Numerically Optimized Control of Analogue Physical Systems

Dissertation
zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Physik
der Universität Hamburg

vorgelegt von

Nicolas Tobias Heimann

Hamburg

2025

Gutachter/innen der Dissertation:	Prof. Dr. Ralf Riedinger Prof. Dr. Dieter Jaksch
Zusammensetzung der Prüfungskommission:	Prof. Dr. Ralf Riedinger Prof. Dr. Dieter Jaksch Prof. Dr. Michael Potthoff Prof. Dr. Roman Schnabel Dr. Christoph Becker
Vorsitzende/r der Prüfungskommission:	Prof. Dr. Michael Potthoff
Datum der Disputation:	16.02.2026
Vorsitzender des Fach-Promotionsausschusses PHYSIK:	Prof. Dr. Wolfgang J. Parak
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Markus Drescher
Dekan der Fakultät MIN:	Prof. Dr.-Ing. Norbert Ritter

Inhaltsverzeichnis

Preface

This cumulative thesis has been conducted from mid October 2021 until end of July 2025 at the Center for Optical Quantum Technologies. It is based on the publications [1–3], presented in the chapters 3, 4 and 5. My work has been funded by the Cluster of Excellence "Advanced Imaging of Matter" (EXC 2056), Project No. 390715994. The work was supervised by Prof. Dr. Ludwig Mathey and as secondary supervisor by Prof. Dr. Ralf Riedinger. I expand on my explicit contributions to each publication included in this thesis in Sect. 3.4, Sect. 4.7 and Sect. 5.4.

Zusammenfassung

Wir untersuchen die numerisch optimierte Steuerung des analogen Verhaltens in physikalischen Systemen für den Entwurf groß angelegter Interferometer in Gravitationswellenobservatorien und die Entwicklung kohärenter Quantentechnologien. Im Laufe der Entwicklung dieser Technologien wurden bemerkenswerte Meilensteine erreicht, darunter hochpräzise Betriebsbereiche, die die Detektion von Gravitationswellen oder die kohärente Steuerung von Quantensystemen ermöglichen. Ganze wissenschaftliche Teilbereiche profitieren von diesen Errungenschaften und eben diese haben zahlreiche Bestrebungen zur Entwicklung industrietauglicher Quantentechnologien ausgelöst.

Im ersten Teil dieser Arbeit entwerfen wir machine learning-assistierte Strategien zur Reduzierung seismischer Störungen in Gravitationswellenobservatorien, die auf groß angelegten Michelson-Interferometern basieren. In einer Zusammenarbeit mit der Forschungsgruppe von Roman Schnabel untersuchen wir Pendelaufhängungssysteme in einem Machbarkeitsnachweis Experiment, das aus einem Pendel mit hohem Qualitätsfaktor (hohem Q-Faktor) und einem dreiachsigen Kraftrückkopplungsseismometer besteht. In der Veröffentlichung [1] demonstrieren wir die Fähigkeiten der machine learning-assistierten multivariaten Zeitreihenprognose zur Vorhersage der Bewegung des Pendels mit hohem Q-Faktor unter Einwirkung von seismischen Störungen. Unser System bietet eine flexible Architektur, um seismische Sensornetzwerke in einem modellfreien Ansatz einzubeziehen. Wir erreichen eine Verringerung der Verschiebungsleistungsspektraldichte für einen breiten Frequenzbereich, einschließlich mechanischer Resonanzen. Darüber hinaus stellen wir eine Verringerung nichtlinearer Instrumentierungsartefakte fest, was die Vorteile nichtlinearer künstlicher neuronaler Netze gegenüber herkömmlichen Ansätzen wie dem Wiener Filter verdeutlicht. Unsere Ergebnisse stützen die Idee einer machine learning-assistierten korrigierenden Vorwärtsstabilisierung in Gravitationswellendetektoren der nächsten Generation unter Verwendung seismischer Sensornetzwerke.

Im zweiten Teil dieser Arbeit untersuchen wir die Optimierung der kohärenten Steuerung in Quantencomputern zum Zweck des Entwurfs von hochpräzisen Qubit-Gattern in Präsenz von experimentell motivierten Verlusten und Einschränkungen. In einer Zusammenarbeit mit der Forschungsgruppe von Klaus Sengstock implementieren wir in der Veröffentlichung [2] eine numerische Plattform für den Entwurf optimierter Quantengatter für Quantencomputer auf Basis von Rydberg-Atomen. Dazu implementieren wir einen hybriden quanten-klassischen Feedbackloop, welcher es ermöglicht, die zugrundeliegende Parametrisierung auf einem klassischen Gerät zu trainieren, basierend auf Gradienten der Verlustlandschaft, die über Feedback von unserer Plattform errechnet werden. Konkret verwenden wir gradient ascent pulse engineering (GRAPE), um Gradienten über endliche Differenzen im Parameterraum zu approximieren. Für feste

Gatterdauern identifizieren wir hochpräzise Implementierungen für einen breiten Bereich von Abständen zwischen den Atomen, der vom Blockaderegime bis zur Schwachkopplungsgrenze reicht. Wir demonstrieren die Robustheit optimierter Implementierungen von Quantengattern gegenüber räumlichen Fluktuationen der gefangenen Atome für den betrachteten Bereich von Abständen der Atome. Darüber hinaus stellen wir in der Veröffentlichung [3] eine quantenoptimale Steuerungsmethode vor, welche lineare Antworttheorie zur Schätzung von Gradienten nutzt. Mit unserer Methode stellen wir eine Erweiterung zu Gradienten basierten optimalen Steuerungsmethoden wie GRAPE vor. Wir schätzen Gradienten anhand endlicher Differenzen im Raum der durch die Steuerungsoperatoren aufgespannt wird, über die lineare Antwort des Systems auf zeitlich lokale Störungen der Steuerungsfelder ab. Dies ermöglicht eine hyperparameterfreie Mehrparameteraktualisierung, die sich die multimodale Überlappung der Störung und die zugrunde liegende Parametrisierung der Steuerungsfelder zunutze macht. Wir zeigen deutliche Verbesserungen hinsichtlich der Konvergenz und der optimalen Genauigkeit der generierten Protokolle für das Beispiel eines Quantengatters auf zwei Qubits, im Vergleich zu Standard-GRAPE. Wir demonstrieren unsere Methode unter Berücksichtigung experimentell motivierter Einschränkungen und Dekohärenz und zeigen, dass die resultierenden Genauigkeiten nahe der geschätzten Werten konvergieren.

Abstract

We study numerically optimized control of analog behavior in physical systems for the design of large scale interferometers in gravitational-wave observatories and the development of coherent quantum technology. Remarkable milestones have been achieved over the course of the development of these technologies, reaching high precision operational regimes that enable the detection of gravitational waves or the coherent control of quantum systems. These achievements advance various scientific fields and have sparked numerous endeavors toward industry-ready quantum technology.

In the first part of this thesis, we design machine learning-assisted strategies for the purpose of seismic noise mitigation in gravitational-wave observatories based on large-scale michelson interferometers. In a collaboration with the experimental group of Roman Schnabel, we study pendulum suspension systems in a proof-of-principle experiment consisting of a pendulum with a high quality factor (high-Q) and a triaxial force-feedback seismometer. In publication [1], we demonstrate the capabilities of machine learning-assisted multivariate time-series-forecasting to predict the motion of the high-Q pendulum subject to witnessed seismic noise. Our system provides a flexible architecture to include seismic sensing networks in a model-free approach. We achieve a reduction in the displacement power spectral density for a wide range of frequencies including mechanical resonances. Additionally, we find a reduction of non-linear instrumentation artifacts rendering advantages of non-linear artificial neural networks over conventional approaches, such as Wiener filtering. Our results support the idea of machine learning-assisted corrective forward stabilization in next-generation gravitational-wave detectors utilizing seismic sensing networks.

In the second part of this thesis, we study the optimization of coherent control in quantum computers for the purpose of the design of high-fidelity qubit gates under experimentally motivated losses and constraints. In a collaboration with the experimental group of Klaus Sengstock, we implement a numerical platform for the design of optimized quantum gates for Rydberg-atom based quantum computers in publication [2]. For this, we implement a hybrid quantum-classical loop that allows to train the underlying parameterization on a classical device, based on gradients of the loss landscape that are obtained via feedback from our platform. Specifically, we use gradient ascent pulse engineering (GRAPE) to approximate gradients via finite differences in the parameter space. For fixed gate durations, we identify high-fidelity implementations for a wide range of interatomic distances, ranging from the Blockade regime to the weak-coupling limit. We demonstrate the robustness of optimized implementations of quantum gates to spatial fluctuations of the trapped atoms for the considered range of interatomic distances. Furthermore, we present a quantum optimal control method that utilizes linear response theory to estimate gradients in publication [3]. With our method, we offer

an extension to gradient-based optimal control methods such as GRAPE. We estimate gradients in terms of finite differences in the space spanned by the control operators via the linear response of the system to time-local perturbations of the control fields. This allows for a multi-parameter update in a hyperparameter-free manner, capitalizing on from the multi-mode overlap of the perturbation and the underlying parameterization of the control fields. We show clear improvements in convergence and optimal fidelity of the generated protocols, for the example of a quantum gate on two qubits, in comparison to standard GRAPE. We demonstrate our method in the presence of experimentally motivated constraints and decoherence, showing that the resulting fidelities converge close to estimated values.

List of Publications

This cumulative dissertation is based on the publications listed below.

- **Publication I**: N. Heimann, J. Petermann, D. Hartwig, R. Schnabel, and L. Mathey, *Predicting the motion of a high-Q pendulum subject to seismic perturbations using machine learning*, Appl. Phys. Lett. 122, 254101 (2023) [1]
- **Publication II**: N. Heimann, L. Broers, N. Pintul, T. Petersen, K. Sponselee, A. Ilin, C. Becker, and L. Mathey, *Noise-Robust Quantum Gate Optimization for Rydberg Architectures*, arXiv:2306.08691 (2023) [2]
- **Publication III**: N. Heimann, L. Broers, and L. Mathey, *Pulse Engineering via Projection of Response Functions*, Phys. Rev. Research 7, 013101 (2025) [3]

1 Introduction

The physical reality is riddled with analog behaviors. Our daily perception aligns with this as many phenomena we encounter appear continuous. The control of analogous phenomena is at heart of many technologies and applications. Even the youngest among us aim for the of control analog systems, such as swings, by utilizing appropriate kinematic inputs to reach operational regimes optimized for fun and thrills. Suspension systems in cars isolating the occupants from bumpy roads are also optimized analog systems with the target to enable a pleasure cruise. Analog behaviors are not exclusive to our macroscopic world. The dynamical behavior of quantum systems can be described as continuous rotations of wave functions in Hilbert spaces. These rotations become intriguingly complex with the size of the system demanding for sophisticated control strategies to explore coherent phenomena that can be exploited for the development of coherent quantum technology.

In this cumulative dissertation, I present my work on numerically optimized control of analogue behavior in physical systems. My research focuses on three different applications of numerically optimized control. This includes the study of seismic noise reduction strategies for large scale michelson interferometer gravitational-wave detectors, as well as the design of optimized strategies to implement quantum gates on quantum computers. Specifically, I present my work on seismic noise mitigation strategies for pendulum suspension systems used in gravitational-wave detectors. I present my work on optimized control for quantum computing with neutral atoms based on Rydberg-mediated interactions. In a more generic setup, I present my work on a novel quantum optimal control method, with the target to implement quantum gates with high fidelity. This cumulate thesis is organized as follows: In this first chapter, I introduce this thesis as well as its chapters.

In Chapter 2, I discuss preliminary notions used within this work. This includes the notion of two-level systems and the control of atomic systems based on coherent control fields.

Chapter 3 covers my work, together with the research group of Roman Schnabel, and Ludwig Mathey, on seismic noise reduction strategies in gravitational-wave interferometry. We introduce the field of gravitational-wave astronomy [4] and elaborate on the challenges in detecting gravitational waves in michelson interferometer observatories [5–10]. A formidable challenge is the simulation of free floating mirror test masses to measure relative changes in the optical path length via interference of laser beams with high accuracy. The expected change of the optical path length in this context is magnitudes smaller than the size of a proton [11], demanding for sophisticated strategies to isolate the mirror test masses from environmental disturbances [12]. These mirror test masses are passively isolated from seismic disturbances [13] via stacked multi stage pendu-

lum suspension systems [14], and further stabilized with feed-forward control strategies based on witnessed noise [15]. In a proof-of-principle experiment, we study pendulum suspended systems systems consisting of a pendulum with a high quality factor (high-Q) subject to witnessed seismic noise. We elaborate on machine learning models that we train to learn the motion the pendulum in a model-free approach. This chapter includes publication [1], where we demonstrate the capabilities of machine learning-assisted time-series-forecasting in predicting the motion of a high-Q pendulum in the presence of witnessed seismic noise. The specific contributions to [1] are listed in Sect. 3.4.

In Chapter 4, we turn towards quantum optimal control strategies. Specifically, it covers my work, together with Lukas Broers and Ludwig Mathey, on pulse engineering for quantum computers. We introduce the field of quantum computing with a historical overview starting with the notion of quantum simulators by Richard Feynman [16]. In theory, a quantum simulator can simulate simulate physical systems in a fully probabilistic manner of quantum mechanics, potentially overcoming limitations of classical information theory. One of these limitations is the demand for exponential physical memory to capture the full probability distribution of the underlying system. Similar to classical logic circuits, quantum algorithms utilize quantum-circuits composed of simple building blocks called quantum gates [17]. Nowadays, different platforms providing control over many noisy qubits have emerged, heralding the era of noisy intermediate scale quantum computers [18]. Variational quantum algorithms [19] target to reach favorable operational regimes in such platforms with a hardware-efficient ansatz. Further, we focus on the field of quantum optimal control [20] for the design of optimized implementations of quantum gates on the Hamiltonian domain [21, 22]. Commonly, gradient-based optimal control strategies are utilized to pulse-engineer control fields. Based on small variations in the space of the parameterization, the gradients can be approximated via finite differences. However, we instead utilize linear response theory to estimate such gradients in terms of the linear response of the system to perturbations in the control fields [23]. This chapter includes publication [3], where we introduce pulse engineering via projection of response functions (PEPR), a quantum optimal control technique where we utilize linear response theory to estimate gradients. The specific contributions to [3] are listed in Sect. 4.7.

My work, together with the research group of Klaus Sengstock, Lukas Boers, and Ludwig Mathey, on quantum optimal control strategies for Rydberg-atom systems is covered in Chapter 5. We begin with an introduction to neutral atom based quantum computing [24–30], utilizing Rydberg-mediated interactions based on the strong dipole-dipole interactions between Rydberg atoms [28, 29, 31]. Further, we elaborate on the dipole-dipole interaction [32] and describe the phenomenon of the Rydberg blockade [33, 34]. We characterize three dynamical regimes, namely the blockade regime, the intermediate regime and the weak-coupling limit. Based on Rydberg-mediated interaction, quantum gates can be implemented in the blockade regime and the weak-coupling limit with local and global pulse sequences [35–40]. We design optimized strategies that implement two-qubit gates in all three regimes. For this, we consider the case of the fermionic Ytterbium ^{171}Yb isotope, which provides natural qubit encodings in long-lived nuclear spin states [41–43]. This chapter includes a preprint [2], where we demonstrate opti-

mized strategies to implement two-qubit gates with neutral atoms discuss there robustness against experimentally motivated noise and decoherence. The specific contributions to [2] are listed in Sect. 5.4.

In Chapter 6, I conclude this work.

2 Preliminary notions

In this chapter, I review the most important aspects of two-level and three-level systems for this work, and the control of such systems based on coherent control fields.

2.1 Two-Level systems

A two-level system is the simplest quantum mechanical system. It can be described in terms of a superposition of two orthogonal states, e.g. $|0\rangle$ and $|1\rangle$. The state of a two-level system reads

$$|\psi\rangle = c_0 |0\rangle + c_1 |1\rangle , \tag{2.1}$$

where c_0 and c_1 are complex coefficients normalized such that $|c_0|^2 + |c_1|^2 = 1$. Let $|0\rangle$ be the ground state and $|1\rangle$ be the excited state, with corresponding energies of E_0 and E_1, respectively. The Hamiltonian reads

$$H = \begin{pmatrix} E_1 & 0 \\ 0 & E_0 \end{pmatrix} . \tag{2.2}$$

The solution to the Schrödinger equation

$$i\hbar\partial_t |\psi\rangle = H |\psi\rangle , \tag{2.3}$$

is given by the time-dependent state

$$|\psi(t)\rangle = c_0 |0\rangle + c_1 e^{-i\omega_{10}t} |1\rangle , \tag{2.4}$$

where $\omega_{10} = (E_1 - E_0)/\hbar$ is the transition frequency between the two basis states. The state of the two-level system can be projected onto the unit sphere in three dimensions, commonly referred to as the Bloch-sphere. To represent the state accordingly, it is parametrized by the polar angle θ and the azimuthal angle ϕ, such that at each point in time it can be written as

$$|\psi(\theta, \phi)\rangle = \cos\left(\frac{\theta}{2}\right) |1\rangle + \sin\left(\frac{\theta}{2}\right) e^{i\phi} |0\rangle . \tag{2.5}$$

The polar angle $\theta \in [0, \pi]$ corresponds to the population of the system and the azimuthal angle $\phi \in [0, 2\pi]$ is the relative phase between the two basis states. Here, we use the convention that $\theta = 0$ corresponds to the excited state. The corresponding cartesian representation is given by the Bloch-vector $\mathbf{r} = (\langle\sigma_x\rangle, \langle\sigma_y\rangle, \langle\sigma_z\rangle)^T$, where σ_i are the Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} , \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \tag{2.6}$$

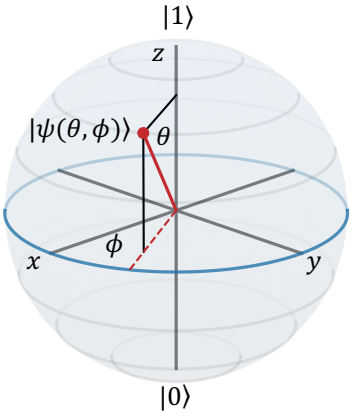


Figure 2.1: **The Bloch-sphere**. The Bloch-vector representation of a state $|\psi(\theta,\phi)\rangle$ is depicted as a red circle. The polar angle θ corresponds the population of the two states, with the ground state at the "south pole", in this convention. The azimuthal angle ϕ is the relative phase difference of the basis states and the length of the Bloch-vector is a measure of coherence. Mixed states are projected onto spheres with radii sub unity.

In Fig. 2.1 we illustrate the two-level state $|\psi(\theta,\phi)\rangle$ for one point in time on the Bloch-sphere. For a pure state, the length of the Bloch-vector is unity, i.e. $|\mathbf{r}| = 1$. The density operator can be written in terms of the Bloch vector

$$\rho = \frac{1 + \mathbf{r}\cdot\boldsymbol{\sigma}}{2}, \tag{2.7}$$

where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)^T$ is the Pauli vector. For mixed states, the length of the Bloch-vector is sub unity.

2.2 The rotating frame

A unitary transformation R defines a rotating frame of reference. A state ψ in this frame reads $\psi_R = R\psi$ and an observable O reads $O_R = ROR^\dagger$. Given a Hamiltonian H, the Schrödinger equation in the rotating frame evaluates to

$$i\hbar\partial_t\psi_R = i\hbar(\partial_t R)R^\dagger\psi_R + RHR^\dagger\psi_R. \tag{2.8}$$

Hence, the dynamics in the rotating frame of R are determined by the modified Hamiltonian

$$\tilde{H} = i\hbar(\partial_t R)R^\dagger + RHR^\dagger, \tag{2.9}$$

and we denote the corresponding unitary transformation generated by $\tilde{H}$ as $U_I(t)$. The Heisenberg picture representation of an observable in the rotating frame reads

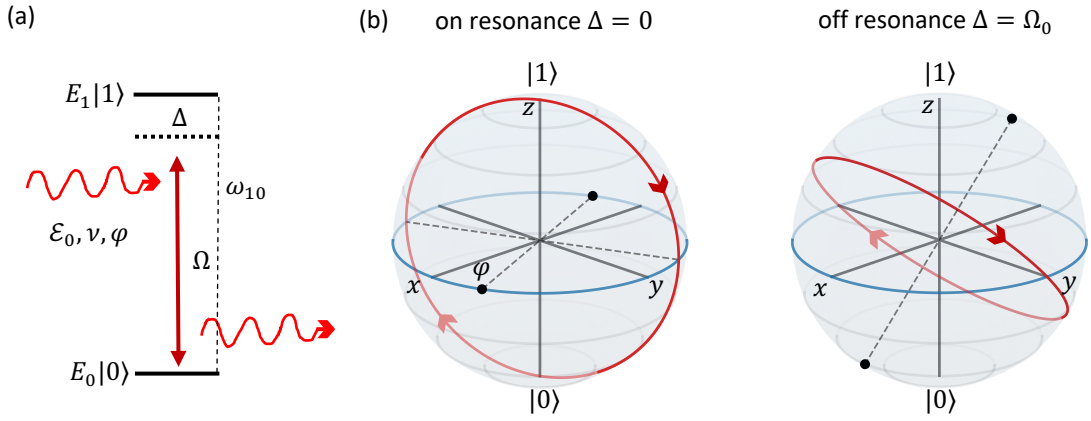


Figure 2.2: **Rabi-oscillations**. (panel a) A monochromatic electric field with a frequency $\nu = \omega_{10} - \Delta$, phase φ and amplitude $\mathcal{E}_0$ stimulates the atom to absorb and emit photons at the rate of the Rabi frequency Ω. (panel b) The resulting oscillatory exchange of population depicted as rotations on the Bloch sphere. On resonance, i.e. for $\Delta = 0$, Rabi oscillations correspond to rotations around a rotational axis that is perpendicular to the z-axis and characterized by the phase φ of the electric field. Off resonance, as shown here for the example of $\Delta = \Omega$, the rotational axis is tilted from the xy-plane which suppresses exchange of population between the ground state and the excited state.

$U_I^\dagger(t) O_R U_I(t)$. Given a hermitian operator H, we refer to $R(t) = \exp(iHt)$ as the rotating frame of H.

In this context, the interaction picture is an important representation of a quantum system in a rotating frame of reference. For the Hamiltonian

$$H = H_0 + V(t), \tag{2.10}$$

the interaction picture representation is defined by the counter-rotating frame of H_0, i.e. $R(t) = U_0^\dagger(t)$, with $U_0(t) = \exp(-iH_0 t/\hbar)$. The modified Hamiltonian in this frame reads

$$\tilde{H}(t) = U_0^\dagger(t) V(t) U_0(t). \tag{2.11}$$

Accrodingly, the representation of an observable O in the interaction picture is given by $O_I(t) = U_0^\dagger(t) O U_0(t)$.

2.3 Single photon Rabi driving

Atomic dipole transitions can be addressed with appropriate electric fields [44]. In the following, we consider two atomic states that are sufficiently decoupled from the remaining part of the spectrum and have sufficiently long coherence times. Then the

system can be treated as a two-level system in the considered subspace, as shown in Fig. 2.2 (a). The atomic transition is driven by a monochromatic electric field with a frequency ν and a wavelength $2\pi/|\mathbf{k}|$,

$$\boldsymbol{\mathcal{E}}(\mathbf{r},t) = \frac{\mathcal{E}_0}{2}\mathbf{e}_\nu e^{-i(\nu t - \mathbf{kr}+\varphi)} + \text{c.c.}\ , \tag{2.12}$$

where $\mathcal{E}_0$ is the strength of the electric field, φ a phase, and $\mathbf{e}_\nu$ defines the polarization axis. The length scales of atoms, given by the Bohr radius $a_0 \approx 10^{-10}$m, are much smaller compared to the typical optical or microwave wavelengths. This allows to consider the electric field to zero-th order in $\mathbf{kr}$ and treat the light–atom interaction as the energy shift of the atomic dipole in the electric field

$$\frac{\hbar\omega_{10}}{2}\sigma_z + \mathbf{d}\boldsymbol{\mathcal{E}}(t), \tag{2.13}$$

where $\mathbf{d}$ is the atomic dipole moment. The electric field frequency is given by $\nu = \omega_{10} - \Delta$, where Δ is a detuning from the atomic transition frequency ω_{10}. Dipole transitions only occur between states of opposite parity. In the following, the states $|0\rangle$ and $|1\rangle$ are assumed parity opposed. In the rotating frame of $\sigma_z\nu/2$, the Hamiltonian reads

$$\frac{\hbar\Delta}{2}\sigma_z + \left[\frac{\hbar\Omega}{2}\left(e^{-i\varphi} + e^{i(2\nu t+\varphi)}\right)|1\rangle\langle 0| + \text{h.c.}\right]. \tag{2.14}$$

$\Omega = d_{10}\mathcal{E}_0/\hbar$ is referred to as the Rabi frequency, characterizing the rate at which the atom is stimulated to absorb and emit single photons. In the adiabatic limit $\Omega \ll \nu$, the fast oscillating terms $\sim \exp(\pm i2\nu t)$ in Eq. 2.14 can be neglected on the time scales of $1/\Omega$. Therefore, it is appropriate to apply the rotating wave approximation and neglect the fast oscillating terms. The resulting Hamiltonian reads

$$H = \frac{\hbar\Delta}{2}\sigma_z + \frac{\hbar\Omega}{2}\cos(\varphi)\sigma_x + \frac{\hbar\Omega}{2}\sin(\varphi)\sigma_y. \tag{2.15}$$

Hence, dipole transitions can be controlled via (i) the electric field strength $\mathcal{E}_0$, (ii) its phase φ, and (iii) the detuning Δ. We note that the cases (ii) and (iii) are redundant in the sense that $\Delta = \partial_t\varphi$, as can be seen by moving into the detuned rotating frame. We illustrate the resulting dynamics in Fig. 2.2. The pulse area is defined as

$$\Theta = \int_{t_0}^{t} d\tau|\Omega(\tau)|. \tag{2.16}$$

A pulse is referred to as a π-pulse if $\Theta = \pi$. For the resonant case, this induces a rotation of π around the rotational axis. It is worth noting that the effect does not depend on the time-dependence of $\Omega(t)$, but rather on its integral value.

Composed systems can be controlled by locally addressing atomic transitions. The resulting dynamics in the full space are intriguing and non-trivial, motivating the utilization of quantum optimal control techniques to design optimized control fields. In Chapter. 4 and Sect. 5 we design such optimized strategies in the context of quantum computing.

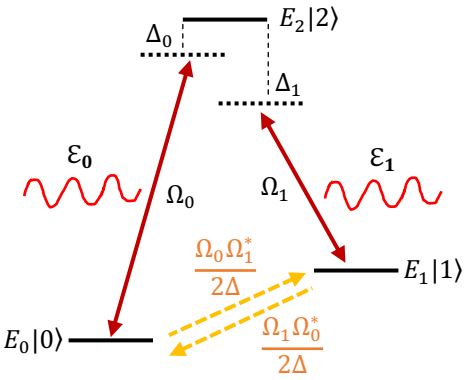


Figure 2.3: **Two-photon stimulated Raman transition.** Two hyperfine states $|0\rangle$ and $|1\rangle$ are dipole-coupled to an additional state $|2\rangle$, with the corresponding Rabi frequencies Ω_0 and Ω_1, respectively. The electric fields $\boldsymbol{\mathcal{E}}_i$ address the transitions between the states $|i\rangle$ and $|2\rangle$ (red arrows). For large detuning, population is transferred from the ground state $|0\rangle$ to the first-excited state $|1\rangle$ through the exchange of virtual photons (orange arrows).

