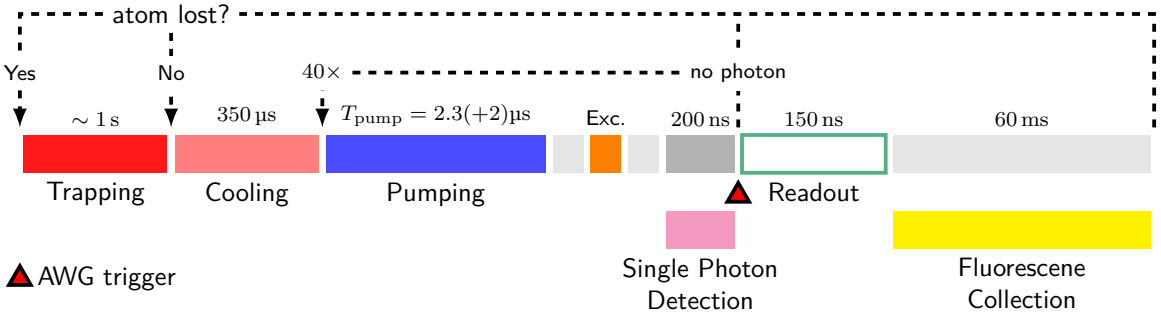


Figure 4.33.: Experimental sequence for determining the contrast: After trapping and cooling (red), the state preparation (blue) and excitation (orange) are performed. Projection of the atomic state is heralded by registering a photon during the single-photon-detection window (magenta). In this case, the AWG is triggered (red triangle) to replay the readout waveform. Simultaneously, cycling and ionization are turned on. Otherwise, preparation and excitation are repeated 40 times before additional cooling is necessary. After the readout window, we check if the atom was ionized by collecting fluorescence for 60 ms. In case the atom is lost, the sequence starts at the trapping.

through appropriate waveplates that transform the photons polarization state from $|\sigma^{\pm}\rangle_{\mathrm{P}}$ to $|V/H\rangle_{\mathrm{P}}$, onto a PBS. Behind each of the two output ports of the PBS we place a APD capable of detecting single photons. By detecting a photon on either of the APDs we have performed the aforementioned projection of the atomic state. We now apply the readout pulse, whose polarization is set such that the atomic state heralded by a detection event with APD 1 or 2 exactly corresponds to the dark and bright state, $|\mathrm{D}\rangle_{\mathrm{ro}}$ and $|\mathrm{B}\rangle_{\mathrm{ro}}$. The cycling and ionization laser are turned on during the entire readout window. Finally, after turning off readout, cycling, and ionization, we determine whether the atom was lost from the trap by checking the fluorescence counts after turning the dipole trap back on.

The repetition rate is strongly limited by the destructive ionization-based readout scheme. If the atom is still present in the trap, we can restart the experimental sequence at the cooling. However, if we loose the atom, we need to trap a new atom before we can enter this stage. The number of events where the dark and bright state is prepared is roughly equal, meaning that the atom is lost in about half of the cases, at least if the contrast we seek to optimize is good. The time scale for trapping alone is $\mathcal{O}(1\,\mathrm{s})$, meaning we only except a low event rate of few Hz, which greatly affects the statistical error (eq. 4.16b). Weighing those disadvantages against the necessity of generating a sufficient amount of data for the optimizer, we decided on measurement duration of several minutes which would leave us with a few hundred events for each data point used in the optimization.

4.3.3. Measurements for Optimizing the Readout Process

4.3.3a. Rectangular Readout Pulse

In similar fashion to the optimization of the pumping process, the first optimization of the pumping process sought to replicate the waveform that is currently used in the experiment. To this end, we generated a rectangular pulse defined by four free parameters: its frequency f, amplitude A, length ΔT and some adjustable offset T_{start} from the beginning of the readout window.

The optimization (fig. 4.34) predicts the best achievable cost at $1.3 \pm 7.4\%$ (contrast 98.7%). The large uncertainty mostly stems from the noise whose fit converges to $\sigma_n = 5.2\%$. Of the parameters, whose cross-sections are displayed in figure 4.35, the offset that is added before the square pulse is irrelevant to the problem. This confirms our intuition for the readout process: Both the ionization and cycling laser fields are approximately uniform during the readout process and transitions out of the $F = 1$ level can only happen through the readout laser. Therefore it should indeed not matter when we turn on the readout laser field, as long as it does not affect the overall overlap between the three fields. This also illustrates M-LOOP's ability to differentiate between relevant and irrelevant parameters. The optimization, in particular the best predicted pulse length of

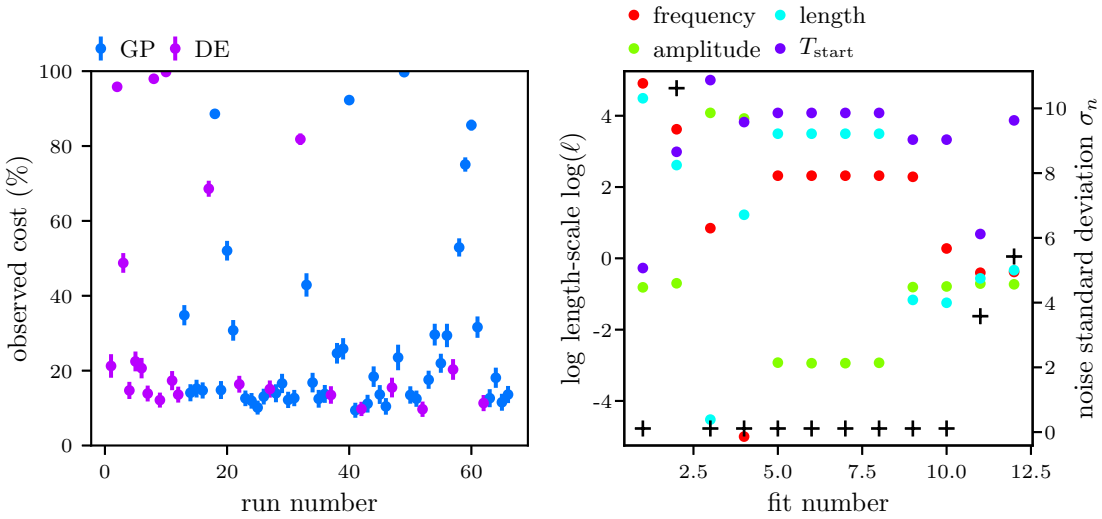


Figure 4.34.: Optimization of the readout pulse using four free parameters. The uncertainty displayed here corresponds to 1σ, not 3σ as in sec. 4.2.

143.6 ns, connect well to the results presented in [87].

While maximizing the contrast is the objective of the optimization and dictates how M-LOOP selects new parameters, the data collected also allows us to make predictions about the ionization probabilities of dark and bright state. This helps us interpreting the results of the optimization. We create a new GP from the parameters that M-LOOP has tested during the optimization and the two sets of redetection probabilities and their respective errors. The hyper-parameters are adjusted using the L-BGFS-B algorithm as before. The GP predicts (fig. 4.36) that the ionization probabilities both states in the beginning rise approximately linearly as a function of pulse length, indicating a constant ionization rate. As ionization of the dark state requires off-resonant excitations via the $F' = 2$ level, whose detuning of $\delta = 816.656(30)\,\text{MHz}$ [64] is much larger than the resonant Rabi frequency, we expect that the rate and thus the slope is much larger for the bright state. While the ionization rate of the bright state decreases as the population of the bright state is depleted, it stays almost constant for the dark state, whose population does not change as much. The GP for the dark state is significantly more certain about its prediction, as the ionization probability for the dark state appears rather insensitive to changes in the parameters.

Examining how the cost depends on the pulse amplitude A we consider the following: As the detuning between $F' = 1$ and $F' = 2$ in $5^2P_{1/2}$ is large, off-resonant excitation from the dark state only increase slowly as a function of the RF amplitude. In contrast, the ionization probability of the bright state rises

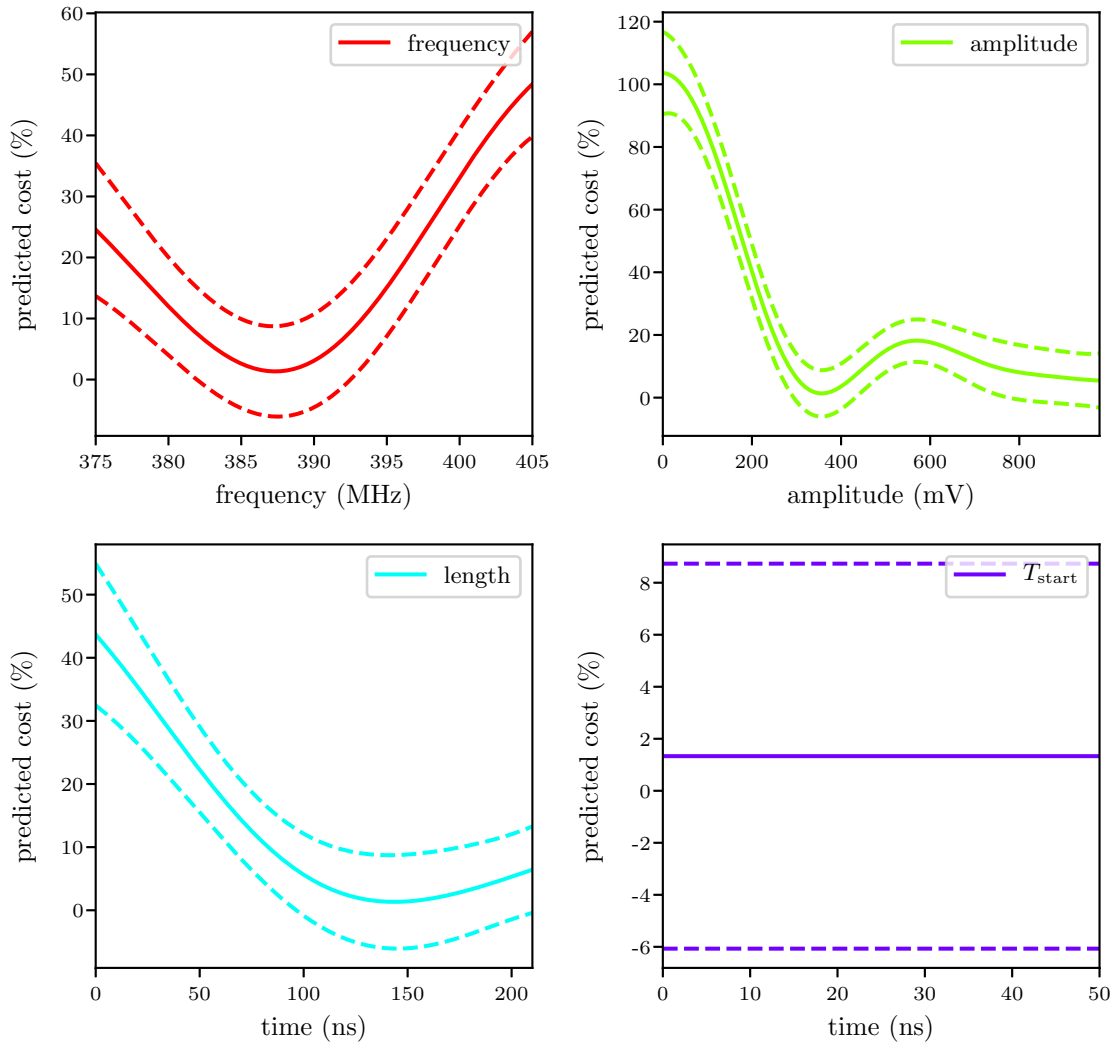


Figure 4.35.: Cross-sections for the parameters used in the optimization of the readout process using a single rectangular pulse. Of the parameters the offset of the square pulse relative to the beginning of the readout window is the least important.

rapidly before reaching a local maximum. For large RF amplitudes both ionization probabilities decrease which might be associated with lower transmissions of the AOM. Together, this leads to the contrast reaching its maximum at $A = 312.3\,\text{mV}$ when the ionization probability for the bright state is almost maximized while it is still low for the dark state.

Independently fitting the redetection probabilities for dark and bright state gives us some intuition about the underlying physics, however, both GP make unphysical predictions. Better results might be obtained by constraining the GP to the physical boundaries of the problem [105, 106], namely redetection probabilities and contrast between 0 and 1. Additionally, the GPs suffer from limited data, large uncertainties