2.4 Two-photon stimulated Raman transitions

The population between long-lived hyperfine states can be driven by the emission and absorption of virtual photons via an additional excited state [45]. To model the system, we consider two long-live hyperfine states $|0\rangle$ and $|1\rangle$, each featuring dipole-allowed transitions to an additional state $|2\rangle$. We consider the electric fields

$$\boldsymbol{\mathcal{E}}_0(\mathbf{r},t) = \frac{\mathcal{E}_0}{2}\mathbf{e}_{\nu,0}e^{-i(\nu_0 t-\mathbf{k}_0\mathbf{r}+\varphi_0)} + c.c., \tag{2.17}$$

$$\boldsymbol{\mathcal{E}}_1(\mathbf{r},t) = \frac{\mathcal{E}_1}{2}\mathbf{e}_{\nu,1}e^{-i(\nu_1 t-\mathbf{k}_1\mathbf{r}+\varphi_1)} + c.c., \tag{2.18}$$

and assume that the $\boldsymbol{\mathcal{E}}_i(\mathbf{r},t)$ selectively address the dipole transition between the state $|i\rangle$ and $|2\rangle$, for $i = 0, 1$. The electric field frequencies are $\nu_i = \omega_{2i} - \Delta_i$, where Δ_i is a detuning from the atomic transition frequency ω_{2i} between the states $|2\rangle$ and $|i\rangle$. Following the same procedure as in Sect. 2.3, the system is transformed into the rotating frame of $|2\rangle\langle 2|\,\nu_0 + |1\rangle\langle 1|\,(\nu_0 - \nu_1)$. In the limit of the rotating wave approximation, the resulting equations of motion read

$$i\dot{c}_2 = \Delta_0 c_2 + \frac{\Omega_1}{2}c_1 + \frac{\Omega_0}{2}c_0, \tag{2.19}$$

$$i\dot{c}_1 = \delta c_1 + \frac{\Omega_1^*}{2}c_2, \tag{2.20}$$

$$i\dot{c}_0 = \frac{\Omega_0^*}{2}c_2, \tag{2.21}$$

where $\delta = \Delta_1 - \Delta_0$ and $\Omega_i = d_{2i}\mathcal{E}_i e^{-i\varphi_i}/\hbar$. For large detuning $\Delta_0 \approx \Delta_1 \equiv \Delta \gg \Omega_0, \Omega_1$, the intermediate level can be adiabatically eliminated using $c_2 = -(\Omega_0 c_0 + \Omega_1 c_1)/(2\Delta)$, such that the dynamics effectively reduce to those of a two-level system. Then, the dynamics reduce approximately to the Rabi driving Eq. 2.15, but with the effective Rabi frequency

$$\Omega_{\text{eff}} = -\frac{\Omega_1^* \Omega_0}{2\Delta}, \tag{2.22}$$

and the effective detuning

$$\Delta_{\text{eff}} = \delta - \frac{|\Omega_0|^2 - |\Omega_1|^2}{4\Delta}. \tag{2.23}$$

We illustrate this system in Fig. 2.3. This type of Raman transitions is used in neutral atom quantum platforms to drive nuclear spin qubits via intermediate states [46], see Sect. 5.3.

3 Seismic noise reduction strategies in gravitational-wave interferometry

The theory of general relativity unifies the concepts of space and time into a single entity and describes gravity as a geometric property of this space-time [47]. The Einstein field equations [48] relate the distribution of mass to the curvature of this space-time. Based on the Einstein field equations, unsolved inaccuracies of classical Newtonian gravity, such as the perihelion of Mercury, could be quantified correctly. Such early verifications of general relativity have paved the way for wide adoption into modern astrophysics. Today, many cosmological objects are well understood relativistically, and have even found their way into pop-science culture, such as black holes, gravitational lenses, or binary pulsars, to name a few.

A particular solution of the Einstein field equations are gravitational waves [49], i.e. oscillations of the gravitational field that transport energy in the form of gravitational radiation. Analogous to electromagnetic radiation, which is created via moving charges, gravitational radiation is created via moving masses. Due to the weakness of the gravitational interaction, direct observation of gravitational radiation was long thought unattainable. The first indirect observation of gravitational waves was achieved by measuring the pulsing radio signal from the Hulse-Taylor binary star system [50] as its long-term oscillatory behavior can be described accurately by the system emitting gravitational radiation and, hence, loosing energy [51, 52]. The discovery of the Hulse-Taylor binary star system was honored with the Nobel Prize in Physics 1993 [53]. However, direct observation of gravitational waves promises to study phenomena beyond the scope of standard astrophysics, which is, for the most part, based on the detection of electromagnetic radiation. Additionally, gravitational waves propagate nearly undisturbed, as gravity cannot be shielded, potentially allowing for the study of far-away astrophysical objects, even before recombination. This opens the field of gravitational-wave astronomy [4] and motivates great collaborative efforts to create gravitational-wave detectors.

Gravitational waves manifest as oscillatory deformations of space that can be detected utilizing interferometers [54], which provide high-precision length measurement based on interference of laser beams. Gravitational-wave observatories, such as LIGO [5, 6] in the United States, Virgo [7, 8] in Italy, and GEO600 [9, 10] in Germany, are large-scale Michelson interferometers that aim for the detection of transient gravitational-wave signals, by measuring relative length changes of the two orthogonal optical paths. In 2015, gravitational radiation that was emitted during the black hole merger GW150914, was measured at both the LIGO observatories [55] in Hanford, WA and Livingston, LA. This first direct detection of a transient gravitational-wave signal was honored by

the Nobel Prize in Physics in 2017 [56]. In the following years, the LIGO and Virgo observatories added more events to their portfolios [57, 58], further establishing the field of gravitational-wave astronomy.

In this chapter, I discuss machine learning-assisted seismic noise reduction strategies for gravitational-wave interferometry. I provide a brief overview of the challenges of gravitational-wave astronomy. Further on, I focus the discussion on the utilization of multi-stage pendulum suspensions and control systems for the mitigation of seismic ground motion. I briefly describe the experimental platform, build by Daniel Hartwig, Jan Petermann and Roman Schnabel, that we utilize as a proof-of-principle model of a pendulum suspended test mass. I introduce the machine learning methods that we have utilized to implement a time-series forecasting model of the suspended test mass in publication [1].

3.1 Challenges in the detection of gravitational waves

Gravitational radiation manifests as contractions and expansions of space in the transverse directions of the wave's advance. Optical interferometers are used to directly detect gravitational waves by measuring the relative change of the optical path lengths of laser beams via interference. The resulting signal is referred to as the strain, a dimensionless quantity, and its amplitude is inversely proportional to the distance to its source, i.e. $\sim 1/r$. The wavelength of a gravitational wave is comparable to length scales of its source and therefore the strain contains information of acoustic nature. For more details, we refer the reader to [4].

Gravitational-wave detectors, such as LIGO and Virgo, operate in a frequency range of $\approx 10 - 10^3$Hz. In this range, the measured strain from cosmic events, such as compact binary mergers, supernovae and pulsars, are not expected to exceed $\approx 10^{-21}$ [11], imposing high precision requirements and, thus, motivating the ongoing efforts to enhance the strain sensitivity in current- [8, 14, 60, 61], and for next-generation gravitational-wave observatories [62–64]. Interferometry-based detectors are subject to quantum noise [65], restricting the measurement precision of the apparatus. However, by injecting squeezed light, the corresponding noise floor contribution can be lowered [61, 66, 67]. Environmental disturbances to the instrumentation also contribute to the noise-level and can potentially induce non-astrophysical transients into the strain [12]. An important class of environmental disturbances are of terrestrial nature [68], such as seismic ground motion and Newtonian noise [69]. For earth-located observatories, seismic noise mitigation strategies [13] are indispensable, and, from here on, we want to focus on the isolation of the mirror test masses against seismic ground motion.

The mirror test masses of the interferometers are passively isolated from seismic disturbances via multi-stage suspension systems [14, 70–72], see Fig. 3.1, as its inertia resists the motion of the suspension point at frequencies above the corresponding characteristic frequencies. However, the desired frequency band is limited by the size of the apparatus and other noise sources that cannot be shielded by this approach, such as thermal noise of the suspensions [73] and Newtonian noise [69]. Active control strategies, such

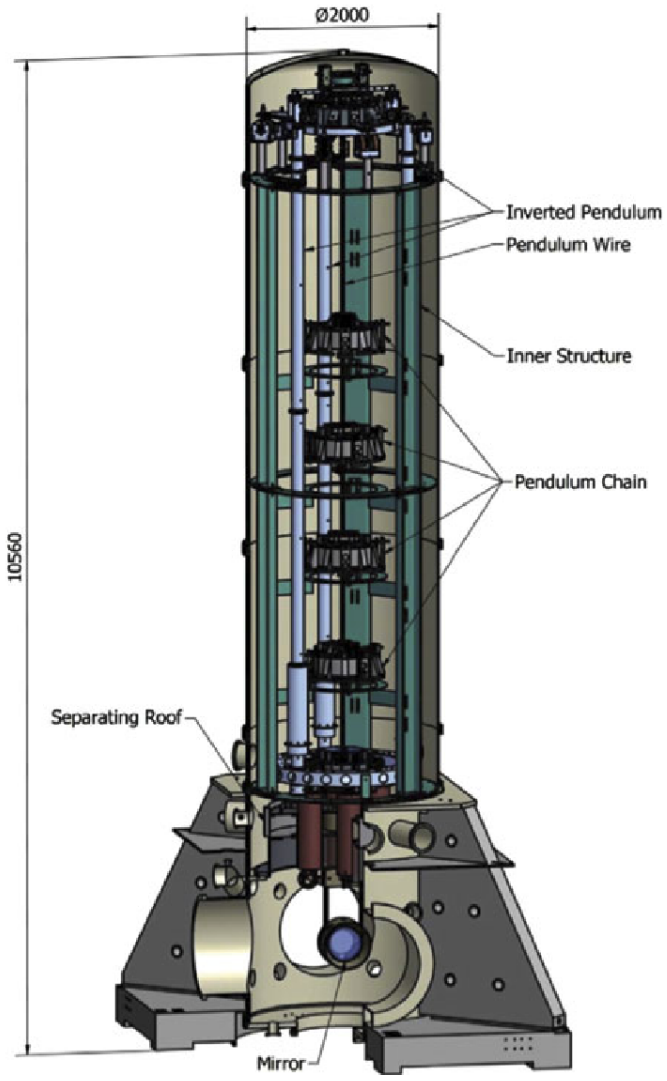


Figure 3.1: **Mirror suspension system of the Virgo detector**. The mirror test mass is suspended via a stacked multi-staged suspension system to passively isolate it from seismic ground motion above the characteristic frequencies of the system. Figure reprinted with permission from [59]. Feed-forward control further allows further isolation of the mirror test mass from witnessed seismic noise. This relies on the design of an appropriate transfer function from the witnessed seismic noise to the displacement of the pendulum suspended mass. This has motivated us to discuss machine learning-assisted seismic noise mitigation strategies of a pendulum suspended test mass, aided by witnessed seismic noise, in publication [1].

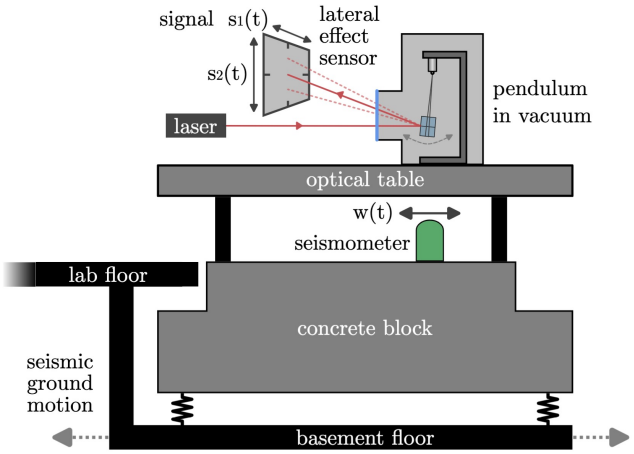


Figure 3.2: **Seismic Isolation Platform**. The proof-of-principle experimental setup that we have used in publication [1] to demonstrate machine learning-assisted multivariate time-series forecasting of the lateral effect sensor signals $s_1(t)$ or $s_2(t)$. Witnessed seismic noise $w(t)$ enhances the predictive capabilities of the model. Adapted from [1].

as feed-back control [74, 75] or feed-forward control [76], are utilized to further decouple the mirror test masses from seismic disturbances [15, 77], and to mitigate Newtonian noise [78]. Feed-forward control enables in-situ mitigation via counteracting forces, typically having less amplitude than in feed-back approaches, potentially reducing the rate of glitches [15]. The coupling between the components of such control systems are described by transfer functions, the design of which is a formidable challenge.

Machine learning techniques serve as powerful tools enhancing the field of gravitational-wave astronomy [79]. Such models have been demonstrated to enhance the strain sensitivity post-facto [78, 80, 81], reducing noise contributions of witnessed noise or instrumentation artifacts. The inherent non-linearity of machine learning models render superior to linear approaches, such as Wiener filtering [82]. Other fruitful applications, such as the classification of transients [83], gravitational-wave signal searches [84], astrophysical interpretation of gravitational-wave sources [85] and gravitational-wave form modeling [86], demonstrate the versatility of machine learning-assisted approaches. This has motivated us to utilize machine learning to predict the motion of a pendulum suspended test mass, subject to witnessed seismic noise, as a step toward machine learning-assisted feed-forward control in gravitational-wave detectors, in publication [1].

3.2 Experimental setup

In this section we elaborate on the seismic isolation platform and the experimental setup that we have used in publication [1], as a proof-of-principle model of a pendulum suspended test mass. The experimental setup was created during the PhD research

projects of Daniel Hartwig [87] and Jan Petermann [88].

At the heart of the setup is a pendulum suspended fused silica test mass inside a vacuum chamber that is located on an optical table. The main pendulum mode has a high-quality factor (high-Q) of $Q_0 = 6 \times 10^4$, and a characteristic frequency of the main pendulum mode of $\approx$ 1.4Hz. Displacements of the test mass are measured by guiding a reflected laser beam from the mirror to a lateral effect sensor that measures the horizontal and vertical light spot on the sensor surface. We note that the lateral effect sensor obeys a non-linear response, inducing non-linear instrumentation artifacts to the measured signal. Such non-linear instrumentation artifacts cannot be captured by linear transfer functions, further motivating the utilization of non-linear machine learning models, as we further elaborate on below.

The optical table is located on a concrete block with a weight of $\approx$ 30t that is suspended via helical steel springs acting as a passive isolation against seismic ground motion on the floor of the laboratory. The characteristic frequencies are $\approx$ 2Hz in the x and y directions, and $\approx$ 3Hz in the z direction. We illustrate this setup in Fig. 3.2. A triaxial seismometer is located on the platform, serving the purpose of witnessing seismic ground motion of the platform. To quantify the seismic isolation performance of the platform, a second triaxial seismometer is utilized to measure seismic ground motion of the laboratory floor for reference. Comparing the corresponding amplitude spectral densities of the two triaxial seismometers indicates that seismic ground motion suppression is achieved in the frequency band above 4Hz. Below this frequency, seismic ground motion is amplified for most frequencies. For more details, we refer the reader to [87, 88].

3.3 Artificial neural network model

Artificial neural networks are machine learning models that mimic biological neural structures via networks of perceptrons acting as feature classifiers that activate based on their receptive fields. The relations between the perceptrons are characterized by weighted connections which are adjusted (trained) with regard to a specific task, so that the model "learns" appropriate feature representations. In publication [1], we have created a multivariate time-series forecasting model, based on feed-forward neural networks [89]. In this chapter, we continue with an introduction to feed-forward neural networks. For a general introduction to the topic, we refer the reader to [90–92], which we follow in this section for a great amount.

3.3.1 Feed-forward neural networks

Feed-forwards neural networks are multi-layer perceptron architectures without circular relations and, thus, function by "feeding" values "forward", through a sequences of non-recurrent layers. We denote the perceptron values of the l-th layer as $x^{(l)}$. Commonly, the relationship between two subsequent layers is affine

$$f^{(l)}(x^{(l-1)}) = A^{(l)}x^{(l-1)} + b^{(l)}, \quad l > 0, \tag{3.1}$$

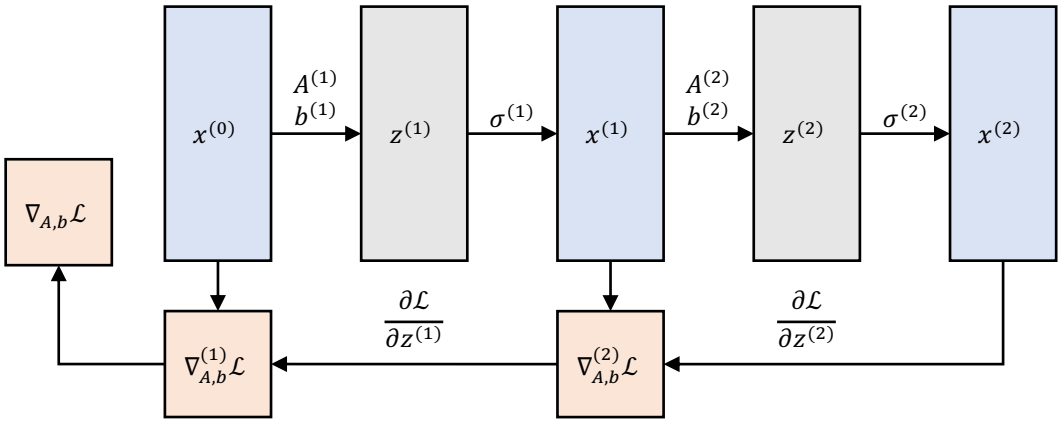


Figure 3.3: **Illustration of backpropagation in a feed-forward neural network**. The input $x^{(0)}$ is propagated "forward" through a sequence of affine maps $z^{(l)} = A^{(l)}x^{(l-1)} + b^{(l)}$, and non-linear activation functions $\sigma^{(l)}$, to compute the output $y = x^{(2)}$, for a network depth of $L = 2$. $A^{(l)}_{ij}$ and $b^{(l)}_i$ are the network weights that characterize the relations between subsequent layers. The gradient of the loss function $\nabla_{A,b}\mathcal{L}$ is calculated by propagating the errors $\partial\mathcal{L}/\partial z^{(l)}$ "backwards".

where $A^{(l)}$ is a linear transformation describing the weighted connections between the $(l-1)$-th and the l-th layer, and $b^{(l)}$ is a bias. We note that a full description of the matter utilizes the notion of tensors, however for the simplicity of the argument, we continue with the notion of vectors and matrices. The values

$$z^{(l)} = f^{(l)}(x^{(l-1)}) \tag{3.2}$$

activate the l-th layer of perceptrons by applying an activation function $\sigma^{(l)}$ element-wise. The values of the l-th layer of perceptrons read

$$x^{(l)}_i = \sigma^{(l)}(z^{(l)}_i). \tag{3.3}$$

For convenience we denote this operation as $\sigma^{(l)} \circ f^{(l)}$. The perceptive field of the perceptron $x^{(l)}_i$ defines the region in its adjacent layer $x^{(l-1)}_i$ that stimulates its activation. It is characterized by the i-th row of the linear map $A^{(l)}$. A perceptive field that is stimulated by all adjacent perceptrons is referred to as global. A layer is referred to as fully-connected, if all its perceptrons have global receptive fields.

The activation functions $\sigma^{(l)}$ serve the purpose of introducing non-linearity, allowing to compose a set of functions $\sigma^{(l)} \circ f^{(l)}$, for $0 < l \leq L$, in a non-affine manner. A feed-forward neural network can be constructed by simply composing these functions

$$F = (\sigma^{(L)} \circ f^{(L)}) \circ \cdots \circ (\sigma^{(1)} \circ f^{(1)}), \tag{3.4}$$

where L is a parameter that is referred to as the depth of the network. We denote the output as $y \equiv x^{(L)} = F(x^{(0)})$, where $x^{(0)}$ is the corresponding input. The remaining layers of perceptrons $x^{(l)}$, for $0 < l \leq L$, are referred to as hidden layers.

3.3.2 Training and backpropagation

Training is the process of adjusting the network weights $A_{ij}^{(l)}$ and $b_i^{(l)}$ to "learn" appropriate feature representations, according to a positive function, the loss function $\mathcal{L}$, quantifying inaccuracies of the classifier. An optimal set of transformation parameters is defined as a global minimum off the loss function

$$A^*, b^* = \arg\min_{A,b} \mathcal{L}. \tag{3.5}$$

Solving this minimization problem is a fundamental challenge of machine learning tasks.

A common class of methods to solve Eq. 3.5 are first order gradient-based optimization algorithms. These methods utilize the first-order gradient of the loss function $\nabla_{A,b}\mathcal{L}$, with respect to the network weights A and b, to iteratively update the network weights in the steepest direction of loss. They converge when the gradient vanishes, indicating a local (or global) extreme value or saddle point. In publication [1], we utilize stochastic gradient descent, based on mini-batches of training data, and an adaptive learning rate according to ADAM [93].

The partial derivatives of the loss, with respect to the network weights, are

$$\frac{\partial \mathcal{L}}{\partial A_{ij}^{(l)}} = \frac{\partial \mathcal{L}}{\partial z_i^{(l)}} x_j^{(l-1)}, \tag{3.6}$$

$$\frac{\partial \mathcal{L}}{\partial b_i^{(l)}} = \frac{\partial \mathcal{L}}{\partial z_i^{(l)}}. \tag{3.7}$$

$\partial \mathcal{L}/\partial z_i^{(l)}$ quantifies the error associated to the l-th layer, in the sense of deviations from optimal loss $\mathcal{L}$. Based on the chain-rule, the errors can be calculated successively, beginning from the last layer. It is

$$\frac{\partial \mathcal{L}}{\partial z_i^{(L)}} = \frac{\partial \mathcal{L}}{\partial y} \frac{\partial \sigma^{(l)}}{\partial z}\bigg|_{z_i^{(L)}}, \tag{3.8}$$

$$\frac{\partial \mathcal{L}}{\partial z_i^{(l)}} = \sum_j \frac{\partial \mathcal{L}}{\partial z_j^{(l+1)}} A_{ji}^{(l+1)} \frac{\partial \sigma^{(l)}}{\partial z}\bigg|_{z_i^{(l)}}. \tag{3.9}$$

Therefore, the gradient of the loss, with respect to the network weights, can be calculated by propagating the errors $\partial \mathcal{L}/\partial z_i^{(l)}$ backwards through the network, see Fig. 3.3. This is the back-propagation algorithm [94].

3.3.3 Convolutional neural networks

The universal approximation theorem [95] states that feed-forward neural networks are universal classifiers, in case they provide an appropriate number of perceptrons. While

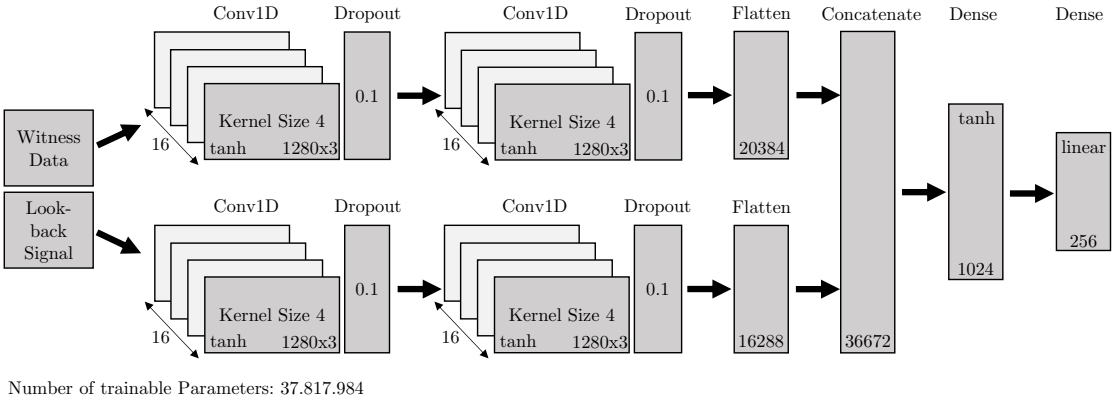


Figure 3.4: **Time-series forecasting based on convolutional neural networks.** In this feed-forward architecture, data from witness sensors (witness data) and history data (look-back signal), are processed via convolutional neural networks. The resulting outputs are concatenated into a single layer acting as an encoder to the desired output space, via a sequence of fully-connected (Dense) layers. Dropout prevents overfitting [97] by randomly dropping perceptrons, avoiding strong co-adaptions that do not generalize to data outside the training data [98]. In publication [1], we have implemented this model using Tensorflow [99]. Adapted from [1].

this guarantees the existence of an appropriate fully-connected neural network model for any learning task, the underlying global perceptive fields are not sensitive to local features, rendering this approach inefficient for classification tasks such as time-series classification, where features appear in highly correlated temporal regions at arbitrary times. Convolutional layers utilize local receptive fields and parameter sharing, forcing equivariant feature classification with respect to translations, making them suitable for time-series classification [96], which we focus on from this point forward.

The fundamental operation of a convolutional layer involves an integral transform over perceptrons $x^{(l-1)}$, with a kernel $K^{(l)}$ of finite size M, that characterizes the receptive fields of the subsequent layer $x^{(l)}$. The basic operation of a convolutional layer is

$$z_i^{(l)} = \sum_{k=i}^{i+M} K_k^{(l)} x_{i+k}^{(l-1)} + b_i^{(l)}. \tag{3.10}$$

Therefore, according to 3.1, the kernel $K^{(l)}$ is the sparse representation of the linear transformation $A^{(l+1)}$, i.e. $A_{ik}^{(l)} = K_{i+k-1}^{(l)}$ if $i + l - 1 \leq M$ and $A_{ik}^{(l)} = 0$. The network parameters are shared across the rows, which is referred to as parameter-sharing. We note that Eq. 3.10 can be extended to multiple kernels for multi-feature classification purposes, and we note that we have left out a discussion on further implementation details, such as padding and striding.

In publication [1], we have implemented a machine learning-assisted multivariate

multi-horizon time-series forecasting model [89, 100–103] that utilizes convolutional and fully-connected layers, see Fig. 3.4.

3.4 Publication I: Predicting the motion of a high-Q pendulum subject to seismic perturbations using machine learning

Nicolas Heimann, Jan Petermann, Daniel Hartwig, Roman Schnabel, and Ludwig Mathey - **Appl. Phys. Lett. 122, 254101 (2023)** [1]. Reproduced from Appl. Phys. Lett. 122, 254101 (2023), with the permission of AIP Publishing.

Motivation

This work was motivated by the ongoing efforts to mitigate seismic disturbances in multi-stage pendulum suspended test masses in gravitational-wave detectors. In a proof-of-principle setup, we have measured the motion of a high-Q pendulum subject to witnessed seismic ground motion. Our goal was to design a machine learning-assisted multivariate time-series forecasting model to predict the motion of the pendulum.

Main findings

In this work, we have studied the predictive capabilities of a machine learning-assisted multivariate time-series forecasting based on convolutional neural networks combined with fully-connected neural networks. We have compared two models: The first model (model I) has only access to the history data of the pendulums motion. The second model (model II) additionally has access to the data of the triaxial seismometer. We found that during training, model II converges to a validation loss reduced by one order of magnitude compared to model I, indicating enhanced predictive capabilities due to witnessed noise data. We have analyzed the amplitude spectral density for three cases: (i) the actual pendulum motion, (ii) the noise reduced pendulum motion, based on model I, and (iii) the noise reduced pendulum motion, based on model II. We found that both model I and model II predict the natural behavior of the pendulum, as well as instrumentation artifacts due to the non-linear response of the lateral effect sensor. We found a broadband noise reduction at low frequencies utilizing model II. We concluded that our results demonstrate the potential of machine learning-assisted corrective forward stabilization for improving the signal-to-noise ratio in next-generation gravitational-wave detectors.

Contribution

I created the underlying methodology and discussed its potential applications with DH. All authors conceptualized this work. DH, JP and RS created the proof-of-principle experiment and the corresponding figure in the manuscript. I performed the numerical analysis and the visualizations, based on the experimental dataset provided by DH, JP and RS. All authors contributed to the discussion and the interpretation of the results, as well as to writing the manuscript. This collaboration was realized under the supervision of RS and LM.