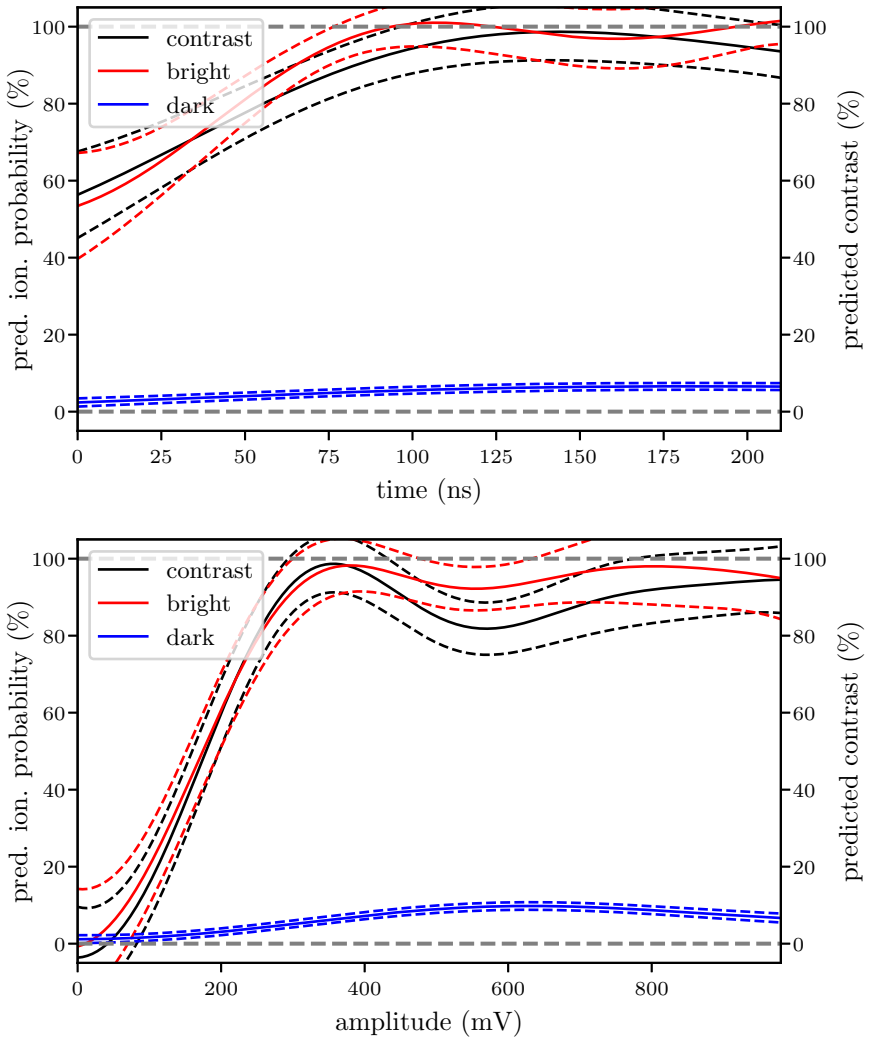


Figure 4.36.: Predicted redetection probabilities for the bright (red) and dark (blue) state, as well as the predicted contrast (black) as a function of the readout square pulse amplitude (bottom) and duration (top), respectively. Each prediction results from an *independent* GP with four parameters. For each cross-section, all other parameters are fixed at the value that yields the best predicted contrast. Dashed lines indicate the physical bounds, 0 and 100 %, which are unbeknown to the GP.

in the provided data and the presence of irrelevant parameters. Nonetheless, we have shown a straight-forward way of using GPs in the experiment outside the context of online optimization, where data is produced by the optimizer itself. In the future, they could find application in the analysis by helping to post-process experimental data or for adjusting parameters like magnetic fields or laser powers based on observations collected from the normal operation of the experiment.

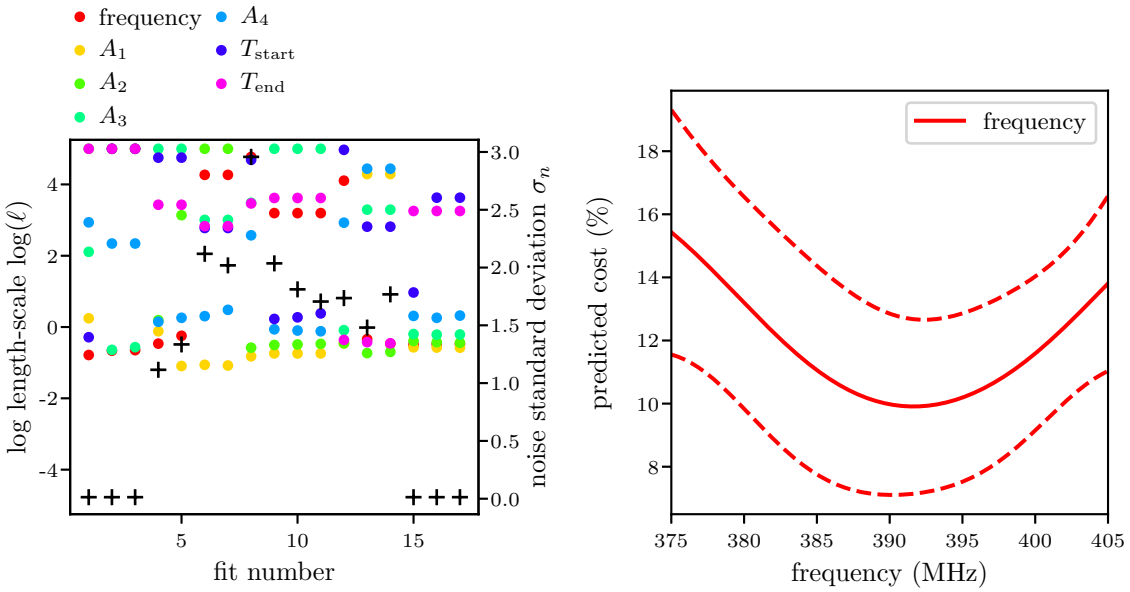
4.3.3b. Modulating the Pulse Amplitude

We have seen previously that the ionization probability of the bright state is very sensitive to the RF amplitude, while the dark state only exhibits a slow ascent. We wanted to investigate whether modulating said amplitude could improve the contrast. Once again we planned on using two different schemes to interpolate the amplitude between several fixed points. The two optimizations that used two (four) equidistant base points (fig. 4.37) with step-wise interpolation alongside the frequency and variable cutoffs at the start and end (T_{start} and T_{end}) of the optimization, remained inconclusive. While M-LOOP once found a similar optimal frequency as before, the overall predictions are poor, which might have several reasons:

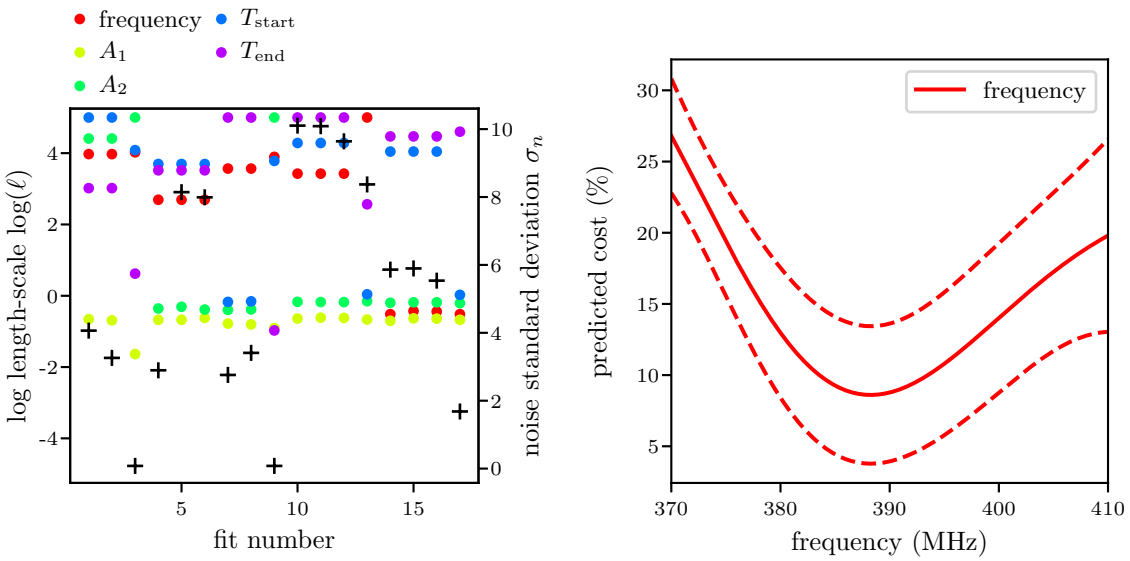
First, we realized that the parameter choice might have been flawed, in particular the use of two independent cut-offs. The previous optimization using a rectangular pulse had show that the landscape of the cost is relatively flat around the optimum and additionally that there is little to no dependence on the offset. While the idea of having two independent cut-offs was to decouple the offset and pulse-length, effectively combined one irrelevant and one relevant parameter, yielding two parameters we could have expected to be hardly relevant and additionally degenerate. Future optimization keeping this in mind would likely be more successful.