Applied Physics Letters ARTICLE pubs.aip.org/aip/apl

Predicting the motion of a high-Q pendulum subject to seismic perturbations using machine learning

Cite as: Appl. Phys. Lett. **122**, 254101 (2023); doi: 10.1063/5.0144593
Submitted: 31 January 2023 · Accepted: 3 June 2023 ·
Published Online: 21 June 2023

Nicolas Heimann,[1,2,a)] Jan Petermann,[1] Daniel Hartwig,[1] Roman Schnabel,[1] and Ludwig Mathey[1,2]

AFFILIATIONS

[1]Zentrum für Optische Quantentechnologien and Institut für Laserphysik, Universität Hamburg, Hamburg 22761, Germany
[2]The Hamburg Centre for Ultrafast Imaging, Luruper Chaussee 149, Hamburg 22761, Germany

Note: This paper is part of the APL Special Collection on Gravitational Wave Detectors.
[a)]**Author to whom correspondence should be addressed:** nicolas.heimann@uni-hamburg.de

ABSTRACT

The seismically excited motion of a high-Q pendulum in gravitational-wave observatories sets a sensitivity limit to sub-audio gravitational-wave frequencies. Here, we report on the use of machine learning to predict the motion of a high-Q pendulum with a resonance frequency of 1.4 Hz that is driven by natural seismic activity. We achieve a reduction in the displacement power spectral density of 40 dB at the resonant frequency 1.4 Hz and 6 dB at 11 Hz. Our result suggests that machine learning is able to significantly reduce seismically induced test mass motion in gravitational-wave detectors in combination with corrective feed-forward techniques.

Pendulum suspensions are used to isolate sensitive experiments from seismic and other environmental disturbances.[1–3] The inertia of the pendulum mass resists the motion of the suspension point at frequencies higher than the pendulum resonance. At resonant frequencies, the movement of the suspension point is amplified. This behavior applies to pendulum-suspended mirrors and is used to passively stabilize the optical path length in high-precision interferometry at audio-band frequencies such as gravitational-wave astronomy.[4,5] Seismic ground motion couples to the pendulum motion in two ways, via the mechanical contact and via the gravitational force due to fluctuations of the gravitational field known as Newtonian noise.[6] This presents a major challenge for the next generation gravitational-wave detectors at sub audio-band signal frequencies as Newtonian noise cannot be shielded[7] and has to be reduced by other strategies. Another source of displacement fluctuations in pendulum suspensions is thermal noise, imposing a high quality (high-Q) factor requirement on the pendulum mode.[8]

Machine learning is a broad and versatile framework for data interpretation and task optimization. Given the data intense operation of gravitational-wave detectors and the significant need to optimize measurements at a high precision, machine learning is a natural toolbox to utilize.[9] Recent applications include noise subtraction[10,11] and the classification of transients.[12] Further applications are gravitational waveform modeling,[13] gravitational-wave signal searches,[14] astrophysical interpretation of gravitational-wave sources,[15] and optimization of sensor placement for Newtonian noise cancellation is more usual.[16] Machine learning is, hence, a growing technology in gravitational-wave astronomy already serving fruitful results over a wide spectrum of challenges.

In this work, we present a machine learning based multivariate time-series forecasting model aided by witnessed seismic noise. We construct a high-Q factor pendulum on a passive isolation platform subject to environmental noise. Our model allows one to reconstruct the motion of the pendulum at frequencies below 25 Hz, and we show that utilizing witnessed seismic noise from a seismometer enhances the predictive capabilities by over a magnitude. We argue that machine learning based active stabilization offers a promising platform to enhance the signal to noise ratio in next generation gravitational-wave detectors.

At the heart of our setup is a fused silica test mass, suspended as a pendulum inside a high vacuum environment with a pendulum mode resonance frequency of $f_0 = 1.435$ Hz and a Q factor of $Q_0 = 6 \times 10^4$. The vacuum mitigates the damping due to friction of the test mass with the surrounding gas as well as coupling of acoustic disturbances. The vacuum chamber is mounted on an optical table located on a passive seismic isolation platform that extends into the basement floor. This platform is composed of a concrete block with a

mass of approximately 30 metric tons suspended on helical springs. In Fig. 1(a), we show an illustration of this setup. Near the resonant frequencies of the platform $f_s^{x,y,z}$, ground motion is amplified, which adds to the excitation of the pendulum modes. The relevant degrees of motion of the test mass are the main-, pitch-, and yaw-mode as illustrated in Fig. 1(b).

We measure the deflection angles ϕ_1 and $(\phi_2 + \phi_2')$ of a 1064 nm laser beam reflected off one surface of the test mass. This measurement is performed by guiding the reflected light to a lateral effect sensor (Thorlabs PDP90A) that measures the horizontal and vertical position of the light spot on the sensor surface, which is proportional to the deflection angles ϕ_1 and $(\phi_2 + \phi_2')$, respectively. The schematics of the signal sensing method are shown in Figs. 1(a) and 1(b). The vertical signal s_2 mainly contains contributions from the main and pitch modes while the horizontal signal s_1 is dominated by the yaw mode. However, small cross coupling contributions are possible due to non-ideal sensor alignment. Instrumentation artifacts arise since the sensor response is non-linear containing spectral contributions at higher harmonics of the resonant modes, i.e., nf_0 and nf_p for the fundamental and pitch modes, respectively, also see Fig. 4. A separate measurement is performed to estimate the sensing noise contribution to the pendulum signal. In this measurement, the path of the laser beam is altered such that it is reflected off a stationary mirror instead of the pendulum, thus containing no contribution from pendulum motion.

The seismically induced motion of the support platform is measured with a triaxial force-feedback seismometer (Nanometrics Trillium 120 QA), which outputs a signal $w(t)$ proportional to the velocity. The x- and y-axes of the seismometer measure the horizontal platform velocity perpendicular and parallel to the main pendulum motion, respectively, while the z-axis measures the vertical velocity.

All signals are digitized with a data acquisition card with 14 bit resolution at 120 samples per second.

We consider the signals $\{s_i(t)\}$ over the look-back window κ and predict the target $y(t) = s_j(t)$ over the look-ahead window τ. Witness channels $\{w_i(t)\}$ are included over the look-back and look-ahead window $\kappa + \tau$ corresponding to a scenario where the witness data are known ahead of time. In Fig. 1(c), we illustrate this data layout. The predicted evolution is, therefore, based on physical knowledge of the system from the past, deterministic instrumentation artifacts as well as witnessed noise. Unwitnessed noise or experimental noise, such as quantization noise or flicker noise, may not be covered.

Two models are employed, one has access to the signals only (model I), while the second one additionally has access to seismic witness channels (model II). For convenience, we use the following notation for an interval of a signal $c_{a:b}(t) = c(t + a)$, where $t \in [0, b - a]$. Our approach defines a multivariate multi-horizon time-series forecasting model[17] based on machine learning[18] to predict the target over the look-ahead time,

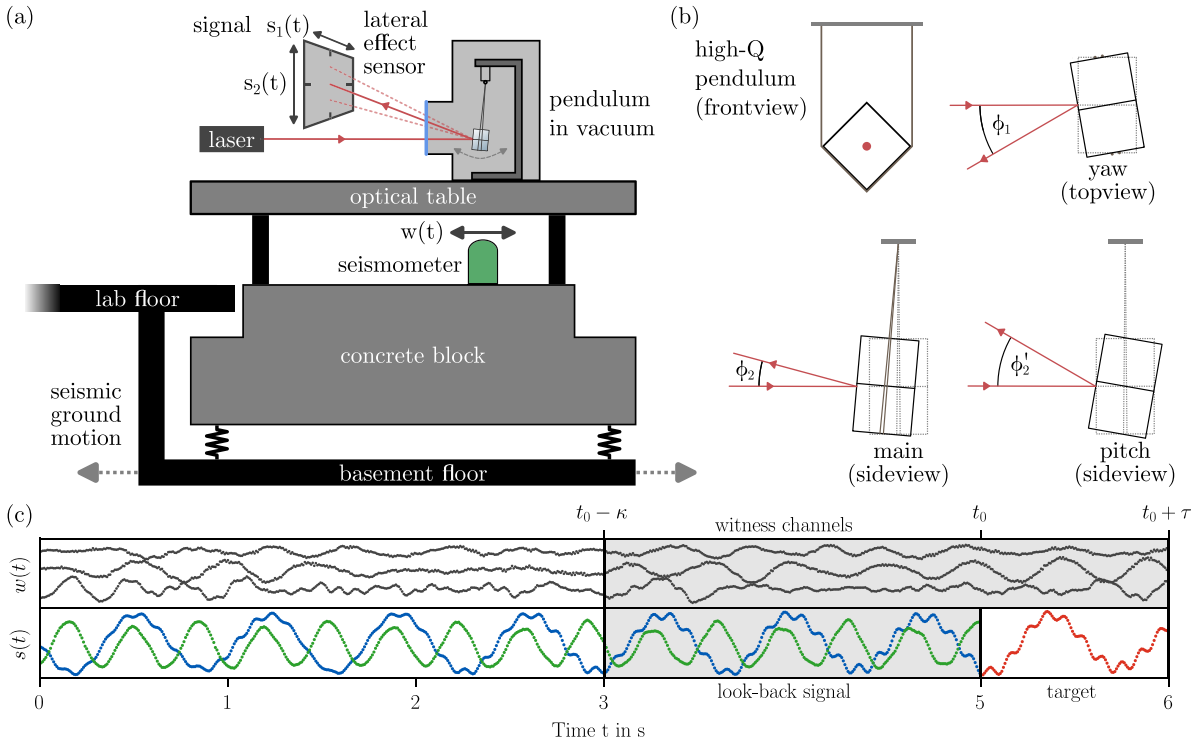


FIG. 1. Sketch of the experiment and the machine learning method. (a) The interferometer is placed on top of an optical table decoupled from the environment by a 30 metric tons suspended concrete platform. (b) Illustration of the relevant modes of the suspended test mass. The dotted lines show the pendulum in its equilibrium position. In the main pendulum mode, the entire pendulum rotates around the suspension point, such that the center of mass is moving. For the pitch and yaw modes, the test mass rotates around its center of gravity. The deflection angles ϕ_1 and $(\phi_2 + \phi_2')$ of the laser beams are measured by a lateral effect sensor. (c) The basic construction of the method invokes look-back signal $s_1(t)$ (green), $s_2(t)$ (blue) from the photodetector and seismic witness channels $w(t)$ (grey) to forecast the target signal $y(t)$ (red).

$$\hat{y}_t^{\mathrm{I}}(u) = \mathcal{F}^{\mathrm{I}}(s_{t-\kappa:t})(u), \tag{1}$$

$$\hat{y}_t^{\mathrm{II}}(u) = \mathcal{F}^{\mathrm{II}}(s_{t-\kappa:t}, w_{t-\kappa:t+\tau})(u), \tag{2}$$

where $u \in [0, \tau]$ is the look-ahead distance, $\mathcal{F}^{\mathrm{I,II}}$ are artificial neural networks (ANNs) with the associated network parameters $\theta^{\mathrm{I,II}}$, $s_{t-\kappa:t}$ is the look-back signal, and $w_{t-\kappa:t+\tau}$ are the witness channels. The windows κ, τ are set such that multiple periods of the characteristic pendulum modes are resolved, encouraging the model to learn feature representations of those. Due to the generic nature of ANNs, no further assumptions about the underlying model are required. Intermediate feature representations are learned from a data-driven training procedure with the objective to infer network parameters corresponding to local minima on high dimensional loss landscapes.

Different network species have been demonstrated to be suitable for time-series forecasting such as fully connected neural networks,[19,20] recurrent neural networks,[21–23] computational reservoirs,[24,25] or convolutional neural networks (CNNs).[26,27] In CNNs, the layers are connected by convolutional operations with parametrized kernels of fixed size.[28] A kernel is defined by local space-invariant interconnections making the inner representations of the network equivariant to translations with respect to the prediction time.[29,30] This allows for smooth translations of the prediction time, which is why we choose CNNs as the main architectural components of $\mathcal{F}^{\mathrm{I,II}}$. We set the activation of the output perceptrons to be hyperbolic tangent functions introducing non-linearity while capturing the oscillatory nature of the time-series. In training, dropout regularization is employed encouraging the network to learn sparse representations and prevent overfitting.[31] Each channel of the multivariate input is represented by an isolated CNN sub-model. Those sub-models are concatenated and post-processed by a sequence of fully connected layers having hyperbolic tangent activations as well, followed by the final layer having linear activation. Further details are provided in the supplementary material.

In this work, we use the three seismic witness channels $w_x(t)$, $w_y(t)$, and $w_z(t)$ from the seismometer, and for the look-back signal, we employ the vertical $s_1(t)$ and horizontal $s_2(t)$ signals from the photodetector, as illustrated in Fig. 1(c). We set the target to be the $s_2(t)$ signal and define the objective to be minimized during training, the loss function, to be the mean squared error of the predicted target

$$\mathcal{L}^{\mathrm{I,II}} = \frac{1}{\tau} ||\hat{y}_t^{\mathrm{I,II}} - (s_2)_{t:t+\tau}||^2. \tag{3}$$

Two disjoint datasets, the training dataset and the validation dataset,[32] are sampled from the experiment such that the prediction times are uniformly distributed encouraging to learn equivariant feature representations with respect to shifts of the prediction time. Based on the training dataset, the associated network parameters are inferred as $\theta = \mathrm{argmin}_\theta \mathcal{L}$ using stochastic gradient descent where the learning rate is dynamically adapted according to ADAM.[33] On the other hand, the validation dataset is used to define the validation-loss according to Eq. (3) allowing us to quantify the training process. We have reserved 20% of the overall record for validation. For each training iteration, the datasets are resampled and served in batches. The validation-loss of both models converges as shown in Fig. 2. Model I approaches $\mathcal{L}^{\mathrm{I}} \approx 10^{-2}$ while model II approaches $\mathcal{L}^{\mathrm{II}} \approx 10^{-3}$, demonstrating that witnessed seismic noise improves the predictive capabilities.

Due to the multi-horizon forecasting over the look-ahead window τ, there exist many predictions $\hat{y}_{t-u}(u)$ at time t corresponding to different look-ahead distances u. This motivates the definition of the prediction as the weighted average,

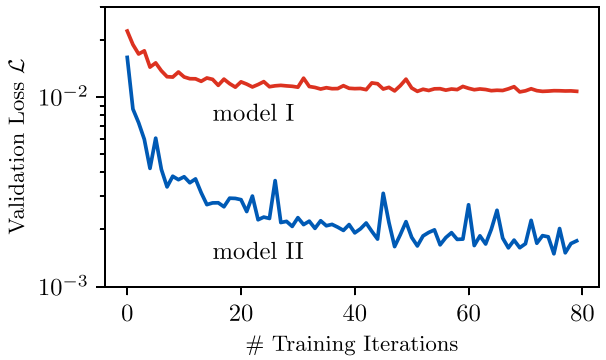


FIG. 2. Validation loss over the number of training iterations. Model I has access only to the look-back signal to form the target prediction, and model II additionally utilizes the seismic witness channels.

$$\tilde{y}^{\mathrm{I,II}}(t) = \int_0^\tau \mathrm{d}u\, x(u) \hat{y}_{t-u}^{\mathrm{I,II}}(u), \tag{4}$$

where $x(u)$ is a normalized weight function on $[0, \tau]$. For the upcoming discussions, we choose the weight function to be uniform $x(u) = 1/\tau$ so that contributions near the prediction time t as well as predictions far into the future $t + \tau$ are weighted equally. The target can be written as

$$s_2(t) = \tilde{y}^{\mathrm{I,II}}(t) + r^{\mathrm{I,II}}(t), \tag{5}$$

where $r^{\mathrm{I,II}}(t)$ is the noise-reduced signal containing unpredicted contributions.

The predictions $\tilde{y}$ according to Eq. (4) were evaluated over the validation dataset. In Fig. 3, we show a single prediction sample, where the target $s_2(t)$ is shown as well, for comparison. The predictions of both models contain periodic structures close to the expected target suggesting that the model synthesizes the phase space initial conditions from the look-back signal, allowing us to integrate the inferred dynamics to obtain the state space evolution. Due to the harmonic nature of the dynamics, we conclude that the underlying CNN acts as a Fourier transformation synthesizing the amplitudes and phases of the modes from the unperturbed motion of the pendulum. This works especially well as the network identifies a discrete number of features corresponding to sharp peaks in the spectrum due to the high-Q factor. Seismic witness data qualitatively improve the predictive capabilities suggesting that the pendulum motion is correlated with the seismic noise.

Next, we show the spectral densities of the pendulum and seismometer signals using Welch's method[34] over samples of size 2^{12}. Figure 4 shows the amplitude spectral density (ASD) of the target $s_2(f)$ and the reduced signals $r^{\mathrm{I,II}}(f) = (s_2 - \tilde{y}^{\mathrm{I,II}})(f)$ from both models according to Eq. (5). Also shown is the velocity spectral density (VSD) of the observed seismic perturbations, where the maximum is near the main pendulum mode f_0. The ASD of the target $s_2(f)$ shows pronounced peaks at the resonances f_0 and f_p. The peaks at multiples of f_0 and f_p correspond to the non-linearity

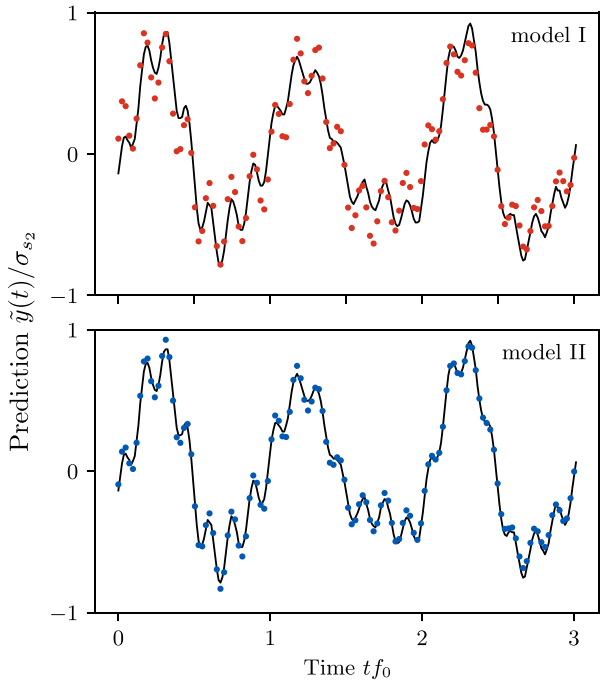


FIG. 3. Example predictions (dots) from the model without (I) and with seismic witness channel (II). The target $s_2(t)$ is depicted as a black continuous line for visual simplicity. The y-axis is in units of the standard deviations σ_{s2}.

of the photodetector, as expected for sinusoidal signals passed through a non-linear element.[35]

To compare the results that we achieve with our machine-learning based method, with an established noise reduction method, we apply linear Wiener filtering to the same data. As we discussed in the supplementary material, we find only a limited success of linear noise filtering, because only linear correlations between the witness channel and the target are included.[36]

Both our models, on the other hand, resolve those non-linear artifacts well as evident by the dips in the ratio of the ASD and the target spectral density, displayed as the ASD/target in Fig. 4(b), at all harmonics of f_0 and f_p. Such monochromatic features are well extrapolated solely from the signal look-back; as on the timescale of the prediction window, their phases and amplitudes are only affected by resonant seismic transients, which occur rarely. Off-resonant excitations, however, are time-local and, hence, not predictable from the signal look-back. Here, the advantage of using the seismic witness channels becomes evident. In the spectral region of high seismic activity (0.5–5 Hz), model I could reduce the root mean square amplitude of the signal by 71%, while utilizing witness channels, model II further lowered the amplitude by a factor of 4.

In the frequency region around the pitch mode (7–13 Hz), there is an asymmetry in the spectral density of the target. Left of the pitch mode resonance f_p, the target ASD is lower than on the right although the spectral density of the seismic background is roughly the same on both sides. Also, on the left side, model II delivered no significant improvement over model I, while on the right side around 11 Hz, witness data allow us to reduce the off-resonant ASD by a factor of 4. We assume this asymmetry is caused by an interaction between the main- and pitch-mode resonances as horizontal suspension point movement has counteracting effects on the rotation of the test mass in the two modes. This would make the reflection angle measurement insensitive to modal excitations at some frequencies.

Near 21 Hz, a peak appears both in the ASD and seismic VSD. This peak could be partially reduced by model I and removed almost completely by model II, showing that high frequency features are resolved as well.

At most frequencies, the noise-reduced spectrum of model II follows the sensing noise closely, indicating that most of the witnessed and predictable noise has been subtracted from the target. The sensing noise level is dominated by quantization noise of the data acquisition system at high frequencies, where it follows a flat line. Toward low frequencies, it shows an increase, likely due to flicker noise in the detector amplification circuits.[37]

Finally, we want to discuss the role of the weight function $x(u)$. Without any further assumption, we have estimated the parameters where model I shows the best results when only the immediate target prediction $\tilde{y}_t^{\mathrm{I}}(0)$ is considered, i.e., $x^{\mathrm{I}}(u) = \delta(u)$, while for model II, the weights decay exponentially $x^{\mathrm{II}}(u) \sim e^{-au/\tau}$. Therefore, witness data allow us to predict subsequent future states. However, finding optimal weights depends on the specific physical application, and further assumptions must be made. For example, witness data and signals might encounter phase differences due to instrumental delay or spatial separation between the measurement devices. Feed forward control also imposes model dependent requirements to the weights. Manufacturing optimal weights is, therefore, a rich and significant task depending on the application's design.

In conclusion, we have presented a machine learning based time-series forecasting model to predict the seismically excited oscillation of a pendulum having a resonance frequency of 1.4 Hz and a Q factor of 6×10^4 with the addition of a three-directional seismic witness sensor. The spectral analysis of the pendulum motion reveals that without the witness channels, machine learning can well predict the displacement amplitudes related to the pendulum resonance. The pendulum motion at off resonance frequencies is dominated by the continuous changes of the seismic field. As expected, the amplitudes at these frequencies could only be predicted when the model had access to the information in the seismic witness channels.

We conclude that our trained neural network learned both the natural behavior of the pendulum and the transfer function from the witnessed seismic noise to the displacement of the pendulum suspended mass, including instrumental artifacts such as non-linearities of the sensor used to measure the pendulum oscillation. Our approach enables flexible multivariate sensor layouts as the model learns the correlations in a model-free approach and no direct measurements of transfer functions have to be performed. The high predictability of the pendulum motion demonstrates the applicability of machine learning for feed-forward suspension control to counteract pendulum excitation through the local seismic field.

In gravitational-wave detectors, the control of pendulum suspensions is a complex effort to strike a balance between stabilization and minimization of introduced noise, requiring advanced control strategies.[38–40] The machine learning approach in combination with external sensors, as demonstrated here, can be used in the future to reduce

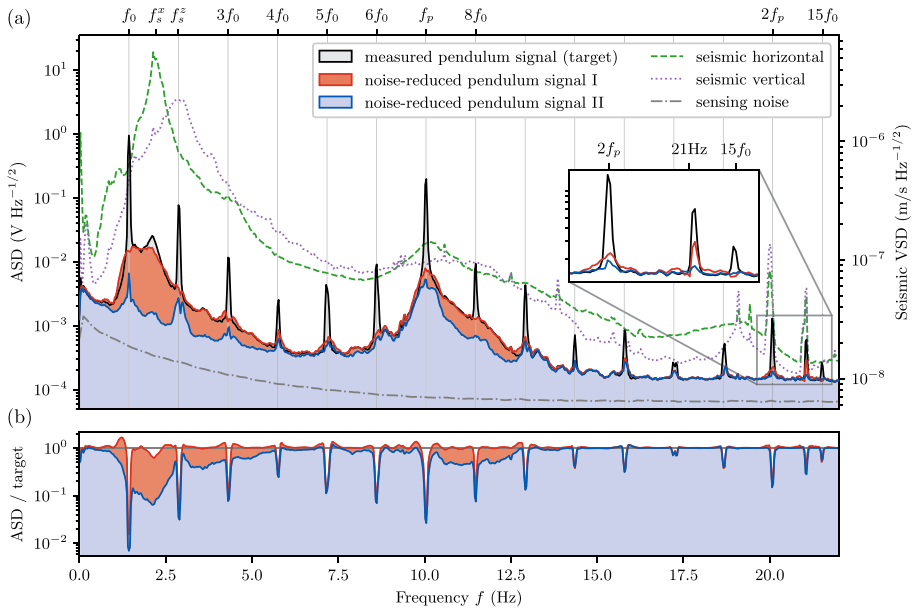


FIG. 4. (a) Amplitude spectral density (ASD) of the measured and noise-reduced pendulum motion (top of colored areas) as well as the velocity spectral density (VSD) of the seismic witness sensors (dashed and dotted lines). The noise-reduced ASDs $r^{\mathrm{I,II}}(f)$ are taken from the difference between the measured pendulum signal s_2 (target) and the predictions $\tilde{y}^{\mathrm{I,II}}$ of each model, see Eq. (5). Both models are able to reduce periodic noise at the resonances nf_0. However, the model utilizing seismic witness channels (noise-reduced pendulum signal II) allows for a more broadband noise reduction close to the sensing noise level (dashed–dotted line). The inset plot depicts high frequency noise reduction where we want to highlight the off-resonant noise reduction near the seismic peak at 21 Hz. (b) Relative ASD of the noise-reduced signals compared to the target signal. Resonant features are reduced by up to two orders of magnitude by both models, while off-resonant noise is reduced by up to one order of magnitude with model II utilizing seismic witness channels.

the seismic excited motion of a pendulum in advance. A feed-forward control loop would exploit the knowledge of the transfer function in question to correctively stabilize, for example, the pendulum's suspension point against incoming disturbances. Alternatively, it could act on the passively isolated platform on which the pendulum suspension is constructed. Or, this control loop would in our proof-of-principle experiment stabilize the plate of our optical table. In all cases, there is the significant advantage that the control loops at the end of the chain, which act directly on the position of the mass of the pendulum, have to do so with less force. It is believed that lower forces on the pendulum will lead to a reduction in the rate of non-Gaussian transients, so-called glitches.[41]

Conventional active stabilization is usually implemented with linear control systems, where the control forces acting on the system are modeled by transfer functions in frequency space.[42] Active stabilization using a machine learning approach in combination with external sensors has the advantage that non-linear disturbances can also be corrected. Alternatively, adaptive filtering techniques allow the forward transfer function to be optimized during operation.[43] In principle, even Newtonian noise can be canceled if mass displacements that cause gravity fluctuations are observed by additional sensors.[44] Our work provides the proof-of-principle that machine learning can be used to predict the motion of systems coupled to an environment. This supports the idea that machine learning-based corrective forward stabilization is a promising platform for improving the signal-to-noise ratio in next-generation gravitational-wave detectors.

See the supplementary material for a detailed description of the proposed machine learning model and a comparison of linear Wiener filtering with the proposed machine learning model.

We thank Lukas Broers and Jim Skulte for very helpful discussions. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) (SFB-925), Project No. 170620586, the Cluster of Excellence "Advanced Imaging of Matter" (EXC 2056), Project No. 390715994, and the Cluster of Excellence "Quantum Universe" (EXC 2121), Project No. 390833306.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Nicolas Tobias Heimann: Conceptualization (lead); Formal analysis (lead); Investigation (lead); Methodology (lead); Software (lead); Visualization (lead); Writing – original draft (equal). **Jan Petermann:** Data curation (equal); Visualization (supporting); Writing – original draft (supporting). **Daniel Hartwig:** Conceptualization (equal); Data curation (lead); Investigation (equal); Methodology (equal); Visualization (equal); Writing – original draft (equal). **Roman Schnabel:** Conceptualization (equal); Funding acquisition (equal);

Methodology (equal); Project administration (equal); Resources (lead); Validation (equal); Writing – review & editing (equal). **Ludwig Mathey:** Conceptualization (equal); Funding acquisition (lead); Investigation (lead); Supervision (lead); Writing – review & editing (lead).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1]N. A. Robertson, G. Cagnoli, D. R. M. Crooks, E. Elliffe, J. E. Faller, P. Fritschel, S. Goßler, A. Grant, A. Heptonstall, J. Hough, H. Lück, R. Mittleman, M. Perreur-Lloyd, M. V. Plissi, S. Rowan, D. H. Shoemaker, P. H. Sneddon, K. A. Strain, C. I. Torrie, H. Ward, and P. Willems, Classical Quantum Gravity **19**, 4043 (2002).
[2]V. Dattilo, Phys. Lett. A **318**, 192 (2003).
[3]Q. Li, C. Xue, J.-P. Liu, J.-F. Wu, S.-Q. Yang, C.-G. Shao, L.-D. Quan, W.-H. Tan, L.-C. Tu, Q. Liu, H. Xu, L.-X. Liu, Q.-L. Wang, Z.-K. Hu, Z.-B. Zhou, P.-S. Luo, S.-C. Wu, V. Milyukov, and J. Luo, Nature **560**, 582 (2018).
[4]J. Aasi, B. P. Abbott, R. Abbott *et al.*, Classical Quantum Gravity **32**, 115012 (2015).
[5]F. Matichard, B. Lantz, R. Mittleman *et al.*, Classical Quantum Gravity **32**, 185003 (2015).
[6]J. C. Driggers, J. Harms, and R. X. Adhikari, Phys. Rev. D **86**, 102001 (2012).
[7]J. Harms, L. Naticchioni, E. Calloni, R. De Rosa, F. Ricci, and D. D'Urso, Eur. Phys. J. Plus **137**, 687 (2022).
[8]G. González, Classical Quantum Gravity **17**, 4409 (2000).
[9]E. Cuoco, J. Powell, M. Cavaglià, K. Ackley, M. Bejger, C. Chatterjee, M. Coughlin, S. Coughlin, P. Easter, R. Essick, H. Gabbard, T. Gebhard, S. Ghosh, L. Haegel, A. Iess, D. Keitel, Z. Márka, S. Márka, F. Morawski, T. Nguyen, R. Ormiston, M. Pürrer, M. Razzano, K. Staats, G. Vajente, and D. Williams, Mach. Learn.: Sci. Technol. **2**, 011002 (2020).
[10]G. Vajente, Y. Huang, M. Isi, J. C. Driggers, J. S. Kissel, M. J. Szczepańczyk, and S. Vitale, Phys. Rev. D **101**, 042003 (2020).
[11]R. Ormiston, T. Nguyen, M. Coughlin, R. X. Adhikari, and E. Katsavounidis, Phys. Rev. Res. **2**, 033066 (2020).
[12]M. Razzano and E. Cuoco, Classical Quantum Gravity **35**, 095016 (2018).
[13]Z. Doctor, B. Farr, D. E. Holz, and M. Pürrer, Phys. Rev. D **96**, 123011 (2017).
[14]P. T. Baker, S. Caudill, K. A. Hodge, D. Talukder, C. Capano, and N. J. Cornish, Phys. Rev. D **91**, 062004 (2015).
[15]P. Graff, F. Feroz, M. P. Hobson, and A. Lasenby, Mon. Not. R. Astronomical Soc. **421**, 169 (2012).
[16]F. Badaracco, J. Harms, A. Bertolini, T. Bulik, I. Fiori, B. Idzkowski, A. Kutynia, K. Nikliborc, F. Paoletti, A. Paoli, L. Rei, and M. Suchinski, Classical Quantum Gravity **37**, 195016 (2020).
[17]R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications: With R Examples*, Springer Texts in Statistics (Springer International Publishing, 2017).
[18]B. Lim and S. Zohren, Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. **379**, 20200209 (2021).
[19]J. Faraway and C. Chatfield, J. R. Stat. Soc.: Ser. C (Appl. Statist.) **47**, 231 (1998).
[20]G. Zhang, Neurocomputing **50**, 159 (2003).
[21]S.-Y. Shih, F.-K. Sun, and H.-Y. Lee, Mach. Learn. **108**, 1421 (2019).
[22]S. Han, Z. Meng, X. Zhang, and Y. Yan, Micromachines **12**, 214 (2021).
[23]J. Shi, M. Jain, and G. Narasimhan, Int. J. Comput. Syst. Eng. **16**, 224 (2022).
[24]G. Tanaka, T. Yamane, J. B. Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, Neural Networks **115**, 100 (2019).
[25]S. Mandal, S. Sinha, and M. D. Shrimali, Phys. Rev. E **105**, 054203 (2022).
[26]J. B. Yang, M. N. Nguyen, P. P. San, X. L. Li, and S. Krishnaswamy, in *Proceedings of the 24th International Conference on Artificial Intelligence IJCAI'15* (AAAI Press, 2015), pp. 3995–4001.
[27]D. M. Durairaj and B. H. K. Mohan, Neural Comput. Appl. **34**, 13319 (2022).
[28]J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, Pattern Recognit. **77**, 354 (2018).
[29]W. Zhang, K. Itoh, J. Tanida, and Y. Ichioka, Appl. Opt. **29**, 4790 (1990).
[30]I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016).
[31]N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, J. Mach. Learn. Res. **15**, 1929 (2014).
[32]F. Sohil, M. U. Sohali, and J. Shabbir, Stat. Theory Relat. Fields **6**, 87 (2022).
[33]D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proceedings of the 3rd International Conference on Learning Representations (2015), arXiv:1412.6980.
[34]P. Welch, IEEE Trans. Audio Electroacoust. **15**, 70 (1967).
[35]T. J. Moir, *Rudiments of Signal Processing and Systems* (Springer International Publishing, Cham, 2022).
[36]L. R. Vega and H. Rey, "A rapid introduction to adaptive filtering," *SpringerBriefs in Electrical and Computer Engineering* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013).
[37]P. Horowitz and W. Hill, *The Art of Electronics*, 3rd ed. (Cambridge University Press, New York, 2015), Chap. 8.1.
[38]K. A. Strain and B. N. Shapiro, Rev. Sci. Instrum. **83**, 044501 (2012).
[39]S. M. Aston, M. A. Barton, A. S. Bell, N. Beveridge, B. Bland, A. J. Brummitt, G. Cagnoli, C. A. Cantley, L. Carbone, A. V. Cumming, L. Cunningham, R. M. Cutler, R. J. S. Greenhalgh, G. D. Hammond, K. Haughian, T. M. Hayler, A. Heptonstall, J. Heefner, D. Hoyland, J. Hough, R. Jones, J. S. Kissel, R. Kumar, N. A. Lockerbie, D. Lodhia, I. W. Martin, P. G. Murray, J. O'Dell, M. V. Plissi, S. Reid, J. Romie, N. A. Robertson, S. Rowan, B. Shapiro, C. C. Speake, K. A. Strain, K. V. Tokmakov, C. Torrie, A. A. van Veggel, A. Vecchio, and I. Wilmut, Classical Quantum Gravity **29**, 235004 (2012).
[40]D. Hartwig, J. Petermann, and R. Schnabel, Eng. Res. Express **4**, 045045 (2022).
[41]R. DeRosa, J. C. Driggers, D. Atkinson, H. Miao, V. Frolov, M. Landry, J. A. Giaime, and R. X. Adhikari, Classical Quantum Gravity **29**, 215008 (2012).
[42]K. J. Åström and R. M. Murray, *Feedback Systems* (Princeton University Press, Princeton, 2009).
[43]J. C. Driggers, M. Evans, K. Pepper, and R. Adhikari, Rev. Sci. Instrum. **83**, 024501 (2012).
[44]J. Harms, Living Rev. Relativity **22**, 6 (2019).

Supplemental Material: Predicting the motion of a high-Q pendulum subject to seismic perturbations using machine learning

Nicolas Heimann,[1,2,*] Jan Petermann,[1] Daniel Hartwig,[1] Roman Schnabel,[1] and Ludwig Mathey[1,2]

[1] *Zentrum für Optische Quantentechnologien and Institut für Laserphysik, Universität Hamburg, 22761 Hamburg, Germany*
[2] *The Hamburg Centre for Ultrafast Imaging, Luruper Chaussee 149, 22761 Hamburg, Germany*

[*] nicolas.heimann@uni-hamburg.de

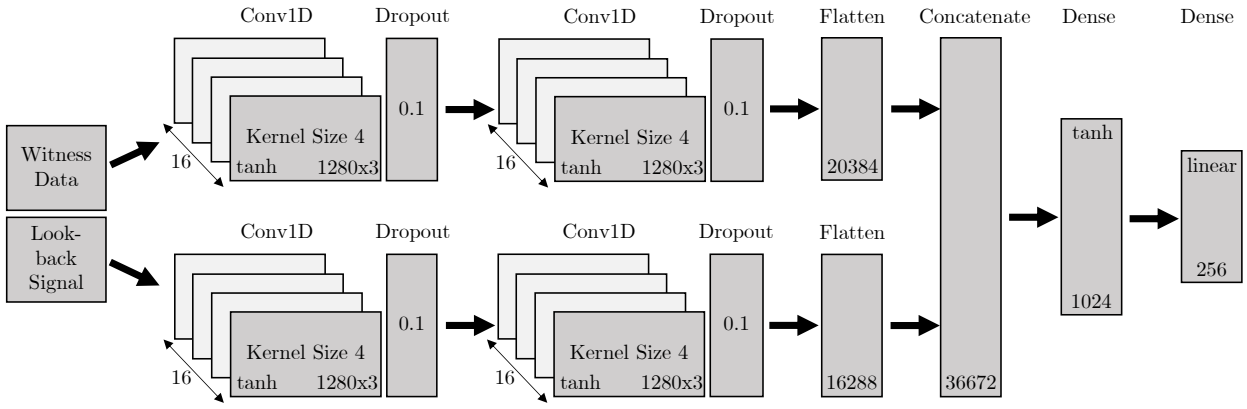


Figure S1. The ANN used in this work. The look-back signal and witness data is processed in two separate CNN sub-models. The subsequent outputs are concatenated and processed by a sequence of dense layers.

I. MODEL DETAILS

In this section we outline details of the artificial neural network (ANN) model used in this work. Let's denote the discretized temporal n-dimensional signals as $c_{t_1:t_2} \in \mathbb{R}^{\lfloor f(t_2-t_1)\rfloor \times n}$ with $t_1 < t_2$ and the sampling frequency f of c. The multi-horizon time-series target prediction of the look-back signal $s_{t-\kappa:t}$ and witness channels $\{w_{t-\kappa_w:t+\tau_w}\}$ over the look-ahead window of size n_y is

$$\hat{y}_t = \mathcal{F}(s_{t-\kappa_s:t}, \{w_{t-\kappa_w:t+\tau_w}\}) \in \mathbb{R}^{n_y}. \tag{S1}$$

where $\mathcal{F}$ is an ANN. In our specific case we have $s = (s_1(t), s_2(t))$ as the two dimensional pendulum signal and $w = (w_x(t), w_y(t), w_z(t))$ the three-directional seismic witness signal. All signals share the same sampling frequency $f = 120\,\mathrm{Hz}$ and the target prediction is

$$\hat{y}_t = \mathcal{F}^{\mathrm{II}}(s_{t-\kappa:t}, w_{t-\kappa:t+\tau}) \in \mathbb{R}^{n_\tau}. \tag{S2}$$

We set $n_\kappa = \lfloor f\kappa \rfloor = 2^{10}$ and $n_\tau = 2^8$. The prediction target is $y_t = (s_2)_{t:t+\tau}$. Temporal signals are assumed to be stationary as the pendulum is contained within a vacuum chamber and hence under stable environmental conditions (such as pressure and temperature). Seismic noise is assumed to be stationary on time scales of the experimental run. Therefore, the temporal signals are standardized as $s_{1,2}, w_{x,y,z} \sim \mathcal{N}(0,1)$ in preprocessing.

The input space is $\mathcal{X}^{\mathrm{II}} = \mathbb{R}^{n_\kappa \times 2} \times \mathbb{R}^{(n_\kappa+n_\tau)\times 3}$ with the corresponding labels $\mathcal{Y}^{\mathrm{II}} = \mathbb{R}^{n_\tau}$. The isolated model has $\mathcal{X}^{\mathrm{I}} = \mathbb{R}^{n_\kappa \times 2}$ using the same labels $\mathcal{Y}^{\mathrm{I}} = \mathcal{Y}^{\mathrm{II}}$. The datasets are sampled from the uniform distribution $U(I)$ of prediction times on the sampling interval I

$$T_I^{\mathrm{I}} = \{((s_{t-\kappa:t}), y_t) | t \sim U(I)\} \tag{S3}$$

$$T_I^{\mathrm{II}} = \{((s_{t-\kappa:t}, w_{t-\kappa:t+\tau}), y_t) | t \sim U(I)\} \tag{S4}$$

The training-dataset $T^{\mathrm{I,II}}_{[0,qt_{\max}]}$ and validation-dataset $T^{\mathrm{I,II}}_{[qt_{\max},t_{\max}]}$ are disjoint where we have $t_{\max} = 60\mathrm{h}$ and set $q = 0.8$. After each training iteration, training and validation data is resampled. The network parameters θ are inferred over a training of 500 training iterations with a batch size of 128 using a training dataset of size 10^5. We introduce the prediction at time t as the weighted sum of all target predictions containing that particular time

$$\tilde{y}_t = \sum_{j<n_\tau} X_j \cdot (\hat{y}_{t-j})_j \tag{S5}$$

where X_j are normalized weights.

$\mathcal{F}^{\mathrm{I,II}}$ was implemented using Tensorflow [1]. Witness channel data $(n_\kappa + n_\tau, 3)$ is the input of two sequential 1D convolutional layers of depth 16 and kernel size 4 having dropout of 0.1. A similar convolution is applied to the look-back signal $(n_\tau, 2)$. The output from the sub-models are concatenated and processed thru a dense layer of size $4n_\tau$ having tanh activation followed by a linear dense layer producing the target prediction of size n_τ. In Figure S1

we provide a visual representation of $\mathcal{F}^{\mathrm{II}}$. The network was trained on a single compute node providing 64GB of RAM. For our concrete model, the number of trainable parameters scales linearly in the number of sub-models, that is approximately 2×10^7 trainable parameters per sub-model. Therefore, large sensing arrays can be implemented on compute clouds providing distributed inference.

II. LINEAR FILTER

In this section, we construct a forward linear prediction Wiener filter [2] and compare it to the proposed non-linear model. We construct a linear model to predict the target $s_2(t+\tau)$ at a look-ahead time of τ utilizing history and witness data over look-back time of $\kappa = 0.5$s. The loss function is given as the mean squared error

$$\mathcal{L}_\tau = \mathbb{E}\left[||s_2(t_i+\tau) - \hat{y}(t_i+\tau)||^2\right], \tag{S6}$$

with the prediction

$$\hat{y}(t_i+\tau) = \sum_{j=0} h^s_j s(t_{i-j}) - \sum_{j=0} h^w_j w(t_{i-j}+\tau), \tag{S7}$$

where $h^s \in \mathbb{R}^{60\times 2}$, $h^w \in \mathbb{R}^{60\times 3}$ are finite impulse response filters of size 60, and $s(t) \in \mathbb{R}^2$ and $w(t) \in \mathbb{R}^3$ are the respective signals. The witness channels $w(t)$ and look-back signal $s(t)$ are as given as in the latter section. The filters h^j are inferred via stochastic gradient descent over the same dataset and training configuration as used for the non-linear model. The residual ASD $(s_2 - \hat{y})(f)$, the target ASD $s_2(f)$ and the VSD $w_i(f)$ are calculated using Welch's method [3] over samples of size 2^{12}.

We consider two examples. In the first example, the look-ahead time is set to the immediate following sample, i.e. $\tau = 1/f$. In the second example, the look-ahead time is set to $\tau = 0.2\,\mathrm{s}$, corresponding to 24 samples. Note that for the proposed non-linear model we use a look-ahead time of $\tau \approx 2\,\mathrm{s}$. Figure S2 shows the ASD of the noise-reduced signal at a look-ahead time of $\tau = 1/f$. The signal can be extrapolated easily in the low frequency spectra. Here, the linear filter reduces the ASD of the signal by 95% for 0.5–5 Hz, while the non-linear model reaches values of 92%. For 5–8 Hz and $f > 13\,\mathrm{Hz}$, the noise-reduced signal is actually enhanced compared to the target. Here, white-noise associated with the finite impulse response filter is greater than the ASD of the target. At the main pendulum mode f_0 and the pitch mode f_p, the ASD is reduced by several orders of magnitude. As expected, the higher harmonics nf_0 corresponding to non-linear instrumentation artifacts are not captured by the linear filter. For look-ahead times of $\tau = 0.2\,\mathrm{s}$, the overall ASD is less reduced, as shown in Figure S3. Here, the linear filter reduces the ASD of the signal by 86% for 0.5–5 Hz. Near the pitch mode, the reduction is less pronounced as in the latter case. The level of white noise is similar in both cases, as it depends on the filter size [2]. Hence, at high frequencies of $f > 15\,\mathrm{Hz}$, the ASD are similar for both examples. We see higher harmonics exceeding white-noise levels and find that other resonances in the seismic VSD induce additional peaks above the target ASD.

Our proposed model provides crucial improvements compared to the discussed linear model. The residual ASD is lower than the target ASD over the whole frequency domain, which is not the case in the linear approach, due to white-noise induced by the finite impulse response filters. Non-linear instrumentation artifacts are captured by the non-linear model, in contrast to the linear model. Our proposed model outperforms the predictive capabilities of the linear model as the linear model shows a less pronounced reduction of the ASD at only 10% of the look-ahead time τ used by the non-linear model. We note that we have employed a low-complexity optimization approach to infer the filter weights h. The success of this approach depends on careful tuning of the step-size used within the stochastic gradient descent method. We emphasize that a more sophisticated linear filter may be more efficient in the regard of noise-reduction.

REFERENCES

[1] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, "Tensorflow: A system for large-scale machine learning," (2016).

[2] L. R. Vega and H. Rey, *A Rapid Introduction to Adaptive Filtering*, SpringerBriefs in Electrical and Computer Engineering (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013).

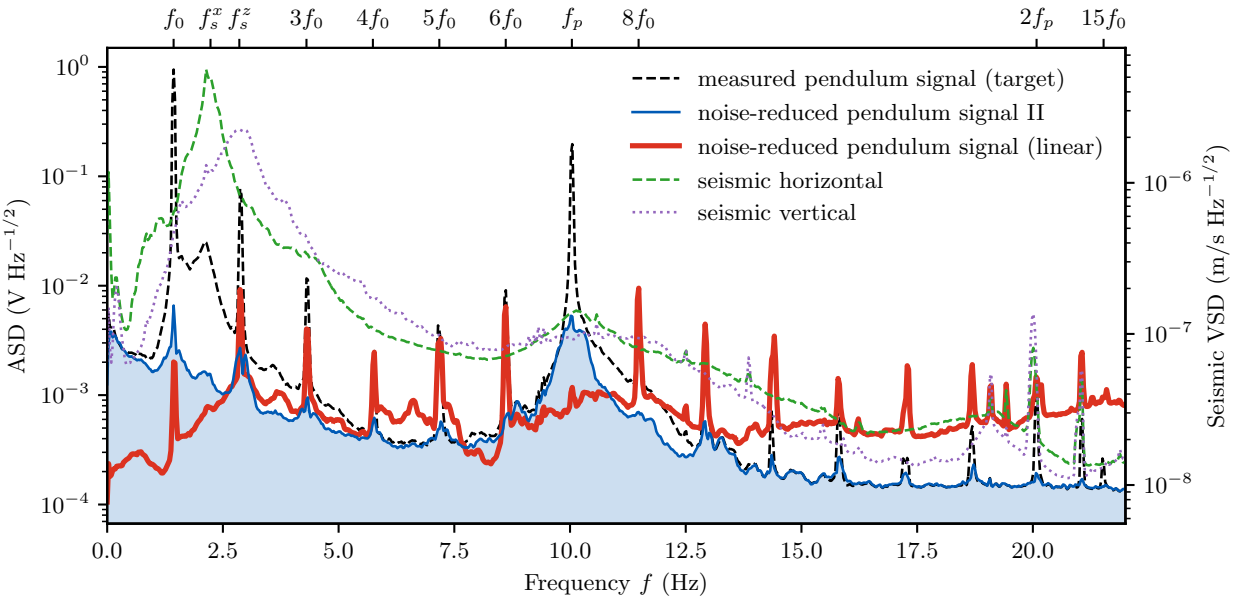


Figure S2. Residual ASD of the linear Wiener filter for the prediction target $s_2(t + 1/f)$ (red line), that is the prediction of the immediate following sample corresponding to the sampling frequency of $f = 120\,\text{Hz}$. The target signal is shown as a black-dashed line and the noise-reduced pendulum signal from the non-linear machine learning model II is shown as a blue line. The horizontal (vertical) VSD of the seismometer is depicted as a green-dashed (purple-dotted) line.

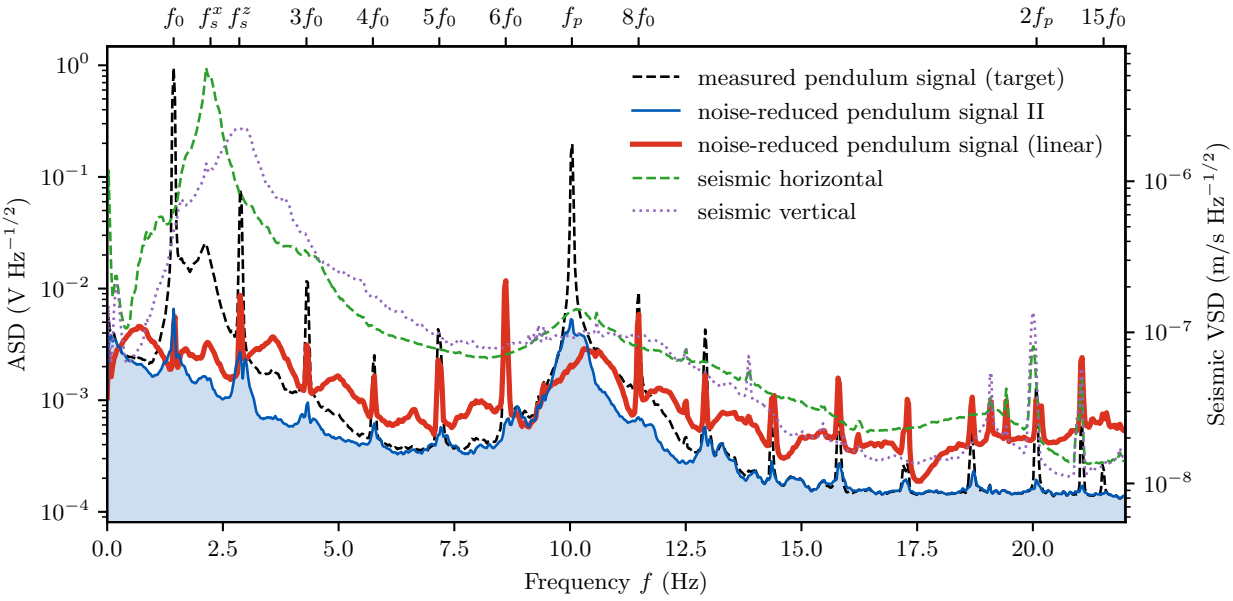


Figure S3. Residual ASD of the linear Wiener filter for the prediction target $s_2(t + 0.2\,\text{s})$ (red line). The look-ahead time of the linear Wiener filter of 0.2 s is small compared to the look-ahead time of the non-linear model, that is 2 s. The target signal is shown as a black-dashed line and the noise-reduced pendulum signal from the non-linear machine learning model II is shown as a blue line. The horizontal (vertical) VSD of the seismometer is depicted as a green-dashed (purple-dotted) line.

[3] P. Welch, IEEE Transactions on Audio and Electroacoustics **15**, 70 (1967).

4 Pulse engineering for quantum computers

Over the past decades, remarkable progress has been made toward scalable coherent technology that performs computations (or simulations) based on the laws of quantum mechanics. Such a device could simulate physical systems in a fully probabilistic manner of quantum mechanics without exponential scaling of relevant measures. Examples of these measures include the physical memory, as it must allow to capture the full probability distribution of the system [16, 104]. A universal quantum computer can, in principle, evaluate any "computable" function or simulate any physical process [105]. Moreover, any computable function can be approximated with arbitrary precision by utilizing a network model of what are today referred to as quantum gates [17]. The discovery of quantum algorithms that cannot be solved classically in a feasible amount of time renders the potential supremacy of quantum computers over classical platforms. The first quantum algorithms discovered to demonstrate this quantum advantage are the Deutsch-Josza algorithm [106], Shor's algorithm [107, 108], and Grover's algorithm [109].

In the perspective of quantum supremacy, the search for suitable [110] physical platforms to build quantum computers gathered much attention. In 1995, Cirac and Zoller proposed trapped ions as a computational platform, serving local control with high fidelity, long coherence times, and interaction via collective quantized motion based on Coloumb repulsion [111]. Shortly after, fundamental logic gates based on trapped ions where demonstrated [112–114]. Over time, other computational platforms have emerged, such as neutral atoms [25, 35, 115], superconducing qubits [116], photonic qubits [117, 118], and more [119, 120]. Coherent quantum systems are inherently sensitive to noise and thus decohere, leading to computational losses, hindering the up scaling current platforms toward quantum supremacy. Utilizing quantum error-correction [121–125], fault-tolerant and large-scale universal quantum computing might be achievable in the future.

In the current era of noisy intermediate-scale quantum (NISQ) computers [18, 126], platforms with hundreds of noisy qubits have been realized [127, 128]. Variational quantum algorithms [19, 129, 130] aim to address the constraints of NISQ-devices by utilizing a classical optimization scheme to train quantum circuits on the logical circuit level. Online learning [131] is a hardware-efficient ansatz, allowing for the design of optimized strategies that respect intrinsic characteristics of the targeted platform, such as imperfections or noise. Leaving the logical circuit level, the device-level protocols, i.e. the Hamiltonian, may be optimized directly, entering the realm of quantum optimal control [20, 22, 132]. Optimized strategies have been demonstrated for many platforms, such as ions [133–136], superconducting qubits [137–140], or neutral atoms [141–145].

In this chapter, I discuss optimal control-assisted strategies to pulse-engineer optimized protocols on digital quantum computers. I introduce the basic notion of quantum

algorithms and discuss aspects of universal quantum computing. Further, I introduce quantum optimal control and discuss the controllability of the system under control. I then link this to the field of quantum computing and elaborate on gradient-based approaches to quantum optimal control, such as gradient ascent pulse engineering, that utilizes finite difference quotients to approximate gradients. Next, I turn toward linear response theory and discuss how gradients can be estimated within this framework. Finally, I describe how we have constructed an optimal control technique by utilizing linear response theory in publication [2].

4.1 Quantum algorithms

A classical digital processor executes algorithms via a sequence of state transitions based on the input state to may produce the output state, i.e. the result of the computation. The atomic unit of information is the bit, which is a binary enumeration of 0 and 1. Throughout the computation, bit strings are transformed sequentially via logical gates, i.e. Boolean functions. Besides the utilization of quantum mechanical principles, quantum algorithms share many similarities with the classical computational model. In this section, we want to elaborate on the concept of quantum algorithms. For more information we refer the reader to [146, 147] which we follow in this section to great extend.

Quantum processors utilize coherent physical systems to perform calculations based on the laws of quantum mechanics. The initial state of the computation $|\psi_{\text{in}}\rangle$ is time-propagated to the final state $|\psi_{\text{out}}\rangle$, according to the Hamiltonian $H(t)$ which describes the underlying physical system, and encodes computational protocols. The corresponding unitary transformation reads

$$U(t_0, t) = \mathcal{T} \exp\left(-\frac{i}{\hbar}\int_{t_0}^{t} d\tau H(\tau)\right). \tag{4.1}$$

$\mathcal{T}$ is the time-ordering operator. These quantum systems can be utilized to simulate other quantum systems, such as many body systems, by engineering an appropriate Hamiltonian $H(t)$. The utilization of said quantum simulators [148] promises to enable the study of physical system that are otherwise intractable by analytic endeavors or classical simulation devices.

In a digital quantum processor, the atomic unit of information is the qubit, which is a coherent binary superposition of the states $|0\rangle$ and $|1\rangle$. This notion can be extended to n-qubit states, i.e. coherent superpositions of 2^n binary string states $|q_1\rangle \otimes \cdots \otimes |q_n\rangle$, where $q \in \{0,1\}^n$ is a binary string of size n. The n-qubit state at time $t_0 + t$ reads

$$|\psi(t_0 + t)\rangle = U(t_0, t)\,|\psi(t_0)\rangle = \sum_{q\in\{0,1\}^n} d_q(t_0 + t)\,|q_1\rangle \otimes \cdots \otimes |q_n\rangle\,. \tag{4.2}$$

$d_q(t)$ is the time-dependent coefficient of the binary string state $|q_1\rangle \otimes \cdots \otimes |q_n\rangle$. We note that the utilization of classical devices to compute the time-evolution of such a n-qubit state demands for an exponentially large physical memory to address 2^n coefficients $d_q(t)$,

fundamentally limiting classical opportunities in the simulation of physical systems. On the other hand, digital quantum processors exploit nature's hidden capacity to keep track of all coefficients throughout the entire computation, regardless of the (potentially large) size of the considered Hilbert space. To obtain information from the n-qubit state, it has to be measured. However, measurement leads to a probabilistic collapse of the state onto the measurement basis. To estimate the state before measurement, quantum state tomography [149, 150] utilizes ensembles of identically prepared states. The involved statistics require for the preparation of ensembles of identical states, as quantum states cannot be cloned from another [151].