Second, it appears as if both optimizations suffered from underestimating the noise level, that is very low compared to the previous optimization despite similar observed uncertainties in all three measurements. Fixing this would have required either more events per data point, or a larger number of data points for the respective optimization, or both. Evidently, that would have prolonged the overall measurement times, that already took 16.3 h and 11.0 h for the optimization with seven and five parameters, respectively.

Unfortunately, due to technical difficulties that already occurred after five days of measurements for the readout and that have not been fully resolved by the time of this writing, none of these ideas were implemented.



(a) Optimization using four base points at $\{0.25, 0.5, 0.75, 1.0\} \times 210\,\text{ns} + T_{\text{start}}$.



(b) Optimization using two base points at $\{0.5, 1.0\} \times 210\,\text{ns} + T_{\text{start}}$.

Figure 4.37.: Optimization of the readout pulse using amplitude modulation with step-wise interpolation between four (a) and two (b) points. Additional parameters are the readout frequency and two independent cut-offs, T_{start} and T_{end}.

4.4. Conclusions from the Optimizations

We have demonstrated a way of integrating AWGs into the existing experimental infrastructure. We have seen that they are able to fully replace the currently employed combination of RF driver and pulse-shaper for operating the AOMs in the experiment without loss of performance (sec. 4.2.3a and 4.3.3a). The necessary adaptation of the experimental software framework, in particular the scheduler, have been thoroughly tested throughout the period in which the optimizations have been performed. Possible issues have been eliminated to the best of our ability and knowledge.

Furthermore, we have shown that *automatized* online optimizations is a valuable addition to other techniques aimed at maintaining a high experimental performance. By employing ML algorithms, in particular DE and GPR, we were first able to replicate the results of previous *manual* optimizations of the initial state preparation and the readout process. New laser sequences for the two processes were developed and investigated. While to this point, no significant improvement beyond what would be possible and expected without using ML for either of the two processes was achieved, implementing novel pulse sequences in the future could be facilitated using the machine learner, with or without an AWG.

The optimizations of the initial state preparation showed that in lab 2 improvements could be achieved, mostly by depopulating the $F = 2$ level. The easiest option to achieve this is to extend the overall pumping time to roughly $\sim 11.0\,\mu s$ (sec. 4.2.3b), which would not require any changes to the current experimental infrastructure. Experimentally we observed a relative increase in single photon detection efficiency of $\sim 20\%$ (fig. 4.12) and and an improvement of the initial state preparation efficiency from 76.2(9) % with $2.3 + 2\mu s$ to 82.6(9) %(fig. 4.13). While extending the total pumping time would approximately half the repetition rate at intermediate distances where this time would be larger than the time it takes the heralding signal to travel between the setups and the BSM ($7.6\,\mu s$ for 700 m optical fiber), this effect is substantially smaller at larger distances ($\mathcal{O}(10\,km)$). Smaller improvements were also possible for pumping schemes where $pump_{2\to1}$ was single-tone frequency modulated (sec. 4.2.3d) with a relative increase of η_{det} by $\sim 2\%$ (fig. 4.21). Using this scheme, the population of $F = 2$ after $2.3\,\mu s$ of pumping could be decreased from 22.24(96) % to 15.4(7) % (Table 4.1). Alternatively, it is possible that increasing the amount of power available to $pump_{2\to1}$ would already be sufficient. Similar tests in lab 1 have not been performed.

The optimizations of the readout process once again confirmed previous experimental results while also demonstrating how our parameter choice can influence the GP's ability to optimize. The destructive ionization-based measurement scheme required for determining the contrast resulted in lower statistics and larger uncertainty compared to the optimizations of the initial state preparation. While

significantly increasing the time per iteration would have been able to combat this issue, this approach would have been prohibitively time-consuming, in particular if more parameters were to be used. Alternatively, open-loop optimizations using simulated data for the readout process might be more convenient. The changes outlined in section 2.4.1c could be applied to modify the simulations from [73, 87].

Lastly, we have demonstrated possible applications of GPs that do not require performing closed-loop online optimization, where GPs were employed to develop statistical models from independently collected data. This might help to understand experimental observations and could enhance the experimental operation and data post-processing.

5. Conclusion and Outlook

In this work, we have discussed how to realize a long-distance entanglement between two individually trapped ^{87}Rb atoms, which is used to perform a loophole-free Bell test. Long-distance entanglement distribution is a key ingredient for future quantum networks, that enable distributed quantum computing and quantum communication. We identified two crucial processes, the initial state preparation and atomic state readout, that could be improved to yield even better performance of the setup.

We have seen that integration of an arbitrary waveform generator (AWG) into the existing experimental framework is possible without losing experimental performance, allowing for quick modification of the employed laser pulse sequences. Automatized online optimization using machine learning (ML) techniques, in particular Gaussian process (GP) regression and differential evolution, has been demonstrated to be an efficient tool within the experimental setup. The changes to the experimental software framework that were necessary for the implementation can also help to improve the overall experimental performance and reduce workload for the people involved in the experiment, for example by automatically scanning and re-adjusting parameters on a regular basis.

We were able to improve the initial state preparation, in particular by extending the total pumping time to $\sim 11\,\mu s$, which would be a viable option in experiments with larger spatial separation of the two laboratories. However, it was not possible to find pumping schemes that would lower populations in $|1, \pm 1\rangle$, which could be related to the prevalence of off-resonant excitations from $|1, 0\rangle$ to $5^2P_{3/2}$ during the optical pumping. We propose to investigate whether an optical pumping scheme utilising the D_1 instead of D_2 transition for $\text{pump}_{1\to1}$ could improve η_{prep} (sec. 2.4.1c). Technically, implementing this proposal should be feasible as there is already a laser locked to the D_1 transition, with optical power available that is currently not used in the experiment. Schemes where a combination of 780 nm and 795 nm light fields for $\text{pump}_{1\to1}$ are used could be optimized with regard to η_{prep} and the total pumping time using M-LOOP.