Quantum circuits naturally generalize the notion of logic circuits known from classical computational theory [17]. The quantum pendant to a logic gate is the quantum gate, carrying a unitary transformation in $\mathrm{SU}(2^n)$ acting on n-qubit states. A quantum circuit is constituted by the composition of individual quantum gates. It reads

$$U = U_N U_{N-1} \dots U_1, \tag{4.3}$$

where N is referred to as the circuit depth, and U_i are quantum gates. An important class of quantum gates are single-qubit gates, carrying a unitary transformation in $\mathrm{SU}(2)$. Single qubit gates $U_i^{(k)}$ can be embedded into a quantum circuit in a local manner with the quantum gate $U_i = U_i^{(1)} \otimes \cdots \otimes U_i^{(n)}$. Multi-qubit entanglement gates carry a unitary transformation $\mathrm{SU}(2^m)$ acting on m-qubit states, and cannot be written as tensor products of single-qubit gates. Gates of this kind can be embedded into the quantum circuit U by acting as the identity on the remaining $n - m$ qubits of the n-qubit state.

In classical computation theory, any Boolean function can be constructed by composing simple two-bit logic gates. Remarkably, the singleton set of the NAND (or the NOR) two-bit logic gate is universal in this sense. Therefore, a classical computer is, in principle, only obliged to provide one of these two logic gates to allow for the implementation of arbitrary Boolean functions. Very similar circumstances apply to quantum circuits, where a universal set of quantum gates is constituted by a single (almost any) two-qubit entanglement gate in combination with single-qubit gates [152–154]. Hence, any unitary transformation $V \in \mathrm{SU}(2^m)$ can be approximated with arbitrary precision by a quantum circuit U based on a universal set of quantum gates. In this sense, the canonical universal entanglement gate is the controlled-NOT (CNOT) gate [155]. It has the effect of flipping the target qubit $|q_2\rangle$, depending on the value of the control qubit $|q_1\rangle$, i.e. $|q_1\rangle \otimes |q_2\rangle \to |q_1\rangle \otimes |q_1 \oplus q_2\rangle$. In the basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$, the CNOT gate has the matrix form

$$CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{4.4}$$

The choice of an appropriate and realizable entanglement gate depends on the com-

putational platform. For example, the controlled-Z gate

$$CZ_\phi = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\phi} \end{pmatrix}, \tag{4.5}$$

can be implemented naturally with trapped ions [111] or neutral atoms [35], for $\phi = \pi$. The CNOT gate can be obtained from the CZ_π gate by applying local qubit gates, i.e. $\text{CNOT} = (I \otimes H)CZ_\pi(I \otimes H)$, where H is the Hadamard gate. This sequential approach renders a constructive way to engineer quantum algorithms. However, losses accumulate, motivating the design of optimized sequences of quantum gates potentially reducing operational costs. Variational quantum algorithms utilize classical optimization strategies to train parameterized quantum circuits with regard to the computational target [19]. Due to the phenomenon of Barren Plateaus, it becomes exponentially hard to train variational quantum algorithms when scaling up algorithmic dimensions, this includes the number of qubits and circuit depth [156]. Optimization of non-local parameterizations have been demonstrated to mitigate this phenomenon [157], indicating the potential for utilization of quantum optimal control.

4.2 Quantum optimal control theory

The active manipulation of quantum dynamical processes on the atomic level, to reach desirable operational regimes via the engineering of control fields, is the subject of quantum optimal control [20, 132]. We consider the quantum optimal control Hamiltonian

$$H_\theta(t) = H_0 + \sum_{j=1}^{n_f} f_j(t) B_j. \tag{4.6}$$

H_0 describes the system that we impose control over via n_f hermitian operators B_j with the respective amplitudes of the control fields $f_j(t)$. We refer to the hermitian operators B_j as the control-operators and $f_j(t)$ as the respective control functions. The control functions $f_j(t)$ are parameterized by the transformation parameters $\theta_{j,k}$, for $0 < k \leq n_{k_j}$.

Whether a target state $|\psi^*\rangle$ is kinematically reachable from a given initial state $|\psi(t_0)\rangle$ is a question of the controllability of the system. The system is considered fully controllable if the Hamiltonian H_0 together with the control operators B_j generate a dynamical Lie algebra spanning the special unitary algebra $\mathfrak{su}(N)$, which generates the special unitary group $\text{SU}(N)$ [158]. In this case, any target unitary transformation $V \in \text{SU}(N)$ can be approximated with arbitrary precision by the unitary transformation $U_\theta(t_0, t) \in \text{SU}(N)$ generated by $H_\theta(t)$, given kinematically appropriate control functions $f_k(t)$. This connects fully controllable systems to universal quantum computing [21, 22] by pulse-engineering kinematically appropriate control functions to assemble quantum gates on the Hamiltonian level.

At the heart of the optimization task is the loss function $\mathcal{L}(\theta)$, quantifying deviations from the desired operational target. It describes a hyper-surface, the loss landscape, in

the space of the transformation parameters θ. The optimal transformation parameters θ^* are inferred via the optimization task

$$\theta^* = \arg\min_{\theta} \mathcal{L}(\theta). \tag{4.7}$$

For hybrid approaches, the transformation parameters θ are iteratively updated according to classical strategies based on gradient information of the loss-landscape obtained via feedback from the underlying quantum system [19]. Such a feedback loop can be implemented numerically, by integrating the equations of motion. Online learning approaches utilize coherent quantum devices to pulse-engineer optimized control fields via measurement-based feedback control, taking into account imperfections of the underlying physical platform [131, 159–161].

In publication [2] and publication [3], we designed optimized realizations of a target quantum gate V. We model dissipative processes via the Lindblad master equation [162]

$$\dot{\rho} = [H, \rho] + \sum_i \gamma_i \left(L_i \rho L_i^\dagger + \frac{1}{2} \left\{ L_i^\dagger L_i, \rho \right\} \right), \tag{4.8}$$

where L_i are the Lindbladians of the dissipative processes, characterized by the dissipation rate γ_i, and ρ is the density matrix of the system. We time-propagate an initial state $\rho(t_0)$ to the state $\rho(t)$ according to the master equation. We quantify the accuracy of the state $\rho(t)$ in approximating the desired target state $\rho^* = V\rho(t_0)V^\dagger$ via the state-fidelity

$$\mathcal{F}(\theta; \rho(t_0)) = \mathrm{Tr}(P\rho^* P\rho(t)). \tag{4.9}$$

P is a projector that serves the purpose of restricting the state-fidelity to a specific subspace. We consider an ensemble of n_ρ initial states $\{\rho(t_0)\}$ and estimate the accuracy of the generated transformation, in approximating the target unitary transformation V, via the ensemble average of the corresponding batch-fidelities

$$\mathcal{F}(\theta) = \frac{1}{n_\rho} \sum_{\rho(t_0)} \mathcal{F}(\theta; \rho(t_0)). \tag{4.10}$$

The corresponding infidelities are given as $1 - \mathcal{F}$.

Gradient ascent pulse engineering (GRAPE) [163] utilizes gradient ascent (or descent) to pulse-engineer optimized control fields in an iterative fashion [163]. The gradient of the loss $\mathcal{L}(\theta)$ with respect to the transformation parameter $\theta_{j,k}$ is approximated via the difference quotient

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta_{j,k}} = \frac{\mathcal{L}(\theta + \epsilon \mathbf{e}_{j,k}) - \mathcal{L}(\theta)}{\epsilon}. \tag{4.11}$$

ϵ is a hyper parameter that we refer to as the finite-difference length. An appropriate choice of ϵ must respect the specific topology of the loss landscape and is usually set to be much smaller than the typical energy scales of the system. The correction to the transformation parameter $\theta_{j,k}$ is given in the direction of steepest descent

$$\theta_{j,k} \to \theta_{j,k} - \alpha \frac{\partial \mathcal{L}(\theta)}{\partial \theta_{j,k}}. \tag{4.12}$$

α is a hyperparameter that we refer to as the learning rate. We discuss appropriate choices of the learning rate α below in Sect. 4.5. In publication [2], we implement stochastic mini-batch gradient descent, based on Eq. 4.10, for an adaptive learning rate α according to the ADAM method [93].

4.3 Linear response theory

Quantum optimal control theory can be extended to the framework of linear response theory by calculating gradients of expectation values via linear response functions [23]. Before we proceed, the framework of linear response theory is expanded upon in this section.

We consider a system that is described by the Hamiltonian $H_0(t)$. We assume the system to be subject to a perturbation in a hermitian operator B, with a time-dependent amplitude of $-\epsilon F(t)$. The Hamiltonian reads

$$H(t) = H_0(t) - \epsilon F(t) B. \tag{4.13}$$

At the time t_0, we assume that the system is in the initial state $\psi(t_0)$. At the time $t > t_0$, the state of the system is $\psi(t) = U(t_0, t)\psi(t_0)$, where $U(t_0, t)$ is the unitary transformation generated by the Hamiltonian $H(t)$ in Eq. 4.13. In the interaction picture, the state reads $\psi_I(t) = U_0^\dagger(t_0, t)\psi(t)$, where $U_0(t)$ is the unitary transformation generated by the unperturbed Hamiltonian $H_0(t)$. The Schroedinger equation of $\psi_I(t)$ implies that its dynamics are governed by the perturbation Hamiltonian in the rotating frame

$$i\hbar\partial_t\psi_I(t) = -\epsilon F(t) B_I(t)\psi_I(t). \tag{4.14}$$

We denote the corresponding unitary time-evolution as $U_I(t)$. The Heisenberg picture representation of an observable A in the rotating frame, up to first order in ϵ, reads

$$U_I^\dagger(t) A_I(t) U_I(t) = A_I(t) + \epsilon \frac{i}{\hbar} \int_0^t dt' [A_I(t), B_I(t')] F(t') + O(\epsilon^2) \tag{4.15}$$

Therefore, the expectation value of A at the time t is

$$\langle A(t) \rangle = \langle A_I(t) \rangle_{\rho(t_0)} + \epsilon \frac{i}{\hbar} \int_{t_0}^\infty dt' \chi_{AB}(t, t') F(t') + O(\epsilon^2), \tag{4.16}$$

with the susceptibility

$$\chi_{AB}(t, t') = \frac{i}{\hbar} \theta(t - t') \langle [A_I(t), B_I(t')] \rangle_{\rho(t_0)}. \tag{4.17}$$

$\langle A_I(t) \rangle_{\rho(t_0)} = \mathrm{Tr}(U_0(t_0, t)\rho(t_0)U_0^\dagger(t_0, t)A)$ is the expectation value of A at the time t, for the unperturbed case. This leads to the finite difference quotient

$$\frac{\Delta A(t)}{\epsilon} = \int_{t_0}^\infty dt' \chi_{AB}(t, t') F(t') + O(\epsilon), \tag{4.18}$$

where $\Delta A(t) = \langle A(t) \rangle - \langle A_I(t) \rangle_{\rho(t_0)}$. Hence, the gradient of the expectation value of A with respect to a perturbation in the operator B is determined by the corresponding linear response.

4.4 Gradient estimation based on time-local perturbations

We consider the Hamiltonian Eq. 4.13 for the case of a time-local perturbation $F(t) = \delta(t - t_r)$ at the perturbation time t_r. The Hamiltonian reads

$$H(t) = H_0(t) - \epsilon\delta(t - t_r)B, \tag{4.19}$$

The gradient of A with respect to the time-local perturbation in the operator B reads

$$\frac{\Delta A(t)}{\epsilon} = \begin{cases} \chi_{AB}(t, t_r) + O(\epsilon) & t_0 \leq t_r \leq t, \\ 0 & \text{otherwise.} \end{cases} \tag{4.20}$$

Therefore, the gradient of A with respect to the time-local perturbation in the operator B is determined by computing the corresponding susceptibility $\chi_{AB}(t, t_r)$ at the perturbation time. For $0 \leq t_r \leq t$, the susceptibility can be evaluated via an online approach as

$$\chi_{AB}(t, t_r) = \mathrm{Tr}(U_0(t_r, t)\delta\rho(t_r)U_0^\dagger(t_r, t)A). \tag{4.21}$$

$\delta\rho(t_r) = i\hbar^{-1}[B, U_0(t_0, t_r)\rho(t_0)U_0^\dagger(t_0, t_r)]$ is the dynamical change of the state, with respect to the perturbation operator B at the perturbation time t_r. This suggests the following kinematic procedure to compute the susceptibility for a given initial state $\rho(t_0)$. First, we time-propagate the initial state $\rho(t_0)$ to the perturbation time and calculate the dynamical change $\delta\rho(t_r)$. Second, we time-propagate the dynamical change $\delta\rho(t_r)$ to the time t. Finally, we compute the expectation value of A with respect to the time-propagated dynamical change at the time t.

Next, we decompose the delta function into the mode functions $f_k(t)$ of a complete orthogonal basis, in the sense of square-integrable functions over the time-interval $[t_0, t]$. We expand a test function $g(t)$ into this basis

$$g(t) = \int_{t_0}^{t} dt' g(t') \underbrace{\sum_k \frac{1}{\langle f_k, f_k\rangle} f_k^*(t') f_k(t)}_{\delta(t-t')}. \tag{4.22}$$

Thus, the decomposition of the delta function into the mode-function reads

$$\delta(t - t_r) = \sum_k g_k(t_r) f_k(t), \tag{4.23}$$

where we have introduced the conjugate mode functions $g_k(t) = f_k^*(\tau)/\langle f_k, f_k\rangle$, in the sense of this decomposition. Therefore, a time-local perturbation can be seen as a variation of the Hamiltonian in the perturbation operator. Specifically, the variation in the control operator B reads $H_0(t) \to H_0(t) - \epsilon \sum_k g_k(t_r) f_k(t) B$. While the decomposition of the delta function is exact, the set of mode functions may not be complete or may only approximate a complete basis set in a numerical context. To illustrate this, we calculate the finite difference quotient $\Delta A \epsilon^{-1}$ in the sense of a variation of $H_0(t)$ in the

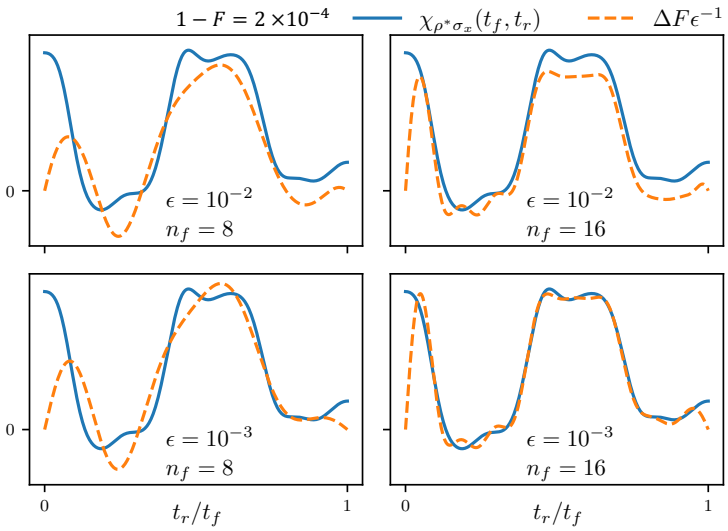


Figure 4.1: **ϵ-free estimation of gradients**. The gradient of the fidelity F, as a function of the perturbation time, for an initial state $\rho_y(0) = 1$, for the case of the Hadamard gate, as described in the supplemental material of Publication [3]. The finite-difference quotient $\Delta F \epsilon^{-1}$ is calculated in the sense of the decomposition of the delta function Eq. 4.23, and it approximates the linear susceptibility $\chi_{\rho^* \sigma_x}(t_f, t_r)$ for appropriate choices of ϵ and number of modes n_f. The susceptibility is calculated in an ϵ-free manner and does not depend on the expressibility of the parameterization.

perturbation operator B. The accuracy of the finite difference quotient $\Delta A \epsilon^{-1}$ depends on an appropriate choice of ϵ and the set of mode functions. In Fig. 4.1 we show the variational and perturbative estimate of the gradient of the fidelity for example of the Hadamard gate as described in the supplemental material of publication [3]. We note that the variational approach requires two dynamical evaluations, while the perturbative approach requires only a single dynamical evaluation. With the perturbative approach, the gradient is estimated in an ϵ-free manner and it does not depend on the expressibility of the control functions.

4.5 Parameter update via projection of response functions

In the last section, we have utilized linear response theory to estimate the gradient $\Delta A \epsilon^{-1}$ via the linear susceptibility $\chi_{AB}(t, t_r)$, for the case of a time-local perturbation in the operator B at the perturbation time t_r. Next, we utilize this to construct a correction to transformation parameters of the quantum optimal control Hamiltonian Eq. 4.6 in a gradient ascent (descent) manner. We impose control over the Hamiltonian Eq. 4.19 in the perturbation operator B over the time-interval $[t_0, t_f]$ and we consider the perturbation-time t_r within this time-interval. The Hamiltonian reads

$$H(t) = \underbrace{H_0(t) + f(t)B}_{H_\theta(t)} - \epsilon \delta(t - t_r) B. \tag{4.24}$$

The control function is $f(t) = \sum_k \theta_k f_k(t)$, where we assume that the mode functions $\{f_k\}$ are orthogonal and square-integrable over the considered time-interval.

We obtain the correction to the transformation parameters by considering the overlap of the perturbation mode functions and the parameterization mode functions. According to the decomposition of the delta function Eq. 4.23, the correction to the k-th transformation parameter is

$$d\theta_k = \alpha g_k(t_r) \chi_{AB}(t_f, t_r), \tag{4.25}$$

where the projection coefficients $g_k(t_r)$ are the conjugate mode functions at the perturbation time t_r. α is a hyperparameter that we refer to as the learning rate. Choosing an appropriate learning rate is a system specific task. In publication [3], we determine the learning rate heuristically as the maximum value that respects the system-specific topology of the loss landscape. In [164], the learning rate is determined heuristically via a parameter-specific approach. The gradient ascent (-) or descent (+) step is

$$\theta_k \to \theta_k \pm \alpha g_k(t_r) \chi_{AB}(t_f, t_r). \tag{4.26}$$

We emphasize that this is a multi-parameter update of the transformation parameters, based on a single dynamical evaluation of the linear response function $\chi_{AB}(t_f, t_r)$. Parameterizations with a large overlap between mode functions and the time-local perturbation render many non-zero projection coefficients. The multi-parameter update capitalizes on this, as the numerical complexity to compute the susceptibility does not

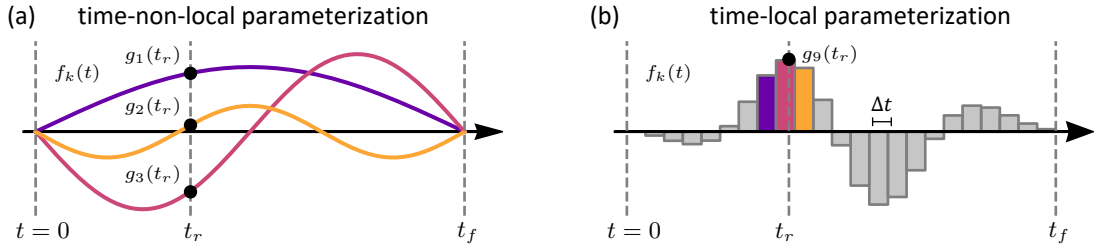


Figure 4.2: **Projection of the response onto the mode functions.** Illustration of the mode functions $f_k(t)$ and the projection coefficients $g_k(t_r)$ for a time-local perturbation at the time t_r. In panel (a), we illustrate the first three mode functions of a Fourier parametrization, and in panel (b), we illustrate the mode functions of a step-wise parametrization with a time-discretization of Δt. The time-local perturbation has large overlap with the time-non-local Fourier-modes and hence many projection coefficients $g_k(t_r)$ render non-zero. In contrast, for the case of the time-local step-wise parametrization, the only non-zero projection coefficient is $g_9(t_r)$. Panel (a) is adapted from [3].

depend on expressibility of the protocols. However, the accuracy of the multi-parameter update depends on the expressibility of the protocols, as described above.

Let us elaborate on the multi-parameter update Eq. 4.26 by discussing two different parameterizations, namely a time-non-local parameterization and a time-local parameterization. First we consider the time-non-local parameterization of Fourier-modes [157], that we also utilize in Publication [3]. Specifically, we consider a truncated set of n_f mode functions $f_k(t) = \sin(\pi k t/t_f)$, with $0 < k \leq n_f$ and the corresponding conjugate mode functions $g_k(t) = (2/t_f)\sin(\pi k t/t_f)$. The overlap between the time-local perturbation and the time-non-local Fourier-modes is large, and therefore many projection coefficients $g_k(t_r)$ render non-zero, see Fig. 4.2 (a). For comparison, we consider a step-wise parameterization with a time-discretization of Δt. The mode functions are $f_k(t) = 1$ if $k\Delta t \leq t < (k+1)\Delta t$, otherwise $f(t) = 0$, where $0 < k \leq n_f$ and $n_f = t_f/\Delta t$. The corresponding conjugate mode functions are $g_k(t) = f_k(t)/\Delta t$. In this case, a single projection coefficient $g_k(t_r)$ renders non-zero, as we show in Fig. 4.2 (b), effectively resulting in a single-parameter update.

4.6 Algorithmic procedure

We extend the Hamiltonian Eq. 4.24 to multiple control operators B_j, for $0 < j \leq n_B$, with the corresponding control functions $f_j(t)$, as given in Eq. 4.6. We consider the case of a single time-local perturbation in a single control operator B_j. Per iteration, we calculate the correction to the transformation parameters $\theta_{j,k}$ of the control function $f_j(t)$. We choose a perturbation operator randomly from the space of control operators and we choose a random perturbation time in the time-interval $[t_0, t_f]$. Specifically, the algorithmic procedure works as follows:

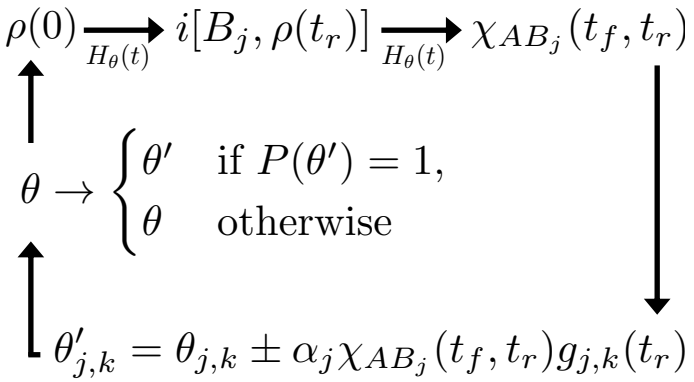


Figure 4.3: **Iterative update of the transformation parameters**. The gradient is determined by the susceptibility of the system that can be obtained via the online approach Eq. 4.21 for a given initial state $\rho(0)$. A potential new set of transformation parameters θ' is created according to the multi-parameter update rule Eq. 4.26. If the constraints-predicate $P(\theta') = 1$ then the transformation parameters are updated as $\theta \rightarrow \theta'$.

1. Determine hyperparameters: α_j.
2. Choose a random control operator B_j, for $0 < j \leq n_B$, and choose a random perturbation time t_r in the time interval $[t_0, t_f]$.
3. Choose an initial state $\rho(t_0)$.
4. Evaluate the susceptibility $\chi_{AB_j}(t_f, t_r)$, according to Eq. 4.21, for $\rho(t_0)$.
5. Calculate the potential update to the transformation parameters $\theta'_{j,k} = \theta_{j,k} \pm d\theta_{j,k}$ according to Eq. 4.26.
6. Update the transformation parameters $\theta \rightarrow \theta'$, if the constraints-predicate $P(\theta')$ evaluates to 1.
7. Go to step (2)

In Fig. 4.3 we illustrate the iterative update of the transformation parameters (3-6).

The constraint-predicate $P(\theta)$ determines whether the control functions meet additional requirements. These requirements may derive from operational boundaries of specific quantum hardware. For example, in an optical control setup, the Rabi frequency is limited by the maximal laser intensity and the strength of the dipole moment of the driven transition. As due to the probabilistic nature of the method, the algorithm does not terminate after a rejection of a potential set of transformation parameters θ'.

4.7 Publication III: Pulse engineering via projection of response functions

Nicolas Heimann, Lukas Broers, and Ludwig Mathey - **Phys. Rev. Research 7, 013101 (2025)** [3]

Motivation

This work was motivated by the growing interest in quantum optimal control methods for the development of coherent quantum technology. First order gradient-based optimal control methods, such as GRAPE, rely on finite differences and large statistics in online utilization, which both are prone to errors. We utilize linear response theory to estimate gradients due to external perturbations of the control fields in an ϵ-free manner. By projecting linear response functions onto the mode functions of the parameterization, we obtain a multi-parameter rule whose numerical complexity does not depend on expressibility of the control fields.

Main findings

In this work we have introduced pulse engineering via projection of response functions (PEPR), an iterative quantum optimal control method that utilizes linear response theory to estimate gradients via time-local perturbations of the control fields. We have demonstrated PEPR for pulse engineering of control fields, for the example of quantum gate optimization of a CNOT gate. Additionally, we have implemented standard GRAPE as a comparative benchmark. We find an improved convergence behavior and fidelity obtained with PEPR compared to standard GRAPE. Additionally, we have demonstrated PEPR in the presence of dissipation, where we find that the infidelity reaches values close to estimated lower boundaries. We have demonstrated PEPR for the case of constrained control fields via a simple predicate function. We conclude that PEPR, as an extension to gradient-based optimal control methods, will support the design and establishment of quantum technology going forward.

Contribution

LM created the methodology, LB and I contributed to the formalization and the conceptualization. I performed the numerical analysis and created the visualizations. All authors contributed to the discussion and interpretation of the results, as well as to writing the manuscript. This project was realized under the supervision of LM.

Pulse engineering via projection of response functions

Nicolas Heimann [1,2,3,*] Lukas Broers [1,2] and Ludwig Mathey [1,2,3]
[1]Zentrum für Optische Quantentechnologien, Universität Hamburg, 22761 Hamburg, Germany
[2]Institut für Quantenphysik, Universität Hamburg, 22761 Hamburg, Germany
[3]The Hamburg Centre for Ultrafast Imaging, 22761 Hamburg, Germany

(Received 20 May 2024; accepted 3 December 2024; published 27 January 2025)

We present an iterative optimal control method of quantum systems, aimed at an implementation of a desired operation with optimal fidelity. The update step of the method is based on the linear response of the fidelity to the control operators, and its projection onto the mode functions of the corresponding operator. Our method extends methods such as gradient-ascent pulse engineering (GRAPE) and variational quantum algorithms, by determining the fidelity gradient in a hyperparameter-free manner, and using it for a multiparameter update, capitalizing on the multimode overlap of the perturbation and the mode functions. This directly reduces the number of dynamical trajectories that need to be evaluated in order to update a set of parameters. We demonstrate this approach, and compare it to the standard GRAPE algorithm, for the example of a quantum gate on two qubits, demonstrating a clear improvement in convergence and optimal fidelity of the generated protocol.

DOI: 10.1103/PhysRevResearch.7.013101

I. INTRODUCTION

Algorithmic control and optimized utilization of quantum computational devices plays a central role in the research of quantum technologies. In recent years, the emergent field of quantum machine learning has brought forth new optimization heuristics that expand the body of quantum optimal control (QOC) [1–3]. One key method is gradient-ascent pulse engineering (GRAPE) [4,5], which is a QOC method that is based on estimating gradients in a space of control parameters to navigate the error surface of a given objective. As such, GRAPE has found utilization in the context of quantum computing [6–9]. Further, variational quantum algorithms (VQAs) [10,11] provide a more recent circuit-based approach to QOC [12–16] in the context of parametrized quantum circuits.