Optimizations of the atomic state readout process were not able to produce schemes that would increase the contrast. The optimizations suffered from high noise levels and long measurement times. In the future, conducting simulation studies could help to improve the understanding and performance of the process. However, we were able to demonstrate the use of GPs outside the context of online optimization.

Lastly, we have identified modification to M-LOOP itself that could potentially improve the performance of the optimizers in scientific experiments.

A. Laser Setup

Different laser beams with frequencies matching the atomic transitions (fig. A.1) are used for control and manipulation of the atomic quantum system. Three diode lasers whose frequency is stabilized using Doppler-free saturation spectroscopy can be split into several beams using PBSs. The individual powers in each beam is set by rotating a HWP in front of the PBS. Then, the laser frequency is fine-tuned to match the desired atomic atomic transition using AOMs, which also switch on and off the respective beam. Currently three beams each are extracted from the cooling and repump laser:

cooling: cooling, cycling, and $\text{pump}_{1\to1}$

repump: repump, excitation, and $\text{pump}_{2\to1}$

There are also two free running laser diodes, the optical dipole trap (ODT) and the ionization laser.

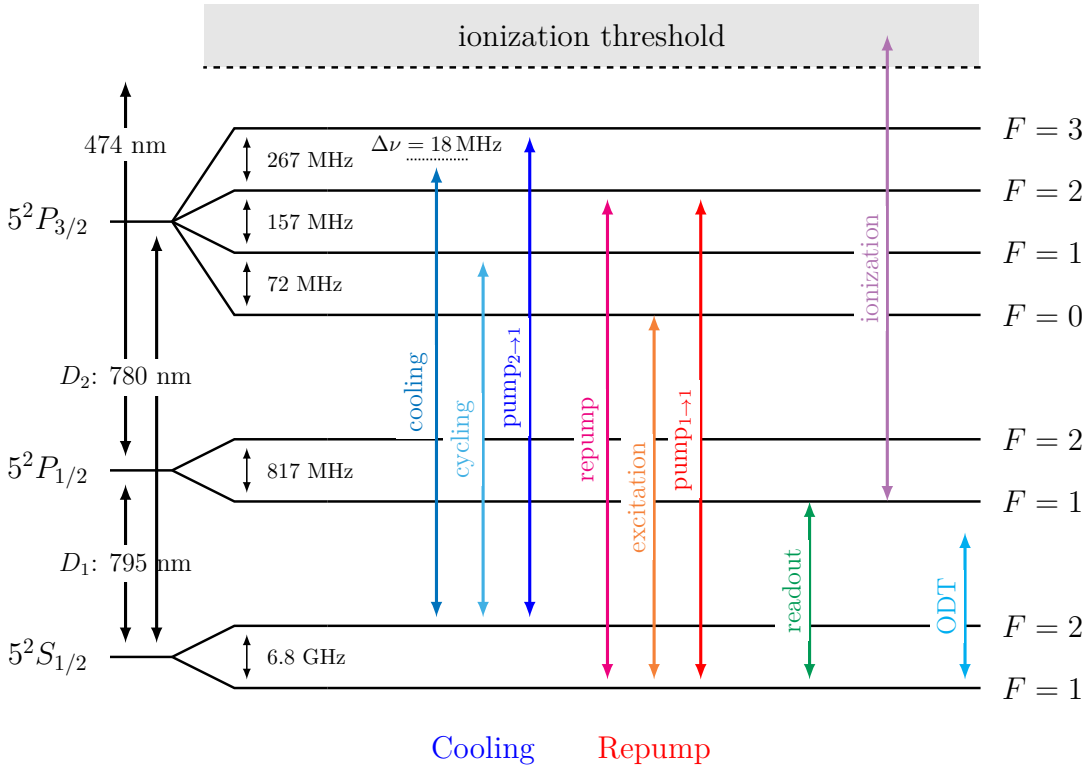


Figure A.1.: Laser frequencies used in the experiment. Several transitions are extracted from cooling (blue) and repump (red) laser.

Bibliography

[1] R. Feynman, *The Feynman lectures on physics*. Reading, Mass: Addison-Wesley Pub. Co, 1963.

[2] M. Planck, "Über eine verbesserung der wienschen spektralgleichung," in *Verhandlungen der Deutschen Physikalischen Gesellschaft 2*, Deutsche Physikalische Gesellschaft, 1900, pp. 202–204.

[3] M. Planck, "Über das gesetz der energieverteilung im norrnalspectrum," in *Verhandlungen der Deutschen Physikalischen Gesellschaft 2*, 1900, p. 237.

[4] A. Einstein, "Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt," en, *Annalen der Physik*, vol. 322, no. 6, pp. 132–148, 1905. DOI: https://doi.org/10.1002/andp.19053220607.

[5] D. Bouwmeester, *The physics of quantum information : quantum cryptography, quantum teleportation, quantum computation*. Berlin New York: Springer, 2000.

[6] V. Giovannetti *et al.*, "Quantum metrology," *Physical Review Letters*, vol. 96, no. 1, 2006. DOI: 10.1103/physrevlett.96.010401.

[7] A. Einstein *et al.*, "Can quantum-mechanical description of physical reality be considered complete?" *Physical Review*, vol. 47, no. 10, pp. 777–780, 1935. DOI: 10.1103/physrev.47.777.

[8] N. Bohr, "Can Quantum-Mechanical Description of Physical Reality be Considered Complete?" *Physical Review*, vol. 48, no. 8, pp. 696–702, 1935. DOI: 10.1103/PhysRev.48.696.

[9] A. Einstein, "QUANTEN-MECHANIK UND WIRKLICHKEIT," *Dialectica*, vol. 2, no. 3-4, pp. 320–324, 1948. DOI: 10.1111/j.1746-8361.1948.tb00704.x.

[10] J. S. Bell, "On the einstein podolsky rosen paradox," *Physics Physique Fizika*, vol. 1, no. 3, pp. 195–200, 1964. DOI: 10.1103/physicsphysiquefizika.1.195.

[11] J. F. Clauser *et al.*, "Proposed experiment to test local hidden-variable theories," *Physical Review Letters*, vol. 23, no. 15, pp. 880–884, 1969. DOI: 10.1103/physrevlett.23.880.

[12] S. J. Freedman and J. F. Clauser, "Experimental test of local hidden-variable theories," *Physical Review Letters*, vol. 28, no. 14, pp. 938–941, 1972. DOI: 10.1103/physrevlett.28.938.

[13] A. Aspect *et al.*, "Experimental test of bell's inequalities using time- varying analyzers," *Physical Review Letters*, vol. 49, no. 25, pp. 1804–1807, 1982. DOI: 10.1103/physrevlett.49.1804.

[14] A. Aspect *et al.*, "Experimental realization of einstein-podolsky-rosen-BohmGedankenexperiment: A new violation of bell's inequalities," *Physical Review Letters*, vol. 49, no. 2, pp. 91–94, 1982. DOI: 10.1103/physrevlett.49.91.

[15] B. Hensen *et al.*, "Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres," *Nature*, vol. 526, no. 7575, pp. 682–686, 2015. DOI: 10.1038/nature15759.

[16] M. Giustina *et al.*, "Significant-loophole-free test of bell's theorem with entangled photons," *Physical Review Letters*, vol. 115, no. 25, 2015. DOI: 10.1103/physrevlett.115.250401.

[17] L. K. Shalm *et al.*, "Strong loophole-free test of local realism," *Physical Review Letters*, vol. 115, no. 25, 2015. DOI: 10.1103/physrevlett.115.250402.

[18] W. Rosenfeld *et al.*, "Event-ready bell test using entangled atoms simultaneously closing detection and locality loopholes," *Physical Review Letters*, vol. 119, no. 1, 2017. DOI: 10.1103/physrevlett.119.010402.

[19] P. St-Hilaire *et al.*, "Real-time holographic display: Improvements using a multichannel acousto-optic modulator and holographic optical elements," S. A. Benton, Ed., SPIE, 1991. DOI: 10.1117/12.44734.

[20] M. Haney and D. Psaltis, "Real-time programmable acoustooptic synthetic aperture radar processor," *Applied Optics*, vol. 27, no. 9, p. 1786, 1988. DOI: 10.1364/ao.27.001786.

[21] B. Braverman *et al.*, "Fast generation and detection of spatial modes of light using an acousto-optic modulator," *Optics Express*, vol. 28, no. 20, p. 29 112, 2020. DOI: 10.1364/oe.404309.

[22] B. Fröhlich *et al.*, "Two-frequency acousto-optic modulator driver to improve the beam pointing stability during intensity ramps," *Review of Scientific Instruments*, vol. 78, no. 4, p. 043 101, 2007. DOI: 10.1063/1.2720725.

[23] M. Endres *et al.*, "Atom-by-atom assembly of defect-free one-dimensional cold atom arrays," *Science*, vol. 354, no. 6315, pp. 1024–1027, 2016. DOI: 10.1126/science.aah3752.

[24] D. Barredo *et al.*, "An atom-by-atom assembler of defect-free arbitrary two-dimensional atomic arrays," *Science*, vol. 354, no. 6315, pp. 1021–1023, 2016. DOI: 10.1126/science.aah3778.

[25] H. Bernien *et al.*, "Probing many-body dynamics on a 51-atom quantum simulator," *Nature*, vol. 551, no. 7682, pp. 579–584, 2017. DOI: 10.1038/nature24622.

[26] G. Carleo *et al.*, "Machine learning and the physical sciences," *Rev. Mod. Phys. 91, 045002 (2019)*, 2019. DOI: 10.1103/RevModPhys.91.045002. arXiv: 1903.10563 [physics.comp-ph].

[27] S. J. Wetzel and M. Scherzer, "Machine Learning of Explicit Order Parameters: From the Ising Model to SU(2) Lattice Gauge Theory," *Physical Review B*, vol. 96, no. 18, p. 184 410, 2017, arXiv: 1705.05582. DOI: 10.1103/PhysRevB.96.184410.

[28] M. Ntampaka *et al.*, "The Role of Machine Learning in the Next Decade of Cosmology," *arXiv:1902.10159 [astro-ph]*, 2021, arXiv: 1902.10159.

[29] J. Werschnik and E. K. U. Gross, "Quantum optimal control theory," 2007. arXiv: 0707.1883 [quant-ph].

[30] T. Caneva *et al.*, "Chopped random-basis quantum optimization," *Phys. Rev. A 84, 022326 (2011)*, 2011. DOI: 10.1103/PhysRevA.84.022326. arXiv: 1103.0855 [quant-ph].

[31] Z. An *et al.*, "Quantum optimal control of multilevel dissipative quantum systems with reinforcement learning," *Physical Review A*, vol. 103, no. 1, p. 012 404, 2021. DOI: 10.1103/PhysRevA.103.012404.

[32] D. Zeidler *et al.*, "Evolutionary algorithms and their application to optimal control studies," *Physical Review A*, vol. 64, no. 2, p. 023 420, 2001. DOI: 10.1103/PhysRevA.64.023420.

[33] J. Kelly *et al.*, "Optimal Quantum Control Using Randomized Benchmarking," *Physical Review Letters*, vol. 112, no. 24, p. 240 504, 2014. DOI: 10.1103/PhysRevLett.112.240504.

[34] P. B. Wigley *et al.*, "Fast machine-learning online optimization of ultra-cold-atom experiments," *Scientific Reports*, vol. 6, no. 1, 2016. DOI: 10.1038/srep25890.

[35] W. Heisenberg, "Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen.," de, *Zeitschrift für Physik*, vol. 33, no. 1, pp. 879–893, 1925. DOI: 10.1007/BF01328377.

[36] M. Born *et al.*, "Zur Quantenmechanik. II.," de, *Zeitschrift für Physik*, vol. 35, no. 8, pp. 557–615, 1926. DOI: 10.1007/BF01379806.

[37] J. J. Sakurai, *Modern quantum mechanics*. Reading, Mass: Addison-Wesley Pub. Co, 1994.

[38] D. F. S. David J. Griffiths, *Introduction to Quantum Mechanics*. Cambridge University Pr., 2018.

[39] D. Bohm, "A suggested interpretation of the quantum theory in terms of "hidden"variables. i," *Physical Review*, vol. 85, no. 2, pp. 166–179, 1952. DOI: 10.1103/physrev.85.166.

[40] P. A. M. Dirac, "A new notation for quantum mechanics," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 35, no. 3, pp. 416–418, 1939. DOI: 10.1017/s0305004100021162.

[41] S. Barnett, *Quantum Information*. OXFORD UNIV PR, 2009.

[42] R. P. Feynman *et al.*, "Geometrical representation of the schrödinger equation for solving maser problems," *Journal of Applied Physics*, vol. 28, no. 1, pp. 49–52, 1957. DOI: 10.1063/1.1722572.

[43] L. Knips, "Efficient quantum state analysis and entanglement detection," Ph.D. dissertation, Ludwig-Maximilians-Universität, Munich, 2019.

[44] J. Von Neumann, *Mathematische Grundlagen der Quantenmechanik,* Berlin: J. Springer, 1932, English translation: J. V. Neumann, *Mathematical Foundations of Quantum Mechanics*. Princeton University Press, 1996.

[45] K. S. M. Redeker, "Entanglement of single rubidium atoms: From a bell test towards applications," Ph.D. dissertation, Ludwig-Maximilians-Universität, Munich, 2020.