Gradient-based optimization methods, such as GRAPE, rely on finite differences and large statistics in online utilization on quantum devices, both of which are prone to errors. The multitude of dynamic trajectories (or runs) that have to be realized, either in the numerical simulation that is employed or on a real device, are a bottleneck of gradient based heuristics [17,18]. While VQAs are kept in high regard as a promising utilization of noisy intermediate-scale quantum (NISQ) devices [19,20], they have been repeatedly shown to display serious shortcomings [21–24]. This has highlighted the necessity for extensions of VQA methods [23,25–27]. In this paper, we demonstrate a pulse-engineering method based on linear-response functions of time-local perturbations for parametrizations of the Hamiltonian that are nonlocal in time. We refer to our method as pulse engineering via projection of response functions (PEPR). The objective of this method is to generate an implementation of a desired operation, such as a quantum algorithm, with optimal fidelity. We assume that the quantum system can be controlled via a set of control terms, composed of a set of operators and control functions. We expand each of these control functions into a set of mode functions, which have an arbitrary time dependence in general, and are temporally nonlocal, in particular. In each update step, we determine the response of the fidelity to one of the control operators. This generates the gradient of the fidelity in a hyperparameter-free manner. Next, we determine the optimal update of the control parameters via the projection of the response function on the modes of the control functions, resulting in a multiparameter update. These key features of the method directly address the bottleneck of growing parameter spaces in gradient-based methods, by reducing the necessary amount of dynamical trajectories to iterate over parameter updates. Another advantage is the above mentioned elimination of a hyperparameter associated with finite difference methods, improving the usability in a practical setting.

This paper is structured as follows. In Sec. II we introduce the PEPR method. In Sec. III we describe the main example to demonstrate the method, which is the optimal implementation of a CNOT gate on a two-qubit system driven by Rabi pulses and coupled via Heisenberg interaction. In Sec. IV we benchmark the optimization performance of PEPR in relation to GRAPE and discuss the optimal solutions obtained in the presence of dissipation and constraints of the system. In Sec. V we conclude our findings.

II. METHOD

The pulse-engineering method PEPR that we propose can be generally applied to conjugate sets of mode functions of the control functions for the perturbation and the parametrization to maximize the projection coefficients. Here, we opt for a

*Contact author: nheimann@physnet.uni-hamburg.de

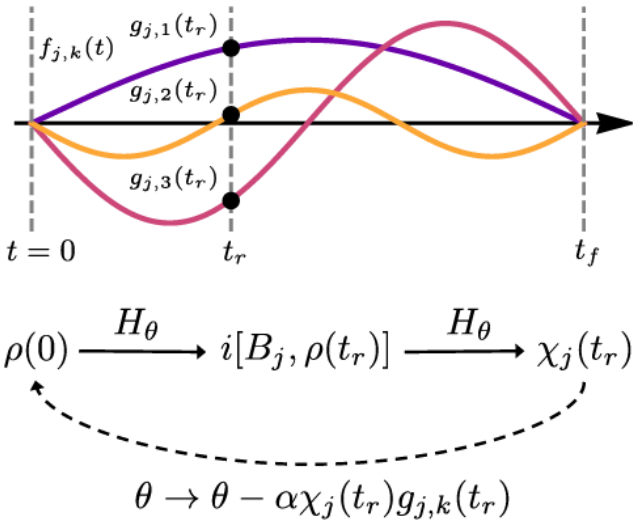


FIG. 1. Illustration of the PEPR method. Panel (a) shows temporally nonlocal mode functions $f_{j,k}(t)$ of the control functions that parametrize the time-dependent Hamiltonian. The objective is to optimize the fidelity of a desired operation, which is done via iterative updates of the trainable parameters. These updates are based on the response to a perturbation at a random time t_r, and the projection of the response onto the mode function $f_{j,k}(t)$, which uses the projection coefficients $g_{j,k}(t_r)$. In panel (b) we display a single update schematically. The random initial state $\rho(0)$ is propagated to the random time t_r. The commutator with one of the control operators B_j is taken. Then, the resulting density operator is propagated to the time t_f, and the response function $\chi_j(t_r)$ is determined. This quantity, together with the projection coefficient $g_{j,k}(t_r)$ and a learning rate α, is used for the update.

time-local perturbation and consequently a temporally nonlocal parametrization based on Fourier modes [27]. We consider this approach to be an extension of GRAPE-like methods and VQAs that utilizes nonlocal parametrizations in an efficient manner and addresses common issues, such as gradient estimation and scaling behavior of gradient-based methods.

We compare the performance of PEPR to that of standard GRAPE in a minimal example of compiling an entangling gate on two qubits. As we discuss below, we find that both the fidelity of the generated protocol, as well as the convergence towards this protocol, is improved in a convincing fashion. We note that QOC has been related to linear-response theory in Ref. [28].

The method works as follows, see Fig. 1 for an illustration. Analogously to variational and optimal control approaches, we consider a Hamiltonian of the form

$$H_\theta = H_0 + \sum_{j=1}^{n_B} \theta_j(t) B_j. \tag{1}$$

The control operators B_j represent the options of external control of the system, with n_B being the number of them. We write general parametrizations of the time-dependent control functions $\theta_j(t)$ as

$$\theta_j(t) = \sum_{k=1}^{n_j} \theta_{j,k} f_{j,k}(t), \tag{2}$$

where $\theta_{j,k}$ are parameters, and $f_{j,k}(t)$ are mode functions of the control functions of the parametrization. The term H_0 in Eq. (1) is the part of the Hamiltonian that cannot be controlled externally.

We use this Hamiltonian to propagate an initial density matrix $\rho(0)$ over the time interval of $t \in [0, t_f]$, such that we obtain the density operator $\rho(t_f)$. The objective of the algorithm is to maximize the fidelity of the state $\rho(t_f)$, compared to the target state that a target transformation V produces, i.e., $\rho^* = V\rho(0)V^\dagger$. We therefore write the state fidelity as

$$F_\theta(\rho) = \mathrm{Tr}(\rho(t_f)\rho^*), \tag{3}$$

where the subscript θ emphasizes the dependence on the parameters which determine the time evolution that produces $\rho(t_f)$. Equivalently, we aim to minimize the infidelity $1 - F_\theta(\rho)$ for any initial state $\rho(0)$. For this purpose, we approach an optimal or near-optimal implementation in an iterative fashion.

Next, we consider a perturbation of the system of the form

$$H_p = -\epsilon\delta(t - t_r)B_j. \tag{4}$$

The perturbation occurs at time t_r, with the time dependence of a δ function. B_j is one of the control operators via which the system can be controlled as presented in Eq. (1). The prefactor ϵ is utilized as a perturbative expansion parameter. We choose the time randomly in the interval $t_r \in [0, t_f]$, and we choose the operator B_j randomly, via a random choice of the index $j = 1, \ldots, n_B$.

We then follow the standard result of linear response theory, in which the change of the expectation value of an observable A, due to a perturbation of the form $H_p = -F(t)B$ is of the general form

$$\Delta\langle A\rangle = \int_{-\infty}^{\infty} dt' \chi_{AB}(t, t')F(t'), \tag{5}$$

with the susceptibility

$$\chi_{AB}(t, t') = \frac{i}{\hbar}\Theta(t - t')\langle[A_I(t), B_I(t')]\rangle, \tag{6}$$

where $A_I(t)$ and $B_I(t')$ are the operators A and B in the interaction picture, respectively.

We apply this approach to the optimization objective mentioned above. In this context, the observable A is the target state ρ^*. The time at which this observable is evaluated is t_f, such that we have $A_I(t_f) = U^\dagger(t_f)\rho^* U(t_f)$, where $U(t_f) = U(0, t_f)$ is the time-evolution operator from time 0 to time t_f. The perturbation contains one of the control operators B_j, and $F(t) = \epsilon\delta(t - t_r)$. We therefore have $B_{j,I}(t_r) = U^\dagger(t_r)B_j U(t_r)$. With this, we write

$$\begin{aligned}\chi_j(t_r) &\equiv \chi_{F_\theta,B_j}(t_f, t_r) \\ &= \frac{i}{\hbar}\mathrm{Tr}([U^\dagger(t_f)\rho^* U(t_f), U^\dagger(t_r)B_j U(t_r)]\rho(0)) \qquad (7) \\ &= \frac{i}{\hbar}\mathrm{Tr}(\rho^* U(t_r, t_f)[B_j, U(t_r)\rho(0)U^\dagger(t_r)]U^\dagger(t_r, t_f)), \qquad (8)\end{aligned}$$

where we note that $t_f > t_r$, such that $\Theta(t_f - t_r) = 1$. Furthermore, using Eq. (5) and $F(t) = \epsilon\delta(t - t_r)$, we obtain

$$\frac{\Delta F_\theta}{\epsilon} = \chi_j(t_r). \tag{9}$$

This expression shows that the gradient of the fidelity in the operator space, spanned by the control operators B_j, is determined by computing the linear response of the fidelity with regard to that operator at a time t_r.

Based on this response function, we determine the optimal update of the parameters $\theta_{j,k}$ as follows, where the index j corresponds to the operator B_j in Eq. (4). We write the projection of the δ function that is used in the perturbation in Eq. (4) onto the mode functions of the control functions of the parametrization as

$$\delta(t - t_r) = \sum_{k=1}^{\infty} g_{j,k}(t_r) f_{j,k}(t). \tag{10}$$

The $g_{j,k}(t)$ are conjugate functions to the $f_{j,k}(t)$ in the sense of a decomposition of the δ function. Based on this decomposition of the time dependence of the perturbation in terms of the mode functions of the control function of the control operator B_j, we obtain the update rule

$$\theta_{j,k} \to \theta_{j,k} - \alpha_0 g_{j,k}(t_r)\chi_j(t_r), \tag{11}$$

where we introduce the parameter α_0, which in similar contexts is referred to as a learning rate or step size. While the decomposition in Eq. (10) is exact, we use the approximation of truncating the sum by the number of modes n_j that are included in the representation in Eq. (2). With this, we identify the term $-\alpha_0 g_{j,k}(t_r)\chi_j(t_r)$ as a correction to the parameters $\theta_{j,k}$. We emphasize that $\chi_j(t_r)$ in Eq. (11) is used to update n_j-many parameters $\theta_{j,k}$ without increasing the numerical complexity of the approach, assuming that the corresponding $g_{j,k}(t_r) \neq 0$. This is in contrast to conventional variational methods, in which the complexity grows with the controllability of the Hamiltonian. Therefore, this method directly benefits from parametrizations in which the mode functions of the control functions have significant overlap with the time dependence of the perturbation, i.e., the δ functions acting at different times.

Note that $\theta_j(t)$ are functions that can be arbitrarily parametrized through choices of $f_{j,k}(t)$ in Eq. (2). Common parametrizations in optimal control contexts use time-local stepwise functions, which is adjacent to parametrized variational quantum circuit methods, or low-dimensional random bases [29,30]. Since this method is an extension of GRAPE, and relies on, and benefits from, the overlap between different functional bases for the perturbation and the parametrization of the Hamiltonian, we refer to it as pulse engineering via the projection of response functions (PEPR). Hence, a central aspect of PEPR is the overlap of parametrization-mode functions of the control functions and the perturbation-mode functions of the control functions.

As an example that implements these considerations, we choose the temporally nonlocal parametrization of Fourier modes [27], which consists of mode functions of the control functions that all have overlap with (almost) any time-local perturbation proportional to $\delta(t - t')$. This particular parametrization is

$$\theta_j(t) = \sum_{k=1}^{n_j} \theta_{j,k} \sin\left(\pi k \frac{t}{t_\mathrm{f}}\right). \tag{12}$$

This means that the decomposition of the δ function in Eq. (10) is now given through the functions

$$f_{j,k}(t) = \sin\left(\pi k \frac{t}{t_\mathrm{f}}\right), \tag{13}$$

$$g_{j,k}(t) = \frac{2}{t_\mathrm{f}} \sin\left(\pi k \frac{t}{t_\mathrm{f}}\right), \tag{14}$$

such that we obtain the update rule

$$\theta_{j,k} \to \theta_{j,k} - \alpha \sin\left(\pi k \frac{t_r}{t_\mathrm{f}}\right)\chi_j(t_r), \tag{15}$$

where we have introduced the effective learning rate $\alpha = \alpha_0 \frac{2}{t_\mathrm{f}}$.

III. MODEL

We demonstrate PEPR for a simple example of the target transformation $V = \mathrm{CNOT}$ on two qubits. The full Hamiltonian reads

$$H(t) = \sum_{j=1}^{2} h_{x,j}(t)\sigma_x^j + h_{y,j}(t)\sigma_y^j + J(t)\vec{\sigma}^1\vec{\sigma}^2, \tag{16}$$

where $\vec{\sigma}$ is the vector of Pauli matrices. The time-dependent functions are equivalent to the $\theta_j(t)$ in Eq. (2). We parametrize these functions as described in Eq. (12) such that

$$h_{x/y,j}(t) = \sum_{k=1}^{n_j} \theta_{x/y,j,k} \sin\left(\pi k \frac{t}{t_f}\right), \tag{17}$$

$$J(t) = \sum_{k=1}^{n_j} \theta_{J,k} \sin\left(\pi k \frac{t}{t_f}\right). \tag{18}$$

Next we describe two generic scenarios for optimization tasks, namely, optimization with and without constraints on the parameters.

For optimization without constraints, we begin by randomly sampling the initial parameters θ from normal distributions, i.e., $\theta_{j,k} \sim \mathcal{N}(0, 1)$. We then follow the PEPR procedure, described above, to iteratively update these parameters in order to identify parameters that produce a time evolution that implements the target transformation V. In each iteration, we initialize the state of the system in a product state $\rho(0) = \rho_1 \otimes \rho_2$, where ρ_1 and ρ_2 are random local pure states, see Supplemental Material [31] for details. We randomly choose the control operator $B_j \in \{\sigma_x^1, \sigma_y^1, \sigma_x^2, \sigma_y^2, \vec{\sigma}^1\vec{\sigma}^2\}$ and evaluate the susceptibility $\chi_j(t_r)$ for a random time $t_r \in [0, t_f]$. We then update the parameters according to Eq. (15).

For optimization with constraints, we discuss the example that the amplitude of the Rabi pulses is smaller than an upper bound $\Omega_{\max}$, so we demand $|h_{x,j}(t) - ih_{y,j}(t)| < \Omega_{\max}$. Similarly, we demand that the magnitude of the Heisenberg coupling $J(t)$ is smaller than an upper bound $J_{\max}$, namely, we demand $|J(t)| < J_{\max}$. We implement the constraints for PEPR and GRAPE as follows.

In an update step of PEPR, we generate the potential update of the parameters, and then check if the new set of parameters violates the constraints or not. If they do, we discard the potential update, and generate a new potential update. If the new set of parameters fulfills the constraints, the parameter update is accepted. To generate initial values for the parameters that fulfill the constraints, we first sample the parameters θ from normal distributions, i.e., $\theta_{x/y,j,k} \sim \mathcal{N}(0,1)$ and $\theta_{J,k} \sim \mathcal{N}(0,1)$. We then check if these potential initial values fulfill the constraints. If they do not, we rescale the parameters via $\theta_{x/y,j,k} \to \theta_{x/y,j,k}/\max(1, \max_t(|h_{x,j}(t) - ih_{y,j}(t)|)/\Omega_{\max})$, and $\theta_{J,k} \to \theta_{J,k}/\max(1, \max_t(|J(t)|)/J_{\max})$.

For an update step of GRAPE, we rescale the control functions via $h_{x/y,j}(t) \to h_{x/y,j}(t)/\max(1, |h_{x,j}(t) - ih_{y,j}(t)|/\Omega_{\max})$ and $J(t) \to J(t)/\max(1, |J(t)|/J_{\max})$. This truncates the maximal value of $|h_{x,j}(t) - ih_{y,j}(t)|$ and $|J(t)|$ at the values of $\Omega_{\max}$ and $J_{\max}$, respectively. To generate initial values for the parameters theta for the GRAPE method, we sample them randomly via $\theta_{x/y,j,k} \sim \mathcal{N}(0,1)$ and $\theta_{J,k} \sim \mathcal{N}(0,1)$. Then, we rescale them in the same way as for the PEPR algorithm.

The difference of implementing the constraints in the GRAPE method derives from PEPR being a stochastic method, such that a rejection of a potential update does not terminate the algorithm, while GRAPE is a deterministic method, in which a rejection terminates the algorithm. Thus, we implement constraints in GRAPE differently, as described.

Up to this point, we have described a method that operates using a state fidelity. By averaging over these fidelities we obtain a metric that is representative of the fidelity of the unitary transformation realized through the time evolution. We therefore approximate the infidelity of the transformation by considering n_F trajectories $\{\rho\}$, obtained from different initial states $\rho(0)$, and averaging over their state fidelities, such that

$$1 - F_\theta = 1 - \frac{1}{n_F} \sum_{\{\rho\}} F_\theta(\rho). \tag{19}$$

$F_\theta(\rho)$ is the state fidelity according to Eq. (3), where the subscript emphasizes the dependence on the parameters θ. Here we choose an empirical sampling size of $n_F = 10$. The result of this method, as well as similar optimization approaches, is intrinsically probabilistic. In particular, the random initial parameters θ are the starting point for the local optimization sequence within the parameter space. In order to present a meaningful comparison, we show many realizations of this algorithm with n_T different initial parameters $\{\theta\}$. We also show the average infidelity over this ensemble of optimization trajectories. We write

$$\langle \log_{10}(1 - F) \rangle = \frac{1}{n_T} \sum_{\{\theta\}} \log_{10}(1 - F_\theta). \tag{20}$$

Here we choose an empirical sampling size of $n_T = 100$. Note that this is the log-mean of the infidelity, which gives higher weight to low-infidelity realizations. We choose this more-involved average as visual support for the set of trajectories in Fig. 2, see below.

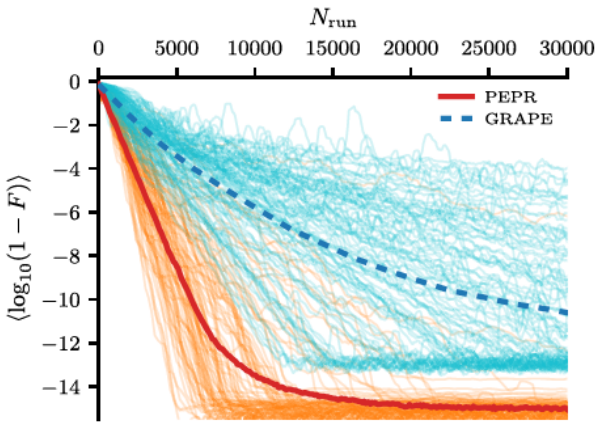


FIG. 2. PEPR vs GRAPE. The average infidelity $\langle \log_{10}(1 - F) \rangle$, as a function of the number of runs N_{run}, obtained using PEPR (red line) and GRAPE (blue dashed). The underlying ensembles of realizations $\{1 - F_\theta\}$ for PEPR and GRAPE are depicted as orange and cyan lines, respectively. PEPR leads to faster convergence with a reduced variance of the underlying realizations, compared to GRAPE. Furthermore, the infidelity that is achieved is consistently smaller for PEPR than for GRAPE. The lower bound of the average infidelity is determined by the accuracy of the used numerical method. Due to the additional error of the gradient estimate, this is more detrimental for GRAPE.

IV. RESULTS

In this section we demonstrate our method in the case of the target transformation of the CNOT gate. As a comparative benchmark, we additionally optimize the parameters based on the GRAPE method, see Supplemental Material [31] for details. In the version of GRAPE used here, we update all the parameters in a single iteration. Therefore, GRAPE requires calculating $5n_j + 1$ time evolutions to estimate the gradient of the infidelity in a single iteration. We emphasize this with regards to our definition of the number of runs N_{run}, which is the number of calculated time evolutions during a sequence of parameter updates.

We first give an example for optimization without constraints. In Fig. 2 we show results of the optimization using both PEPR and GRAPE. We use an optimized set of hyperparameters $\alpha_{\text{PEPR}} = 0.5$, $\alpha_{\text{GRAPE}} = 1.2$, and $\epsilon = 10^{-7}$ for each method to provide an unbiased comparison, as discussed in the Supplemental Material [31]. The optimization trajectories generated using PEPR converge faster to low values of the infidelity and show a reduced variance for a given number of runs N_{run}, compared to GRAPE. The lower bound is determined by the accuracy of the numerical integration method. Here, we use standard fourth-order Runge-Kutta with a time discretization of $h = 2^{-14} t_f$. The lower bound of the optimization trajectories using GRAPE additionally depends on the finite difference length ϵ. Note that the standard GRAPE results show a large variance in the quality of optimization trajectories, compared to PEPR, where we more reliably find fast-converging high-fidelity solutions. The average infidelity over the ensemble of optimization trajectories $\langle \log_{10}(1 - F) \rangle$ also reflects this convergence behavior. We note that the variance $\text{Var}(1 - F)$ obtained using PEPR is reduced,

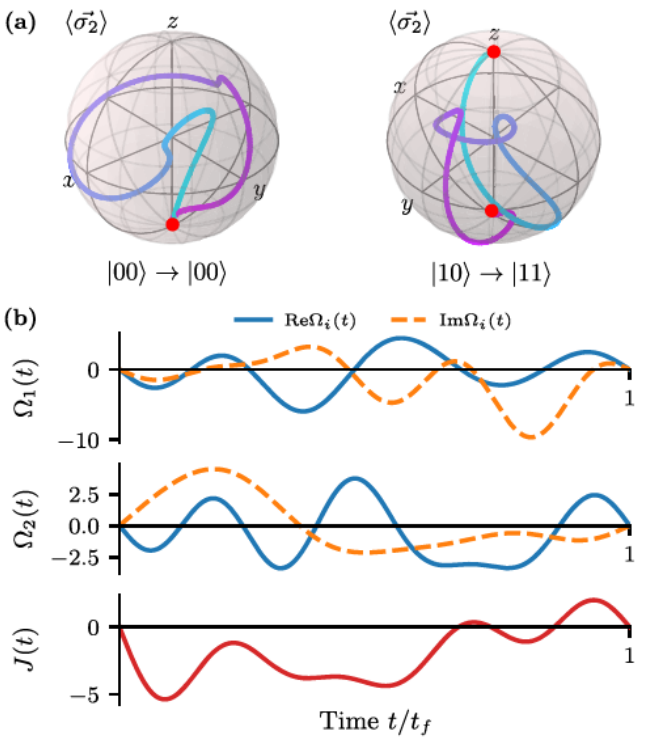


FIG. 3. PEPR-based high-fidelity protocol. In panel (a) we show the time evolution of the Bloch vector of the target qubit $\langle\vec{\sigma}\rangle$ under a high-fidelity implementation of the CNOT transformation for $|00\rangle \to |00\rangle$ and $|10\rangle \to |11\rangle$. The time t is depicted as the line color and the initial and final states are depicted as red dots. In panel (b) we show the corresponding control functions $\Omega_i(t) = h_{x,i}(t) - ih_{y,i}(t)$ and $J(t)$. We note that the amplitudes $|\Omega_i(t)|$ and $J(t)$ are within reasonable boundaries for the given energy scales of the system, defined by the final time t_f.

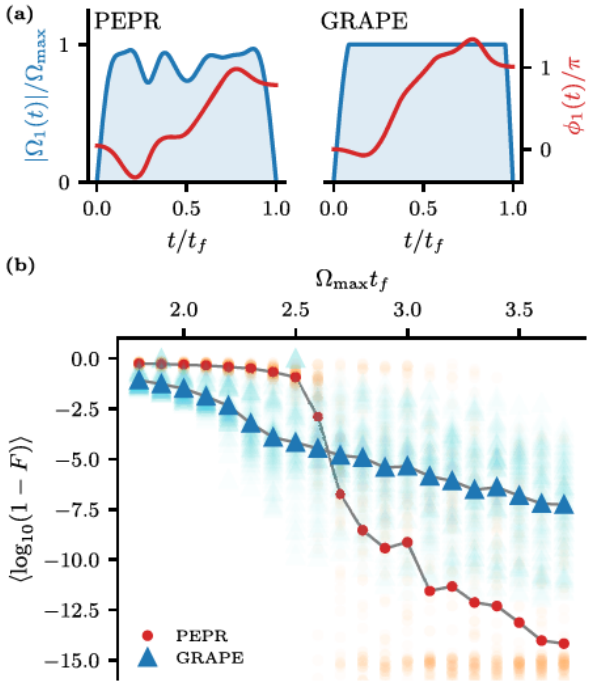


FIG. 4. Optimization under constraints. In panel (a) we show high-fidelity Rabi protocols of the first qubit $\Omega_1(t) = |\Omega_1(t)|e^{i\phi_1(t)}$ obtained using PEPR and GRAPE optimization under constraints. Specifically, we use $\Omega_{\max}t_f = J_{\max}t_f = 2.7$, and $N_{\mathrm{run}} = 30\,000$. In both cases, we obtain protocols close to phase-only protocols as the corresponding pulse areas are close to the maximal value of $\Omega_{\max}t_f$. In panel (b) we show the average infidelity $\langle\log_{10}(1-F)\rangle$ as a function of the available pulse area $t_f\Omega_{\max}$ and with a maximal interaction strength of $J_{\max} = \Omega_{\max}$, for $N_{\mathrm{run}} = 30\,000$, obtained using PEPR (red dots) and GRAPE (blue triangles). The underlying ensembles of realizations $\{1 - F_\theta(N_{\mathrm{run}})\}$ obtained using PEPR and GRAPE are depicted as orange dots and cyan rectangles, respectively.

compared to GRAPE, as discussed in the Supplemental Material [31].

We show an example for a high-fidelity implementation of the CNOT gate generated via PEPR in Fig. 3. We show the time evolution of the Bloch vector of the target qubit $\langle\vec{\sigma}^2(t)\rangle = (\langle\sigma_x^2(t)\rangle, \langle\sigma_y^2(t)\rangle, \langle\sigma_z^2(t)\rangle)^T$ for the example of the initial states $|00\rangle$ and $|10\rangle$. The trajectory of $\langle\vec{\sigma}^2(t)\rangle$ is continuous and efficient on the time scale of t_f. We note that the amplitudes of the control functions are within reasonable boundaries, resulting in a realistic protocol.

As a second example, we demonstrate optimization under constraints. We use the constraint that we discussed above, i.e., $|h_{x,j}(t) - ih_{y,j}(t)| < \Omega_{\max}$ and $|J(t)| < J_{\max}$. In Fig. 4(a) we show the control functions $\Omega_1(t) = h_{x,1}(t) - ih_{y,1}(t)$ of high-fidelity protocols, obtained using PEPR and GRAPE, for $N_{\mathrm{run}} = 30\,000$ runs, and for $\Omega_{\max}t_f = J_{\max}t_f = 2.7$. The infidelity for the PEPR method is 7.64×10^{-7}, and for GRAPE it is 1.12×10^{-4}, for this example. In both cases, the pulse area $\Theta_j = \int_0^{t_f} |\Omega_j(t)|dt$ is close to the maximal value of $\Omega_{\max}t_f$, therefore the protocols approximate phase-only protocols, i.e., $\Omega_1(t) \approx \Omega_{\max}e^{i\phi_1(t)}$.

To elaborate on the properties of protocols under constraints, we show the infidelities obtained after $N_{\mathrm{run}} = 30\,000$, in Fig. 4(b), and the average infidelity, as a function of $\Omega_{\max}t_f$. For the constraints we choose $\Omega_{\max} = J_{\max}$. We find that for PEPR, the infidelities are high up to around $\Omega_{\max}t_f \approx 2.6$. For larger values, the infidelities fall of quickly to small values of 10^{-15}, constrained primarily by the numerical accuracy of the ODE solver. In comparison, we find that GRAPE has lower average infidelities for $\Omega_{\max}t_f < 2.6$, while showing a large variance of the obtained infidelities, but has larger infidelities for $\Omega_{\max}t_f > 2.6$. This demonstrates that PEPR has an intrinsically better performance in finding high-quality implementations.

To expand the scope of the method, we apply it to a system with dissipation. As a concrete example, we include dephasing with the dissipation rate γ_z, via the Lindblad master equation

$$\frac{\partial\rho}{\partial t} = -\frac{i}{\hbar}[H_\theta(t), \rho] + \gamma_z \sum_i \mathcal{D}[\sigma_z^i]\rho. \tag{21}$$

We apply the PEPR method to this system, as described above. In Fig. 5 we show the average infidelity as a function of the number of runs N_{run} for different values of the dissipation rate γ_z. For comparison, we reproduce the optimization process for vanishing dissipation from Fig. 2, displayed in red. In all cases, the average infidelity $\langle\log_{10}(1-F)\rangle$ converges

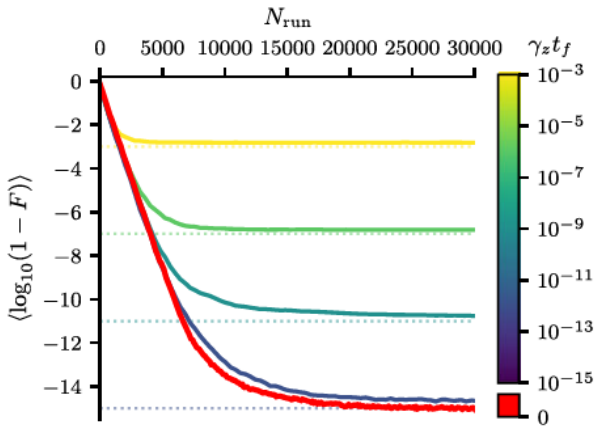


FIG. 5. Optimization for a system with dissipation. The average infidelity $\langle \log_{10}(1-F) \rangle$, as a function of the number of runs N_{run}, obtained via PEPR, for different values of the dissipation rate γ_z. In the initial stage of optimization, the average infidelity is equal to the average infidelity of the nondissipative case (red), which reproduces the data shown in Fig. 2. In the later stage of optimization, the average infidelity converges to values close to the estimated lower bound of $\gamma_z t_f$ (dotted).

for $N_{\text{run}} \gtrsim 2\times10^4$ runs. We observe that the dissipation rate introduces a lower bound on the infidelity, as indicated by the dashed lines, which are $\gamma_z t_f$. However, we find that the method generates optimal protocols in the presence of dissipation, in addition to the dissipationless case discussed above.

V. CONCLUSION

We have presented a quantum optimization method based on linear-response theory, and further based on projecting the response onto control parameters. Maximizing these projection coefficients results in a more efficient update rule that drastically reduces the number of dynamical realizations of the time evolution necessary to update a set of parameters. Due to the nature of our approach, we refer to it as pulse engineering via the projection of response functions (PEPR). We understand this method in the context of the current resurgence of quantum optimal control theory under the moniker of quantum machine learning. PEPR is adjacent to GRAPE and VQAs, but extends them efficiently to time-nonlocal parametrizations that directly benefit from the projection feature of the method. In the endeavor to maximize the overlap of perturbation and parametrization bases, we consider a parametrization that utilizes a low-frequency Fourier-mode expansion.

In a minimal proof-of-concept example we demonstrate the benefits of PEPR over standard GRAPE. In a direct comparison of optimizing the CNOT gate, we find that PEPR produces high-fidelity solutions more reliably. Importantly, PEPR is more efficient in its optimization as it consistently requires less evaluations of time-evolution trajectories in order to optimize its protocols. The evaluation of such time evolutions in order to navigate the parameter space is an important bottleneck of gradient-based methods. Furthermore, we demonstrate that PEPR is efficient in optimizing systems with dissipation, as well as for optimization under constraints. Therefore, this improvement of quantum optimal control methods will support the design and establishment of quantum technology going forward.

ACKNOWLEDGMENTS

This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)-SFB-925-Project No. 170620586 and the Cluster of Excellence "Advanced Imaging of Matter" (EXC 2056) Project No. 390715994. The project is cofinanced by ERDF of the European Union and by "Fonds of the Hamburg Ministry of Science, Research, Equalities and Districts (BWFGB)."

[1] A. P. Peirce, M. A. Dahleh, and H. Rabitz, Optimal control of quantum-mechanical systems: Existence, numerical approximation, and applications, Phys. Rev. A **37**, 4950 (1988).

[2] C. Brif, R. Chakrabarti, and H. Rabitz, Control of quantum phenomena: Past, present and future, New J. Phys. **12**, 075008 (2010).

[3] C. Koch, U. Boscain, T. Calarco, G. Dirr, S. Filipp, S. Glaser, R. Kosloff, S. Montangero, T. Schulte-Herbrueggen, D. Sugny, and F. Wilhelm, Quantum optimal control in quantum technologies. Strategic report on current status, visions and goals for research in Europe, EPJ Quantum Technol. **9**, 19 (2022).

[4] J. P. Palao and R. Kosloff, Optimal control theory for unitary transformations, Phys. Rev. A **68**, 062308 (2003).

[5] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, Optimal control of coupled spin dynamics: Design of NMR pulse sequences by gradient ascent algorithms, J. Magn. Reson. **172**, 296 (2005).

[6] P. Rebentrost and F. K. Wilhelm, Optimal control of a leaking qubit, Phys. Rev. B **79**, 060507(R) (2009).

[7] D. J. Egger and F. K. Wilhelm, Optimized controlled-z gates for two superconducting qubits coupled through a resonator, Supercond. Sci. Technol. **27**, 014001 (2014).

[8] S. Jandura and G. Pupillo, Time-optimal two- and three-qubit gates for Rydberg atoms, Quantum **6**, 712 (2022).

[9] N. Heimann, L. Broers, N. Pintul, T. Petersen, K. Sponselee, A. Ilin, C. Becker, and L. Mathey, Quantum gate optimization for Rydberg architectures in the weak-coupling limit, arXiv:2306.08691

[10] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, New J. Phys. **18**, 023023 (2016).

[11] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625 (2021).

[12] J. Li, X. Yang, X. Peng, and C.-P. Sun, Hybrid quantum-classical approach to quantum optimal control, Phys. Rev. Lett. **118**, 150503 (2017).

[13] A. Choquette, A. Di Paolo, P. K. Barkoutsos, D. Sénéchal, I. Tavernelli, and A. Blais, Quantum-optimal-control-inspired ansatz for variational quantum algorithms, Phys. Rev. Res. **3**, 023092 (2021).

[14] A. B. Magann, C. Arenz, M. D. Grace, T.-S. Ho, R. L. Kosut, J. R. McClean, H. A. Rabitz, and M. Sarovar, From pulses to circuits and back again: A quantum optimal control perspective on variational quantum algorithms, PRX Quantum **2**, 010101 (2021).

[15] O. R. Meitei, B. T. Gard, G. S. Barron, D. P. Pappas, S. E. Economou, E. Barnes, and N. J. Mayhall, Gate-free state preparation for fast variational quantum eigensolver simulations, npj Quantum Inf. **7**, 155 (2021).

[16] R. de Keijzer, O. Tse, and S. Kokkelmans, Pulse based variational quantum optimal control for hybrid quantum computing, Quantum **7**, 908 (2023).

[17] L. Bittel, J. Watty, and M. Kliesch, Fast gradient estimation for variational quantum algorithms, arXiv:2210.06484.

[18] D. Wierichs, J. Izaac, C. Wang, and C. Y.-Y. Lin, General parameter-shift rules for quantum gradients, Quantum **6**, 677 (2022).

[19] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[20] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, Noisy intermediate-scale quantum algorithms, Rev. Mod. Phys. **94**, 015004 (2022).

[21] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nat. Commun. **9**, 4812 (2018).

[22] L. Bittel and M. Kliesch, Training variational quantum algorithms is NP-hard, Phys. Rev. Lett. **127**, 120502 (2021).

[23] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, Connecting ansatz expressibility to gradient magnitudes and barren plateaus, PRX Quantum **3**, 010313 (2022).

[24] E. R. Anschuetz and B. T. Kiani, Quantum variational algorithms are swamped with traps, Nat. Commun. **13**, 7760 (2022).

[25] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, An initialization strategy for addressing barren plateaus in parametrized quantum circuits, Quantum **3**, 214 (2019).

[26] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, Nat. Commun. **12**, 1791 (2021).

[27] L. Broers and L. Mathey, Mitigated barren plateaus in the time-nonlocal optimization of analog quantum-algorithm protocols, Phys. Rev. Res. **6**, 013076 (2024).

[28] A. Castro and I. V. Tokatly, Quantum optimal control theory in the linear response formalism, Phys. Rev. A **84**, 033410 (2011).

[29] P. Doria, T. Calarco, and S. Montangero, Optimal control technique for many-body quantum dynamics, Phys. Rev. Lett. **106**, 190501 (2011).

[30] T. Caneva, T. Calarco, and S. Montangero, Chopped random-basis quantum optimization, Phys. Rev. A **84**, 022326 (2011).

[31] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.7.013101 for the specific implementation details of PEPR and GRAPE that we use in this work, a discussion of the variance $\mathrm{Var}(1-F)$ for PEPR and GRAPE, and an implenentation of PEPR for the case of the Hadamard gate.

Pulse Engineering via Projection of Response Functions: Supplemental Material

Nicolas Heimann,[1,2,3,*] Lukas Broers,[1,2] and Ludwig Mathey[1,2,3]

[1] *Zentrum für Optische Quantentechnologien, Universität Hamburg, 22761 Hamburg, Germany*
[2] *Institut für Quantenphysik, Universität Hamburg, 22761 Hamburg, Germany*
[3] *The Hamburg Centre for Ultrafast Imaging, 22761 Hamburg, Germany*

I. GRADIENT ASCENT PULSE ENGINEERING

The objective of GRAPE [1] is to minimize a certain loss function defined for a quantum system by inferring an optimal set of parameters θ_{opt} that define the dynamics of that system. Here, we consider the Hamiltonian

$$H_\theta = H_0 + \sum_{j=1}^{n_B} \theta_j(t) B_j, \tag{1}$$

as described in Eq. 1 in the main-text, which produces the formal time-evolution operator

$$U_\theta(0, t_f) = e^{-\frac{i}{\hbar}\int_0^{t_f} H_\theta(t)dt}. \tag{2}$$

There are various ways to construct a loss function to evaluate the dynamics with. For instance, given a target transformation V, we can define the state-infidelity

$$1 - F = \text{Tr}(V\rho_0 V^\dagger \rho(t_f)), \tag{3}$$

where ρ_0 is some initial state and $\rho(t_f) = U(0, t_f)\rho_0 U^\dagger(0, t_f)$ is the state obtained from time-propagating the initial state ρ_0 from the initial time $t = 0$ to t_f. The central idea behind GRAPE is to update the parameters θ by gradient ascent (or descent) with respect to the loss function $\mathcal{L}$. It is possible to approximate the gradient of the loss function $\nabla\mathcal{L}$ by the finite differences

$$\frac{\partial \mathcal{L}}{\partial \theta_{j,k}} = \frac{\mathcal{L}(\theta + e_{kj}\epsilon) - \mathcal{L}(\theta)}{\epsilon}, \tag{4}$$

where $e_{j,k}$ is the unit-vector corresponding to the parameter $\theta_{j,k}$. This allows us to update the parameters as

$$\theta \to \theta - \alpha\nabla\mathcal{L}. \tag{5}$$

α is a hyperparameter commonly reffered to as the learning rate in gradient descent contexts. We note that rather than utilizing the full gradient, the parameters $\theta_{j,k}$ can also be updated individually. This requires two evaluation of the loss function per parameter, i.e. $\mathcal{L}(\theta_{j,k} + \epsilon)$ and $\mathcal{L}(\theta_{j,k})$. Therefore, it is common to update all the parameters within a single iteration of the update rule. This reduces the computational costs from $2|\theta|$ to $|\theta| + 1$ time-evolutions, where $|\theta|$ is the number of parameters.

II. CNOT GATE

In this section we describe the model we use in the main-text to demonstrate the performance of PEPR to obtain high-fidelity realizations of the CNOT gate. The qubits are locally controlled as

$$H_1(t) = h_{x,1}(t)\sigma_x^1 + h_{y,1}(t)\sigma_y^1, \tag{6}$$
$$H_2(t) = h_{x,2}(t)\sigma_x^2 + h_{y,2}(t)\sigma_y^2, \tag{7}$$

* nheimann@physnet.uni-hamburg.de

where $\sigma_j^1 = \sigma_j \otimes 1$, and $\sigma_j^2 = 1 \otimes \sigma_j$ are the local Pauli matrices. The interaction-term between the qubits is

$$H_{\text{int}} = J(t)\vec{\sigma}_1\vec{\sigma}_2, \tag{8}$$

where $\vec{\sigma}_1\vec{\sigma}_2 = \sigma_x \otimes \sigma_x + \sigma_y \otimes \sigma_y + \sigma_z \otimes \sigma_z$, and $J(t)$ is the time-dependent strength of the interaction. We write the resulting Hamiltonian as

$$H_\theta = H_1 + H_2 + H_{\text{int}} = \begin{pmatrix} J & h_{x,2} - ih_{y,2} & h_{x,1} - ih_{y,1} & 0 \\ h_{x,2} + ih_{y,2} & -J & 2J & h_{x,1} - ih_{y,1} \\ h_{x,1} + ih_{y,1} & 2J & -J & h_{x,2} - ih_{y,2} \\ 0 & h_{x,1} + ih_{y,1} & h_{x,2} + ih_{y,2} & J \end{pmatrix}. \tag{9}$$

As outlined in the main-text, we consider the parameterized protocols

$$h_{p,i}(t) = \sum_k \theta_{p,i,k} \sin\left(\pi k \frac{t}{t_f}\right) \tag{10}$$

$$J(t) = \sum_k \theta_{J,k} \sin\left(\pi k \frac{t}{t_f}\right), \tag{11}$$

where $p \in \{x, y\}$, $i \in \{1, 2\}$. We explicitly write the density operator of the system as

$$\rho = \begin{pmatrix} \rho_1 & \rho_4 - i\rho_5 & \rho_6 - i\rho_7 & \rho_{10} - i\rho_{11} \\ \rho_4 + i\rho_5 & \rho_2 & \rho_8 - i\rho_9 & \rho_{12} - i\rho_{13} \\ \rho_6 + i\rho_7 & \rho_8 + i\rho_9 & \rho_3 & \rho_{14} - i\rho_{15} \\ \rho_{10} + i\rho_{11} & \rho_{12} + i\rho_{13} & \rho_{14} + i\rho_{15} & \xi - \rho_1 - \rho_2 - \rho_3 \end{pmatrix}, \tag{12}$$

where $\rho_i \in \mathbb{R}$, and $\xi \in \{0, 1\}$ defines the value of trace of the state. We note that $\xi = 0$ is required to time-propagate $i[B_j, \rho]$, as $\text{Tr}([B_j, \rho]) = 0$. For numerical purpose, we choose to represent ρ as a real-valued vector

$$\vec{\rho} = (\rho_1, ..., \rho_{15})^T. \tag{13}$$

We include dissipation in the form of pure-dephasing of the individual qubits. The time-evolution of the state is governed by the Lindblad master equation

$$\frac{\partial \rho}{\partial t} = -\frac{i}{\hbar}[H_\theta(t), \rho] + \gamma_z \sum_i \mathcal{D}[\sigma_z^i]\rho, \tag{14}$$

where $\mathcal{D}[L]\rho = L\rho L^\dagger - \frac{1}{2}\{L^\dagger L, \rho\}$. The equations of motion read

$$\partial_t\rho_1 = 2(-h_{y,2}\rho_4 + h_{x,2}\rho_5 - h_{y,1}\rho_6 + h_{x,1}\rho_7) \tag{15}$$
$$\partial_t\rho_2 = 2(-h_{y,1}\rho_{12} + h_{x,1}\rho_{13} + h_{y,2}\rho_4 - h_{x,2}\rho_5 + 2J\rho_9) \tag{16}$$
$$\partial_t\rho_3 = -2(h_{y,2}\rho_{14} - h_{x,2}\rho_{15} - h_{y,1}\rho_6 + h_{x,1}\rho_7 + 2J\rho_9) \tag{17}$$
$$\partial_t\rho_4 = h_{x,1}\rho_{11} + h_{y,2}(\rho_1 - \rho_2) - 2\gamma_z\rho_4 - 2J\rho_5 + 2J\rho_7 - h_{y,1}(\rho_{10} + \rho_8) + h_{x,1}\rho_9 \tag{18}$$
$$\partial_t\rho_5 = -h_{y,1}\rho_{11} + h_{x,2}(-\rho_1 + \rho_2) + 2J\rho_4 - 2\gamma_z\rho_5 - 2J\rho_6 + h_{x,1}(-\rho_{10} + \rho_8) + h_{y,1}\rho_9 \tag{19}$$
$$\partial_t\rho_6 = h_{x,2}\rho_{11} + h_{y,1}(\rho_1 - \rho_3) + 2J\rho_5 - 2\gamma_z\rho_6 - 2J\rho_7 - h_{y,2}(\rho_{10} + \rho_8) - h_{x,2}\rho_9 \tag{20}$$
$$\partial_t\rho_7 = h_{x,1}(-\rho_1 + \rho_3) - 2(J\rho_4 - J\rho_6 + \gamma_z\rho_7) + h_{x,2}(-\rho_{10} + \rho_8) - h_{y,2}(\rho_{11} + \rho_9) \tag{21}$$
$$\partial_t\rho_8 = -h_{y,1}\rho_{14} + h_{x,1}\rho_{15} + h_{y,1}\rho_4 - h_{x,1}\rho_5 + h_{y,2}(-\rho_{12} + \rho_6) + h_{x,2}(\rho_{13} - \rho_7) - 4\gamma_z\rho_8 \tag{22}$$
$$\partial_t\rho_9 = h_{x,1}\rho_{14} + h_{y,1}\rho_{15} - 2J\rho_2 + 2J\rho_3 - h_{x,1}\rho_4 - h_{y,1}\rho_5 + h_{x,2}(-\rho_{12} + \rho_6) + h_{y,2}(-\rho_{13} + \rho_7) - 4\gamma_z\rho_9 \tag{23}$$
$$\partial_t\rho_{10} = -4\gamma_z\rho_{10} - h_{x,2}\rho_{13} - h_{y,1}\rho_{14} - h_{x,1}\rho_{15} + h_{y,1}\rho_4 + h_{x,1}\rho_5 + h_{y,2}(-\rho_{12} + \rho_6) + h_{x,2}\rho_7 \tag{24}$$
$$\partial_t\rho_{11} = -4\gamma_z\rho_{11} - h_{y,2}\rho_{13} + h_{x,1}\rho_{14} - h_{y,1}\rho_{15} + h_{x,1}\rho_4 + h_{y,1}\rho_5 + h_{x,2}(\rho_{12} - \rho_6) + h_{y,2}\rho_7 \tag{25}$$
$$\partial_t\rho_{12} = -h_{x,2}\rho_{11} - 2\gamma_z\rho_{12} + 2J\rho_{13} - 2J\rho_{15} + h_{y,1}(-1 + \rho_1 + \rho_2 + \rho_3) + h_{y,2}(\rho_{10} + \rho_8) + h_{x,2}\rho_9 \tag{26}$$
$$\partial_t\rho_{13} = h_{y,2}\rho_{11} - 2J\rho_{12} - 2\gamma_z\rho_{13} + 2J\rho_{14} - h_{x,1}(-1 + \rho_1 + \rho_2 + \rho_3) + h_{x,2}(\rho_{10} - \rho_8) + h_{y,2}\rho_9 \tag{27}$$
$$\partial_t\rho_{14} = -2(J\rho_{13} + \gamma_z\rho_{14} - J\rho_{15}) + h_{y,2}(-1 + \rho_1 + \rho_2 + \rho_3) + h_{y,1}(\rho_{10} + \rho_8) - h_{x,1}(\rho_{11} + \rho_9) \tag{28}$$
$$\partial_t\rho_{15} = h_{y,1}\rho_{11} + 2J\rho_{12} - 2J\rho_{14} - 2\gamma_z\rho_{15} - h_{x,2}(-1 + \rho_1 + \rho_2 + \rho_3) + h_{x,1}(\rho_{10} - \rho_8) - h_{y,1}\rho_9. \tag{29}$$

We initialize the state of the system in a product state

$$\rho(0) = \rho_1 \otimes \rho_2, \tag{30}$$

of the individual initial qubit states $\rho_i = \frac{1}{2}(1 + \rho_x^i \sigma_x^i + \rho_y^i \sigma_y^i + \rho_z^i \sigma_z^i)$, where the vector components are sampled as $\rho_x^i, \rho_y^i, \rho_z^i \sim \mathcal{N}(0,1)$. We consider the CNOT target transformation

$$V = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \tag{31}$$

and the corresponding fidelity therefore reads

$$F_\theta(\vec{\rho}(0), \vec{\rho}(t_f)) = \mathrm{Tr}\left(\rho_f^\dagger V \rho(0) V^\dagger\right) \tag{32}$$

$$\begin{aligned} =&\rho_1(t_f)\rho_1(0) + 2\rho_6(t_f)\rho_{10}(0) + 2\rho_7(t_f)\rho_{11}(0) + 2\rho_8(t_f)\rho_{12}(0) + 2\rho_9(t_f)\rho_{13}(0) \\ &+ 2\rho_{14}(t_f)\rho_{14}(0) - 2\rho_{15}(t_f)\rho_{15}(0) + \rho_2(t_f)\rho_2(0) + \rho_3(0) \\ &- (\rho_1(t_f) + \rho_2(t_f))\rho_3(0) - \rho_3(t_f)(-\mathrm{Tr}(\vec{\rho}(t_f)) + \rho_1(0) + \rho_2(0) + 2\rho_3(0)) \\ &+ 2(\rho_4(t_f)\rho_4(0) + \rho_5(t_f)\rho_5(0) + \rho_{10}(t_f)\rho_6(0) + \rho_{11}(t_f)\rho_7(0) + \rho_{12}(t_f)\rho_8(0) + \rho_{13}(t_f)\rho_9(0)). \end{aligned} \tag{33}$$

Instead of evaluating this expression directly, we draw a random time $t_r \in [0, t_f]$ and propagate the initial state to that time which gives us $\rho(t_r)$. We then consider the time-local perturbation proportional to one of the accessible control operators of the Hamiltonian H_θ, i.e. $B_j \in \{\sigma_x^1, \sigma_x^2, \sigma_y^1, \sigma_y^2, \vec{\sigma_1}\vec{\sigma_2}\}$. In the vector representation $\vec{\rho}$, the corresponding perturbations evaluate as

$$\begin{aligned} i[\sigma_x^1, \rho] \to&(-2\rho_7, -2\rho_{13}, 2\rho_7, -\rho_{11} - \rho_9, \rho_{10} - \rho_8, 0, \rho_1 - \rho_3, -\rho_{15} + \rho_5, \\ &- \rho_{14} + \rho_4, \rho_{15} - \rho_5, -\rho_{14} + \rho_4, 0, \rho_1 + 2\rho_2 + \rho_3 - 1, \rho_{11} + \rho_9, -\rho_{10} + \rho_8)^T, \end{aligned} \tag{34}$$

$$\begin{aligned} i[\sigma_x^2, \rho] \to&(-2\rho_5, 2\rho_5, -2\rho_{15}, 0, \rho_1 - \rho_2, -\rho_{11} + \rho_9, \rho_{10} - \rho_8, -\rho_{13} + \rho_7, \rho_{12} \\ &- \rho_6, \rho_{13} - \rho_7, -\rho_{12} + \rho_6, \rho_{11} - \rho_9, -\rho_{10} + \rho_8, 0, \rho_1 + \rho_2 + 2\rho_3 - 1)^T, \end{aligned} \tag{35}$$

$$\begin{aligned} i[\sigma_y^1, \rho] \to&(2\rho_6, 2\rho_{12}, -2\rho_6, \rho_{10} + \rho_8, \rho_{11} - \rho_9, -\rho_1 + \rho_3, 0, \rho_{14} - \rho_4, -\rho_{15} + \\ &\rho_5, \rho_{14} - \rho_4, \rho_{15} - \rho_5, -\rho_1 - 2\rho_2 - \rho_3 + 1, 0, -\rho_{10} - \rho_8, -\rho_{11} + \rho_9)^T, \end{aligned} \tag{36}$$

$$\begin{aligned} i[\sigma_y^2, \rho] \to&(2\rho_4, -2\rho_4, 2\rho_{14}, -\rho_1 + \rho_2, 0, \rho_{10} + \rho_8, \rho_{11} + \rho_9, \rho_{12} - \rho_6, \rho_{13} - \\ &\rho_7, \rho_{12} - \rho_6, \rho_{13} - \rho_7, -\rho_{10} - \rho_8, -\rho_{11} - \rho_9, -\rho_1 - \rho_2 - 2\rho_3 + 1, 0)^T, \end{aligned} \tag{37}$$

$$\begin{aligned} i[\vec{\sigma_1}\vec{\sigma_2}, \rho] \to&(0, -4\rho_9, 4\rho_9, 2\rho_5 - 2\rho_7, -2\rho_4 + 2\rho_6, -2\rho_5 + 2\rho_7, 2\rho_4 - 2\rho_6, 0, 2\rho_2 - \rho_3 \\ &2, 0, 0, -2\rho_{13} + 2\rho_{15}, 2\rho_{12} - 2\rho_{14}, 2\rho_{13} - 2\rho_{15}, -2\rho_{12} + 2\rho_{14})^T. \end{aligned} \tag{38}$$

After propagating these resulting operators from t_r to t_f, we evaluate the susceptibility of the fidelity in Eq. 33 with respect to the perturbation of B_j at time t_r, which is

$$\chi_j(t_r) = \frac{i}{\hbar}\,\mathrm{Tr}\left(V\rho(0)V^\dagger U_\theta(t_r, t_f)\Big[B_j, U_\theta(t_r)\rho(0)U_\theta^\dagger(t_r)\Big]U_\theta^\dagger(t_r, t_f)\right). \tag{39}$$

We finally update the parameters according to PEPR as described in Eq.15 in the main-text. It is

$$\theta_{j,k} \to \theta_{j,k} - \alpha \sin\left(\pi k \frac{t_r}{t_f}\right)\chi_j(t_r). \tag{40}$$

Note, that this update requires only a single time-evolution to obtain χ_j, which is used to update n_j-many parameters $\theta_{j,k}$. This provides better scaling with respect to the number of parameters, compared to GRAPE.

The expressibility of the control functions is determined by the number of modes n_j. In Fig. 1 we show the average infidelity $\langle \log_{10}(1-F)\rangle$, as a function of the number of runs N_{run} based on PEPR and GRAPE for different values of n_j. For a value of $n_j = 2$, the average infidelity obtains high values, indicating insufficient expressibility of the control functions. For larger values of $n_j > 2$, the average infidelity converges to a lower bound determined by numerical accuracy. We find that a value of $n_j = 8$ leads to an efficient convergence behavior of the average infidelity, for both PEPR and GRAPE.

We determine an optimal set of hyperparameters of $\alpha_{\mathrm{PEPR}} = 0.5$ for PEPR, and $\alpha_{\mathrm{GRAPE}} = 1.2$, and $\epsilon = 10^{-7}$ for GRAPE by comparing the convergence behavior of the average infidelity, as described in Eq. 20 in the main-text, for different values of the hyperparameters. In Fig. 2 we show the results that motivate this particular choice of hyperparameters.

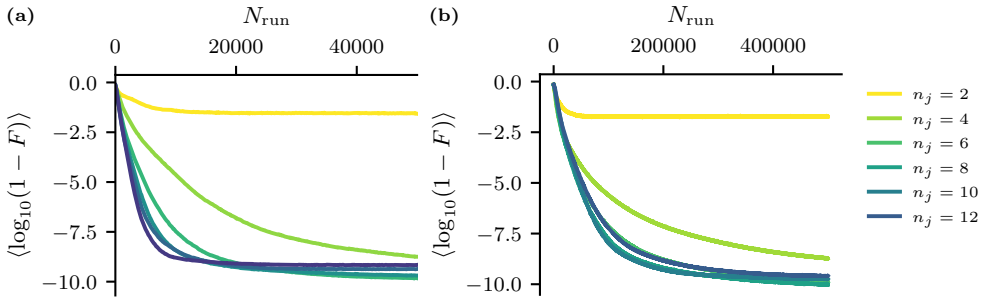


Figure 1. **Number of Modes.** The average infidelity $\langle \log_{10}(1-F)\rangle$, as a function of the number of runs $N_{\rm run}$, for different values of the number of modes n_j per control function. The lower bound of the infidelity is determined by the time-discretization h of the 4th-order Runge-Kutta method. Panel **(a)** shows the results for PEPR with a learning rate of $\alpha_{\rm PEPR} = 0.3$. Panel **(b)** shows the results for GRAPE with a learning rate of $\alpha_{\rm GRAPE} = 0.2$ and $\epsilon = 10^{-6}$.