[46] J. Barrett *et al.*, "Quantum nonlocality, bell inequalities, and the memory loophole," *Physical Review A*, vol. 66, no. 4, 2002. DOI: 10.1103/physreva.66.042111.

[47] N. Brunner *et al.*, "Bell nonlocality," *Reviews of Modern Physics*, vol. 86, no. 2, pp. 419–478, 2014. DOI: 10.1103/revmodphys.86.419.

[48] D. Elkouss and S. Wehner, "(nearly) optimal p values for all bell inequalities," *npj Quantum Information*, vol. 2, no. 1, 2016. DOI: 10.1038/npjqi.2016.26.

[49] E. Schrödinger, "Quantisierung als Eigenwertproblem," en, *Annalen der Physik*, vol. 384, no. 4, pp. 361–376, 1926. DOI: https://doi.org/10.1002/andp.19263840404.

[50] D. Manzano, "A short introduction to the lindblad master equation," *AIP Advances*, vol. 10, no. 2, p. 025 106, 2020. DOI: 10.1063/1.5115323.

[51] H. Carmichael, *An Open Systems Approach to Quantum Optics*. Springer Berlin Heidelberg, 1993. DOI: 10.1007/978-3-540-47620-7.

[52] H.-J. Briegel and B.-G. Englert, "Quantum optical master equations: The use of damping bases," *Physical Review A*, vol. 47, no. 4, pp. 3311–3329, 1993. DOI: 10.1103/physreva.47.3311.

[53] J. Hofmann, "Heralded atom-atom entanglemen," Ph.D. dissertation, Ludwig-Maximilians-University Munich, 2014.

[54] U. Gaubatz *et al.*, "Population transfer between molecular vibrational levels by stimulated raman scattering with partially overlapping laser fields. a new concept and experimental results," *The Journal of Chemical Physics*, vol. 92, no. 9, pp. 5363–5376, 1990. DOI: 10.1063/1.458514.

[55] P. Marte *et al.*, "Coherent atomic mirrors and beam splitters by adiabatic passage in multilevel systems," *Physical Review A*, vol. 44, no. 7, R4118–R4121, 1991. DOI: 10.1103/physreva.44.r4118.

[56] M. Fleischhauer *et al.*, "Electromagnetically induced transparency: Optics in coherent media," *Reviews of Modern Physics*, vol. 77, no. 2, pp. 633–673, 2005. DOI: 10.1103/RevModPhys.77.633.

[57] J. D. Jackson, *Classical Electrodynamics*. WILEY, 1998.

[58] M. Zukowski *et al.*, ""Event-ready-detectors" bell experiment via entanglement swapping," *Physical Review Letters*, vol. 71, no. 26, pp. 4287–4290, 1993. DOI: 10.1103/physrevlett.71.4287.

[59] J.-W. Pan *et al.*, "Experimental entanglement swapping: Entangling photons that never interacted," *Physical Review Letters*, vol. 80, no. 18, pp. 3891–3894, 1998. DOI: 10.1103/physrevlett.80.3891.

[60] J. Hofmann *et al.*, "Heralded Entanglement Between Widely Separated Atoms," en, *Science*, vol. 337, no. 6090, pp. 72–75, 2012. DOI: 10.1126/science.1221856.

[61] W. Rosenfeld *et al.*, "Towards high-fidelity interference of photons emitted by two remotely trapped rb-87 atoms," *Optics and Spectroscopy*, vol. 111, no. 4, pp. 535–539, 2011. DOI: 10.1134/s0030400x11110233.

[62] C. K. Hong *et al.*, "Measurement of subpicosecond time intervals between two photons by interference," *Physical Review Letters*, vol. 59, no. 18, pp. 2044–2046, 1987. DOI: 10.1103/physrevlett.59.2044.

[63] V. Fock, "Konfigurationsraum und zweite quantelung," *Zeitschrift für Physik*, vol. 75, no. 9-10, pp. 622–647, 1932. DOI: 10.1007/bf01344458.

[64] D. A. Steck, *Rubidium 87 d line data*, available onlne at, revision 2.2.1, 2019. eprint: https://steck.us/alkalidata/rubidium87numbers.pdf.

[65] D. Taray, "Towards increasing atomic coherence times with a standing wave dipole trap," M.S. thesis, Ludwig-Maximilians-University Munich, 2020.

[66] M. Körber *et al.*, "Decoherence-protected memory for a single-photon qubit," en, *Nature Photonics*, vol. 12, no. 1, pp. 18–21, 2018. DOI: 10.1038/s41566-017-0050-y.

[67] S. Chu *et al.*, "Experimental observation of optically trapped atoms," *Physical Review Letters*, vol. 57, no. 3, pp. 314–317, 1986. DOI: 10.1103/physrevlett.57.314.

[68] E. L. Raab *et al.*, "Trapping of neutral sodium atoms with radiation pressure," *Physical Review Letters*, vol. 59, no. 23, pp. 2631–2634, 1987. DOI: 10.1103/physrevlett.59.2631.

[69] T. Hänsch and A. Schawlow, "Cooling of gases by laser radiation," *Optics Communications*, vol. 13, no. 1, pp. 68–69, 1975. DOI: 10.1016/0030-4018(75)90159-5.

[70] S. Chu *et al.*, "Three-dimensional viscous confinement and cooling of atoms by resonance radiation pressure," *Physical Review Letters*, vol. 55, no. 1, pp. 48–51, 1985. DOI: 10.1103/physrevlett.55.48.

[71] N. Ortegel, "State readout of single rubidium-87 atoms for a loophole-free test of bell's inequality," Ph.D. dissertation, Ludwig-Maximilans-University Munich, 2016.

[72] K. Redeker, *Detektion von Ionisationsfragmenten von Rb-87 Atomen*, Diploma Thesis, 2012.

[73] J. Volz, "Atom-photon entanglement," Ph.D. dissertation, Ludwig-Maximilans-University, Munich, 2006.

[74] D. Burchardt, "A rigorous test of bell's inequality and quantum teleportation employing single atoms," Ph.D. dissertation, Ludwig-Maximilans-University Munich, 2018.

[75] P. Koschmieder, "Efficient control of magnetic fields in single atom experiments," M.S. thesis, Ludwig-Maximilians-University Munich, 2019.

[76] J. Dalibard and C. Cohen-Tannoudji, "Laser cooling below the doppler limit by polarization gradients: Simple theoretical models," *Journal of the Optical Society of America B*, vol. 6, no. 11, p. 2023, 1989. DOI: 10.1364/josab.6.002023.

[77] R. Grimm *et al.*, "Optical dipole traps for neutral atoms," in *Advances In Atomic, Molecular, and Optical Physics*, Elsevier, 2000, pp. 95–170. DOI: 10.1016/s1049-250x(08)60186-x.

[78] M. Fürst *et al.*, "High speed optical quantum random number generation," *Optics Express*, vol. 18, no. 12, p. 13 029, 2010. DOI: 10.1364/oe.18.013029.

[79] M. Weber, "Optical dipole traps for neutral atoms," Ph.D. dissertation, Ludwig-Maximilians-University Munich, 2005.

[80] N. Schlosser *et al.*, "Collisional blockade in microscopic optical dipole traps," *Physical Review Letters*, vol. 89, no. 2, 2002. DOI: 10.1103/physrevlett.89.023005.

[81] R. S. Garthoff, "Optimized single photon collectionfor quantum applications employing single atoms," not yet published, Ph.D. dissertation, Ludwig-Maximilians-University Munich, 2021.

[82] T. van Leent *et al.*, "Long-distance distribution of atom-photon entanglement at telecom wavelength," *Physical Review Letters*, vol. 124, no. 1, 2020. DOI: 10.1103/physrevlett.124.010510.

[83] M. Bock, "Polarization-preserving quantum frequency conversion for trapped-atom based quantum networks," final defence and publication pending, Ph.D. dissertation, Saarland University Saarbrücken, 2020.

[84] W. Rosenfeld, "Experiments with an entangled system of a single atom and a single photon," Ph.D. dissertation, Ludwig-Maximilians-University Munich, 2008.

[85] U. Volz and H. Schmoranzer, "Precision lifetime measurements on alkali atoms and on helium by beam–gas–laser spectroscopy," en, *Physica Scripta*, vol. T65, pp. 48–56, 1996. DOI: 10.1088/0031-8949/1996/T65/007.

[86] P. v. d. S. Harold J. Metcalf, *Laser Cooling and Trapping*. Springer New York, 2001.

[87] R. S. Garthoff, "Optimierte erzeugung und nachweis von atom-photon-zuständen zur verschränkung räumlich getrennter atome," this work is only availbe in German, M.S. thesis, Ludwig-Maximilans-University Munich, 2015.

[88] L. Maisenbacher, "Precision spectroscopy of the 2s-np transitions in atomic hydrogen," Ph.D. dissertation, Ludwig–Maximilians–Universität, 2020.

[89] A. Beyer *et al.*, "The rydberg constant and proton size from atomic hydrogen," *Science*, vol. 358, no. 6359, pp. 79–85, 2017. DOI: 10.1126/science.aah6677.

[90] S. A. Lee *et al.*, "Doppler-free two-photon transitions to rydberg levels: Convenient, useful, and precise reference wavelengths for dye lasers," *Optics Letters*, vol. 3, no. 4, p. 141, 1978. DOI: 10.1364/ol.3.000141.

[91] M. Krug, "Ionization based state read out of a single rb-87 atom," Ph.D. dissertation, Ludwig-Maximilians-University Munich, 2018.

[92] T. Rieger and T. Volz, *Doppler-free saturation spectroscopy*, accessed 2021-01-10, Max-Planck-Institut for Quantum Optics, Garching.

[93] E. A. Donley *et al.*, "Double-pass acousto-optic modulator system," *Review of Scientific Instruments*, vol. 76, no. 6, p. 063 112, 2005. DOI: 10.1063/1.1930095.

[94] M. R. Hush, *Machine-learning online optimization package*, https://m-loop.readthedocs.io/en/latest/index.html, 2016.

[95] C. E. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. MIT Press Ltd, 2005.

[96] M. Kuss, "Gaussian process models for robust regression,classification, and reinforcement learning," Ph.D. dissertation, Technische Universität Darmstadt, 2006.

[97] F. Leclercq, "Bayesian optimization for likelihood-free cosmological inference," *Physical Review D*, vol. 98, no. 6, p. 063 511, 2018. DOI: 10.1103/PhysRevD.98.063511.

[98] K. V. Mardia *et al.*, *Multivariate analysis*, English. London; New York: Academic Press, 1979, OCLC: 6164035.

[99] D. K. Duvenaud, "Automatic model construction with gaussian processes - kernels," Ph.D. dissertation, University of Cambridge, 2014.

[100] C. A. Micchelli *et al.*, "Universal kernels," *Journal of Machine Learning Research*, vol. 7, no. 95, pp. 2651–2667, 2006.

[101] R. Fletcher, *Practical methods of optimization*. Chichester New York: Wiley, 1987, L-BFGS.

[102] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965. DOI: 10.1093/comjnl/7.4.308.

[103] R. Storn and K. Price, "Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces," en, *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, 1997. DOI: 10.1023/A:1008202821328.

[104] B. Amstrup *et al.*, "Genetic Algorithm with Migration on Topology Conserving Maps for Optimal Control of Quantum Systems," *The Journal of Physical Chemistry*, vol. 99, no. 14, pp. 5206–5213, 1995. DOI: 10.1021/j100014a048.

[105] L. Swiler *et al.*, "A survey of constrained gaussian process: Approaches and implementation challenges," *Journal of Machine Learning for Modeling and Computing*, vol. 1, no. 2, 2020. DOI: 10.1615/jmachlearnmodelcomput.2020035155.

[106] M. M. Noack and J. A. Sethian, "Advanced stationary and non-stationary kernel designs for domain-aware gaussian processes," 2021. arXiv: 2102.03432 [stat.ML].

Acknowledgments

At the end of this thesis I want to thank the people that, through direct and indirect support have helped to make this thesis possible.

First, I want to thank Prof. Dr. Harald Weinfurter for the opportunities I have been presented as part his group at LMU. Thank you to Tim van Leent, Florian Fertig, Wei Zhang, Kai Redecker, and Robert Garthoff who have introduced me patiently to the realm of quantum optics and to the world of single rubidium atoms. Your critical thoughts, knowledge, experience, time and trust in my abilities have been vital for my progress during this challenging times. My thanks also extend to Martin Zeitlmair, Mathias Bock, and all other members of the XPQ group, colleagues, and people of the chair of laserspectroscopy at LMU.

Special thanks to my friends, especially Lotti, Ines, and Nico, who have been by my side throughout six year of studies and particularly during this last one. I cannot imaging this getting through this time without you, both academically and personally. Thank you to the many people I have met along the way: Max, Elias, Anna, Derya, Elia, Dominik, and Ismail, to name just a few. Writing these lines mere hours before finally submitting my thesis, please forgive my brevity. I am almost certain that these will not be the last (written) acknowledgments of my life, and I am well aware that there is room for improvements.

Lastly, I want to thank my family, my two siblings Elisabeth and Valentin and my parents Barbara and Gisbert. From a young age, you have fostered and supported my curiosity, that eventually brought me to this point.

Declaration of Authorship

Eigenständigkeitserkärung

Hiermit erkläre ich, Sebastian Eppelt, die vorliegende Arbeit sebstständig verfasst zu haben und keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel benutzt zu haben.

München, den 8. März 2021 Unterschrift