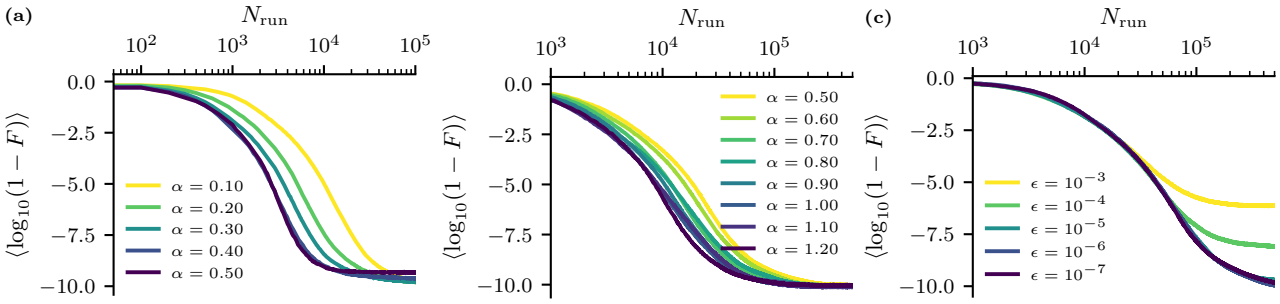


Figure 2. **Hyperparameters**. The average infidelity $\langle \log_{10}(1-F)\rangle$, as a function of the number of runs $N_{\rm run}$ for different values of the hyperparameters with $n_j = 8$. The lower bound of the infidelity is determined by the time-discretization h of the 4th-order Runge-Kutta method. Panel **(a)** shows the average infidelity based on PEPR for different values of the learning rate α. Panel **(b)** shows the average infidelity based on GRAPE for different values of the learning rate α. Panel **(c)** shows the average infidelity based on GRAPE for different values of the finite difference length ϵ.

III. VARIANCE OF THE INFIDELITY

In this section we show the variance of the ensemble of optimization trajectories obtained using PEPR and GRAPE. The unbiased sample variance over the ensemble of realizations is

$$\mathrm{Var}(1-F) = \frac{n_T}{n_T - 1}\sum_{\theta}(F_\theta - \langle F_\theta\rangle)^2 \tag{41}$$

In Fig. 3, we show the variance of the ensembles of realizations for the underlying data used in Fig. 2 in the main-text. PEPR leads to a reduced variance, compared to GRAPE.

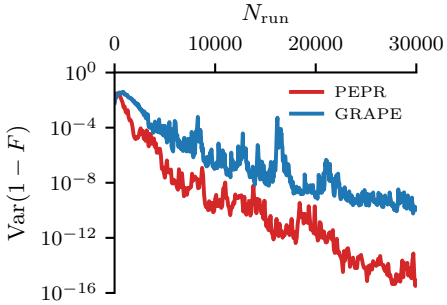


Figure 3. **Variance of the infidelity.** The variance of the infidelity $\mathrm{Var}(1-F)$ obtained using PEPR (red line) and GRAPE (blue line), as a function of the number of runs $N_{\rm run}$. PEPR leads to a reduced variance, compared to GRAPE.

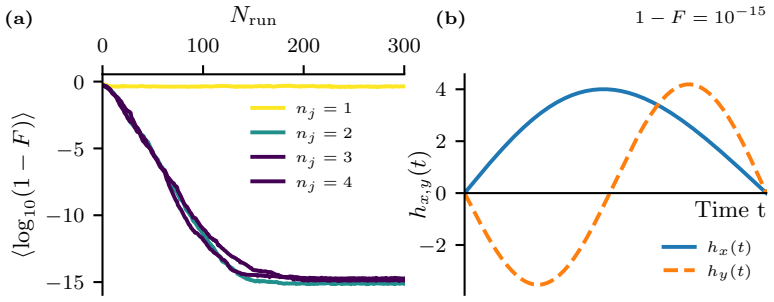


Figure 4. **Results of PEPR for the Hadamard gate**. Panel **(a)** shows the average infidelity $\langle \log_{10}(1-F) \rangle$, as a function of the number of runs N_{run}, for different values of $n_j = 1, 2, 3, 4$ and for a fixed learning rate of value $\alpha = 2.5$. The lower bound of the infidelity is determined by the time-discretization h of the 4th-order Runge-Kutta method. Panel **(b)** shows the control functions $h_x(t)$ and $h_y(t)$ of an example for a high-fidelity implementation for $n_j = 2$.

IV. HADAMARD GATE

Additionally to the optimization of the CNOT gate we demonstrate PEPR here for the example of the Hadamard gate on a single qubit. For this, we consider the Hamiltonian

$$H_\theta(t) = h_x(t)\sigma_x + h_y(t)\sigma_y, \tag{42}$$

with $h_x(t) = \sum_k \theta_{x,k} \sin(\pi k t)$ and $h_y(t) = \sum_k \theta_{y,k} \sin(\pi k t)$. We write the density operator of the system as

$$\rho = \frac{1}{2}\begin{pmatrix} \xi + \rho_z & \rho_x - i\rho_y \\ \rho_x + i\rho_y & \xi - \rho_z \end{pmatrix}, \tag{43}$$

where $\rho_i \in \mathbb{R}$, and $\xi \in \{0, 1\}$ defines the trace of the operator. We note that $\xi = 0$ is required to capture commutator-objects such as $i[B_j, \rho]$ after the perturbation with the control operator B_j, as $\text{Tr}([B_j, \rho]) = 0$. For numerical purposes, we represent ρ as the real-valued vector

$$\vec{\rho} = (\rho_x, \rho_y, \rho_z)^T \tag{44}$$

The dynamics of the state obey the von-Neumann equation $i\hbar\dot{\rho} = [H, \rho]$ and the equations of motion read

$$\dot{\rho}_x = 2h_y\rho_z, \tag{45}$$
$$\dot{\rho}_y = -2h_x\rho_z, \tag{46}$$
$$\dot{\rho}_z = 2h_x\rho_y - 2h_y\rho_x. \tag{47}$$

We initialize the state of the system as $\vec{\rho}(0) \sim \mathcal{N}^3(0, 1)$. We denote the time-propagated state over $[0, t_f]$ as $\rho(t_f)$. We consider the example of the Hadamard target transformation

$$V = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{48}$$

The fidelity to reach the target state $V\rho(0)V^\dagger$, given the initial state $\rho(0)$, is

$$F_\theta = \text{Tr}(V\rho(t_0)V^\dagger\rho(t_f)) = \frac{1}{2}(\xi + \rho_z(t_f)\rho_x(t_0) + \rho_x(t_f)\rho_z(t_0) - \rho_y(t_f)\rho_y(t_0)). \tag{49}$$

The control operators are $B_j \in \{\sigma_x, \sigma_y\}$. In the vector representation $\vec{\rho}$, the corresponding perturbations evaluate as

$$i[\sigma_x, \rho] \to (0, \rho_z, -\rho_y)^T, \tag{50}$$
$$i[\sigma_y, \rho] \to (-\rho_z, 0, \rho_x)^T. \tag{51}$$

We update the parameters according to PEPR as described in Eq. 15 in the main-text. It is

$$\theta_{j,k} \to \theta_{j,k} - \alpha \sin\left(\pi k \frac{t_r}{t_f}\right)\chi_j(t_r) \tag{52}$$

with the susceptibility under a randomly chosen perturbation B_j at the random time $t_r \in [0,1]$

$$\chi_j(t_r) = \frac{i}{\hbar}\,\mathrm{Tr}\left(V\rho(0)V^\dagger U_\theta(t_r,t_f)\Big[B_j, U_\theta(t_r)\rho(0)U_\theta^\dagger(t_r)\Big]U_\theta^\dagger(t_r,t_f)\right). \tag{53}$$

U_θ denotes the unitary time-evolution operator generated by the Hamiltonian Eq. 42.

In Fig. 4 (a) we show the average infidelity $\langle \log_{10}(1-F)\rangle$, as a function of the number of runs N_{run} for an empirically determined optimal learning rate of $\alpha_{\mathrm{PEPR}} = 2.5$. The minimal number of modes to obtain high-fidelity protocols in this simple example is $n_j > 1$. The average infidelity converges to values of $\langle \log_{10}(1-F)\rangle(N_{\mathrm{run}}) \to 10^{-15}$, for $N_{\mathrm{run}} > 200$. The lower bound of the infidelity is determined by the time-discretization h of the 4th-order Runge-Kutta method. We show an example for a high-fidelity implementation of the Hadamard gate with $n_j = 2$ in Fig. 4 (b).

[1] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, Optimal control of coupled spin dynamics: design of NMR pulse sequences by gradient ascent algorithms, Journal of Magnetic Resonance **172**, 296 (2005).

5 Quantum optimal control strategies for Rydberg-atom systems

In recent years, quantum computing based on neutral atoms utilizing Rydberg-mediated interactions has reached remarkable milestones as both a scientific platform and an emerging industry-ready quantum technology. The strong dipole-dipole interaction between nearby Rydberg atoms enables a wide range of quantum information processing [24–27] and quantum simulation [28–30] opportunities. Optical tweezer arrays trapping neutral atoms can be assembled nearly defect-free and provide versatile spatial control [165–169]. Laser-driven atomic transitions enable single qubit control [170, 171] and multi-qubit entanglement gates [40, 172, 173] with high fidelity. Nowadays, hundreds of atoms can be trapped in arbitrary geometries, demonstrating scalability towards quantum supremacy. Large scale digital quantum processors, utilizing hundreds of qubits, have been implemented [127]. For an overview on quantum computing with neutral atoms we refer the reader to [174].

The exploration for fault-tolerant operational regimes paving the way toward quantum supremacy is an active area of research. Error correcting codes [175, 176] introduce redundancies to protect the logical domain against operational losses. Mid-circuit erasure conversions [177] allow for the detection of losses via fluorescence in disjoint subspaces without measuring the logical domain. The utilization of variational quantum algorithms [19, 129, 130] or quantum optimal control theory [20, 22, 132, 178] further give rise to robust [142–145, 179] and time-optimal [141] entanglement gates.

In this chapter, I discuss the design of optimized protocols for the implementation of entanglement gates based on neutral atoms and Rydberg-mediated interactions. I elaborate on Rydberg mediated interactions originating from dipole-dipole interaction of highly excited Rydberg states. Next, I discuss specific implementations of fast quantum gates with neutral atoms. I provide details about the computational platform based on the fermionic Ytterbium isotope ^{171}Yb that has inspired our model in publication [2]. In an additional section, I discuss an improved noise model that we have implemented in response to referee reports received during the submission process of the manuscript.

5.1 Dipole-dipole interaction between Rydberg atoms

In a Rydberg atom [31], one or more valance electrons are in highly excited electronic states, with large principle quantum numbers n. The size of a Rydberg atom scales as n^2, as their geometric cross section scales as n^4. Remarkably, the size of the atom can reach the micrometer regime for sufficiently large n, resulting in large electronic dipole

moments $\sim n^2$ of the ionic core and the highly excited valence electrons. For more details, we refer the reader to [24, 32], on which this section is based on.

The dipole-dipole potential between two Rydberg atoms is

$$V = \frac{e^2}{4\pi\epsilon_0}\frac{\mathbf{d}_1\mathbf{d}_2 - 3(\mathbf{d}_1\mathbf{e}_r)(\mathbf{d}_2\mathbf{e}_r)}{r^3}, \tag{5.1}$$

where $\mathbf{d}_1$ and $\mathbf{d}_2$ are the displacements of the valance electrons to their corresponding ionic cores, and $\mathbf{e}_r$ is the unit vector along the relative coordinate $\mathbf{r}$ characterizing the the separation between the dipoles. This interaction scales as n^4, as $|\mathbf{d}_i|$ scales with n^2. In the following, it is assumed that the Rydberg atoms are separated such that their electronic wave functions do not overlap significantly. Dipole transitions occur between states of opposite parity and hence the dipole-dipole potential in Eq. 5.1 has non-zero matrix elements for dipole-allowed transitions $|R_1R_2\rangle \rightarrow |R_1'R_2'\rangle$, where $|R_i\rangle$ and $|R_i'\rangle$ are Rydberg states. In the subspace spanned by $\{|R_1R_2\rangle, |R_1'R_2'\rangle\}$, the Hamiltonian has the matrix form

$$\begin{pmatrix} 0 & \frac{\hbar C_3(\theta)}{r^3} \\ \frac{\hbar C_3(\theta)}{r^3} & \delta_F \end{pmatrix}, \tag{5.2}$$

where $\delta_F = (E_1' + E_2') - (E_1 + E_2)$ is the Föster defect, E_i and E_i' are the energies of the Rydberg states $|R_i\rangle$ and $|R_i'\rangle$ respectively. $C_3(\theta) \sim n^4$ is an anisotropic interaction coefficient. Transitions into other Rydberg-Rydberg states having larger Föster defects $\delta_F' \gg \delta_F$ are neglected. The corresponding energy eigenvalues reads as follows

$$E_\pm = \frac{\delta_F}{2} \pm \frac{1}{2}\sqrt{\delta_F^2 + 4\frac{\hbar^2 C_3^2(\theta)}{r^6}}. \tag{5.3}$$

In the limit of $\delta_F \gg \hbar C_3(\theta)/r^3$, the bare state energies shift in the lowest perturbative order by V_{vdW} for $|R_1R_2\rangle$ and $-V_{\mathrm{vdW}}$ for $|R_1'R_2'\rangle$, where

$$V_{\mathrm{vdW}} = \frac{\hbar C_6(\theta)}{r^6} \tag{5.4}$$

is the van-der-Waals type interaction strength and $C_6(\theta) = \hbar C_3^2(\theta)/\delta_F$ is the interaction coefficient characterizing the Rydberg-mediated interaction. In the other limit $\delta_F \ll C_3(\theta)/r^3$ which occurs close to Föster resonance [180], the energy shifts asymptotically reach $C_3(\theta)/r^3$ leading to a mixing of the bare states. The crossover between the these two regimes is defined via $V(r_{\mathrm{vdW}}) = \delta_F$.

Rydberg-mediated interactions are commonly utilized to generate correlated states via the phenomenon of the Rydberg blockade. Let us consider two electronic states, namely the groundstate $|g\rangle$ and a Rydberg state $|R\rangle$ coupled with a Rabi frequency of Ω. For two atoms initially in the ground-state, the dynamics can be described in the subspace spanned by $\{|RR\rangle, |+\rangle, |gg\rangle\}$, where $|+\rangle = (|gR\rangle + |Rg\rangle)/\sqrt{2}$. The Hamiltonian reads

$$V_{\mathrm{vdW}}\,|RR\rangle\langle RR| + \frac{\hbar\sqrt{2}\Omega}{2}\left(|+\rangle\langle 11| + |RR\rangle\langle +| + \mathrm{h.c.}\right). \tag{5.5}$$

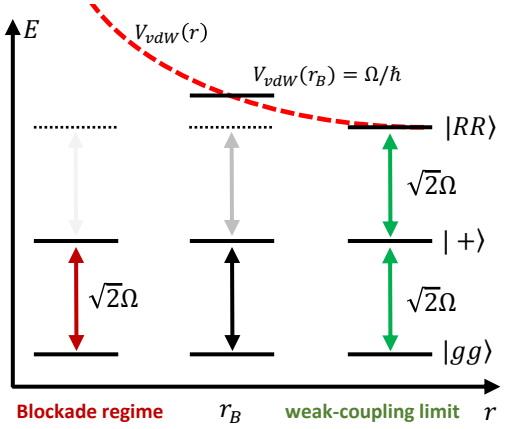


Figure 5.1: **Rydberg-mediated interaction**. The dipole-dipole interaction between Rydberg states induces an energy shift $V_{\mathrm{vdW}} = \hbar C_6/r^6$ of the Rydberg-Rydberg state $|RR\rangle$. In the blockade regime, transitions into $|RR\rangle$ are blocked, allowing for full population of the state $|+\rangle$. In the weak-coupling limit, population transfer into $|RR\rangle$ results in a dynamical phase accumulation at a rate of $V_{\mathrm{vdW}}/\hbar$.

For a sufficiently large van-der-Waals energy $V_{\mathrm{vdW}}/\Omega \gg 1$, the Rydberg-Rydberg state $|RR\rangle$ is shifted far out of resonance. In this limit, the amplitude of the Rydberg-Rydberg state $|RR\rangle$ is constant to lowest order in Ω/V_{vdW}, with $c_{RR} = -\sqrt{2}\Omega/(2V_{\mathrm{vdW}})c_+$, where $c_a = \langle a|\psi\rangle$. In this case, the dynamics can be described in the subspace spanned by $\{|+\rangle, |gg\rangle\}$. To lowest order in Ω, the effective Hamiltonian reads

$$\frac{\hbar\sqrt{2}\Omega}{2}|+\rangle\langle gg| + \text{h.c.}\ , \tag{5.6}$$

resulting in an oscillation of population between the ground state $|gg\rangle$ and the correlated state $|+\rangle$, while transitions into the Rydberg-Rydberg state $|RR\rangle$ are suppressed. This mechanism of "blocking" transitions into the Rydberg-Rydberg state $|RR\rangle$ is commonly referred to as Rydberg blockade, and it has been demonstrated with neutral atoms [33, 34]. In publication [2], we design optimized protocols for the implementation of entanglement gates in three dynamically distinct regimes, see Fig 5.1. We refer to the regime of $V_{\mathrm{vdW}}/\Omega \gg 1$ as the Blockade regime. In the opposite limit of $V_{\mathrm{vdW}}/\Omega \ll 1$, transitions into the Rydberg-Rydberg $|RR\rangle$ state are only partially blocked allowing for dynamical phase accumulation of the Rydberg-Rybderg state $|RR\rangle$ at a rate of $V_{\mathrm{vdW}}/\hbar$. We refer to this regime as the weak-coupling limit. The intermediate regime is characterized by the blockade radius r_B, at which the two energy scales equal, i.e. $V_{\mathrm{vdW}}(r_B)/\hbar = \Omega$.

The Rydberg blockade mechanism can be generalized to the case of many Rydberg atoms such that only a single Rydberg state can be excited in the nearby vicinity characterized by r_B. Hard combinatorial problems can be tackled with such ensembles of neutral atoms. A prominent example is the maximum independent set problem on unit

disks graphs [161, 181, 182].

5.2 Entanglement gates

Entanglement gates with neutral atoms via Rydberg-mediated interactions can be implemented by utilizing local driving fields for conditionally transferring population between qubits and auxiliary Rydberg states [35]. In both, the Blockade regime and the weak-coupling limit, sequences of $n\pi$ pulses can be constructed to implement the CZ_π entanglement gate with high fidelity. In the Blockade regime, a π-pulse is applied to the control qubit, conditionally transferring population into the Rydberg state $|R\rangle_c$. Subsequently, a 2π-pulse is applied to the target qubit, resulting in a conditional π phase shift, depending on the population of $|R\rangle_c$ blocking transitions into the Rydberg state $|R\rangle_t$. Finally, a π-pulse is applied to the control atom to rotate it back into the logical subspace. The resulting transformation approximates the CZ_π gate, up to global phase. This protocol has been demonstrated experimentally with high fidelity [36–39]. A similar protocol exists in the weak-coupling limit, where both atoms are conditionally driven into the Rydberg state simultaneously by a global π pulse. After the duration required for the Rydberg-Rydberg state $|RR\rangle$ to dynamically accumulate a π phase, the first pulse is applied again.

Another protocol implements the CZ_π gate based on two off-resonant global pulses in the Blockade regime [40]. This global protocol implements the CZ_π gate in $\approx 68\%$ operational time compared to the local protocol discussed above. Utilizing quantum optimal control, the time-optimal operational regime can be reached by optimizing dynamical phase protocols of the global driving field [141]. In publication [2] we design optimized protocols that implement CNOT gate by additionally utilizing local qubit control via two photon Raman driving.

5.3 Computational platform

Alkaline earth(-like) metals have two outer valence electrons rendering many useful electronic properties for the purpose of quantum computing with neutral atoms based on Rydberg-mediated interactions. The nuclear spins of the $^1\mathrm{S}_0$ ground state (g) and the $^3\mathrm{P}_0$ metastable state (m) on an optical clock transition [183] serve as qubit encodings that are insensitive to magnetic field fluctuations and having long coherence times [184, 185]. Additionally, the optical clock qubit (o) can be encoded on the optical clock transition between (m) and (g), motivating the abbreviation "omg" in the context of such platforms [41]. Different atomic species have been trapped in optical tweezer arrays, such as Caesium [186], Rubidium [165, 166, 173], Strontium [187–189] or Ytterbium [42, 43, 170]. Specifically, the fermionic Ytterbium isotope ^{171}Yb, with its nuclear spin of $I = 1/2$ can naturally be utilized to encode the (g) and (m) qubits in the hyperfine states $|^1S_0, m_F = \pm 1/2\rangle$ and $|^3P_0, m_F = \pm 1/2\rangle$, respectively. These hyperfine qubits are locally controlled by Raman driving via intermediate states. In publication [2], we implement a numerical two-qubit platform that is inspired by the omg architecture for

^{171}Yb [41]. In such platforms magic trapping allows to trap the Rydberg state with the same trap as the ground state [190, 191].

5.4 Publication II: Quantum Gate Optimization for Rydberg Architectures in the Weak-Coupling Limit

Nicolas Heimann, Lukas Broers, Nejira Pintul, Tobias Petersen, Koen Sponselee, Alexander Ilin, Christoph Becker, and Ludwig Mathey - **arXiv:2306.08691 (2022) [2] - Initially submitted to Phys. Rev. Research., currently being revised for resubmission.**

Motivation

Optimized strategies, robust against experimental noise and decoherence, pave the way toward favorable operational regimes necessary for fault-tolerant quantum computing. The goal of this project was to pulse engineer optimized protocols for implementing the CNOT gate in the presence of experimentally motivated constraints, noise and decoherence in different dynamical regimes, ranging from the Blockade regime to the weak-coupling limit.

Main findings

In this work we have implemented a hybrid quantum-classical optimization platform for the design of optimized protocols implementing the CNOT gate, for a wide range of interatomic distances and in the presence of dissipation. With a fixed gate duration of 1μs, we have determined the critical van-der-Waals interaction strength sufficient to implement the CNOT gate in our setup. We have studied the susceptibility of our optimized protocols to spatial fluctuations and have found robust implementations in the blockade regime and the weak-coupling limit. However, we have found that in the intermediate regime the protocols are highly susceptible to spatial fluctuations. In the revised version of the manuscript, see Sect. 5.5, we have implemented an improved noise model based on the Langevin dynamics of trapped atoms in the presence of thermal fluctuations. We have demonstrated that an in-situ optimization in the presence of spatial fluctuations leads to more robust implementations for all considered interatomic distances. This potentially enables gate implementations with strongly suppressed next-nearest interactions, resulting in more straight-forward implementations.

Contribution

CB, LM, LB and I conceptualized this work. NP, TP, KS, KI and CB provided information about the specific experimental boundaries and realizations of the method and guided with the development of experimentally feasible protocols. LB and I created the numerical model and the GRAPE optimizer. I performed the numerical analysis, and created the visualizations. I created the noise model and performed the robustness analysis. LB performed the analytical calculations on the dynamical Lie algebra. All authors contributed to the discussion and interpretation of the results, as well as to

writing the manuscript. This collaboration was realized under the supervision of CB and LM.

Quantum Gate Optimization for Rydberg Architectures in the Weak-Coupling Limit

Nicolas Heimann,[1,2,3,*] Lukas Broers,[1,2] Nejira Pintul,[1,2] Tobias Petersen,[1,2] Koen Sponselee,[1,2] Alexander Ilin,[1,2,3] Christoph Becker,[1,2] and Ludwig Mathey[1,2,3]

[1] *Zentrum für Optische Quantentechnologien, Universität Hamburg, 22761 Hamburg, Germany*
[2] *Institut für Quantenphysik, Universität Hamburg, 22761 Hamburg, Germany*
[3] *The Hamburg Centre for Ultrafast Imaging, 22761 Hamburg, Germany*

We demonstrate machine learning assisted design of a two-qubit gate in a Rydberg tweezer system. Two low-energy hyperfine states in each of the atoms represent the logical qubit and a Rydberg state acts as an auxiliary state to induce qubit interaction. Utilizing a hybrid quantum-classical optimizer, we generate optimal pulse sequences that implement a CNOT gate with high fidelity, for experimentally realistic parameters and protocols, as well as realistic limitations. We show that local control of single qubit operations is sufficient for performing quantum computation on a large array of atoms. We generate optimized strategies that are robust for both the strong-coupling, blockade regime of the Rydberg states, but also for the weak-coupling limit. Thus, we show that Rydberg-based quantum information processing in the weak-coupling limit is a desirable approach, being robust and optimal, with current technology.

I. INTRODUCTION

Rydberg tweezer arrays have evolved into an intriguing and promising platform for quantum computing [1–3] and quantum simulation [4, 5]. These devices support the preparation of scalable, nearly defect-free systems [6–8], high fidelity single-qubit operations [9] and the implementation of two-qubit gates via Rydberg states [10–14]. This includes the quantum gate design based on Rydberg blockade, that corresponds to the widely explored regime of strong van-der-Waals interaction strength and small interatomic distances. Furthermore, qubit architectures based on alkaline-earth and alkaline-earth-like atoms [15], such as strontium [16, 17] and ytterbium atoms [18–20], have desirable features such as long-lived decoupled nuclear spin states that are suitable to be used as qubit states, as well as single-photon Rydberg transitions for implementing fast two-qubit gates. The existence of a meta-stable clock state further allows for elaborate qubit schemes allowing novel error correction strategies and shelving operations for non-destructive mid-circuit readout [21]. Further design options include triple magic trapping of qubit and Rydberg states [22], and local Rydberg control via manipulation of inner shell electrons [23].

Optimization methods, such as quantum machine learning and quantum optimal control are a powerful and versatile approach of operating and controlling quantum dynamics in a way that is optimal or near-optimal according to a desired metric. In particular, variational quantum algorithms [24–26] are a class of algorithms which utilize a generalized quantum circuit with parameterized gates to transform the synthesis of quantum algorithm solutions into an optimization problem. This approach can be extended towards quantum optimal control [27–31] and has been utilized in different noisy intermediate-scale quantum devices [32], such as trapped ions [33–36], superconducting qubits [37–40] and neutral atoms [41–45]. Recently, time-optimal gates have been constructed using quantum optimal control [41] and realized experimentally [46].

In this paper, we demonstrate machine learning assisted design of a controlled-not (CNOT) gate in Rydberg tweezer systems. The logical qubit states are implemented in two hyperfine states of the atoms, which are controlled via Raman pulses. Additionally we consider a Rydberg state in each atom, which can be Rabi-driven from one of the hyperfine states. We demonstrate that using either a global Rabi protocol, driving the Rydberg transition of all atoms, and individual Raman protocols, driving the hyperfine transition of individual atoms, or a global Raman protocol and individual Rabi protocols, are sufficient to support universal quantum computing. We focus on the case of a global Rabi protocol and individual Raman protocols. The parameters of the atomic states and the magnitudes of the Rabi and Raman protocols, as well as an applied magnetic field, are modeled after ^{171}Yb atom tweezers. However, we emphasize that our analysis and results are directly applicable to all Rydberg tweezer systems, as they include realistic conditions of operation of current devices. We consider a fixed total operation time, and determine fidelity-optimal implementations based on a hybrid quantum-classical optimizer algorithm. We use the van-der-Waals interaction strength as a variable parameter. We identify the minimal van-der-Waals interaction that supports an implementation of a CNOT gate with high fidelity, and find that the fidelity saturates beyond that magnitude. We determine the robustness of our optimal implementations with respect to fluctuations of the distance between the atoms. We find that the implementations are not only robust in the blockade but also in the weak-coupling limit. We propose this regime to be utilized for robust optimal quantum computing under realistic conditions with current technology.

This paper is organized as follows. In Sect. II, we introduce the model and method used throughout the

* nheimann@physnet.uni-hamburg.de

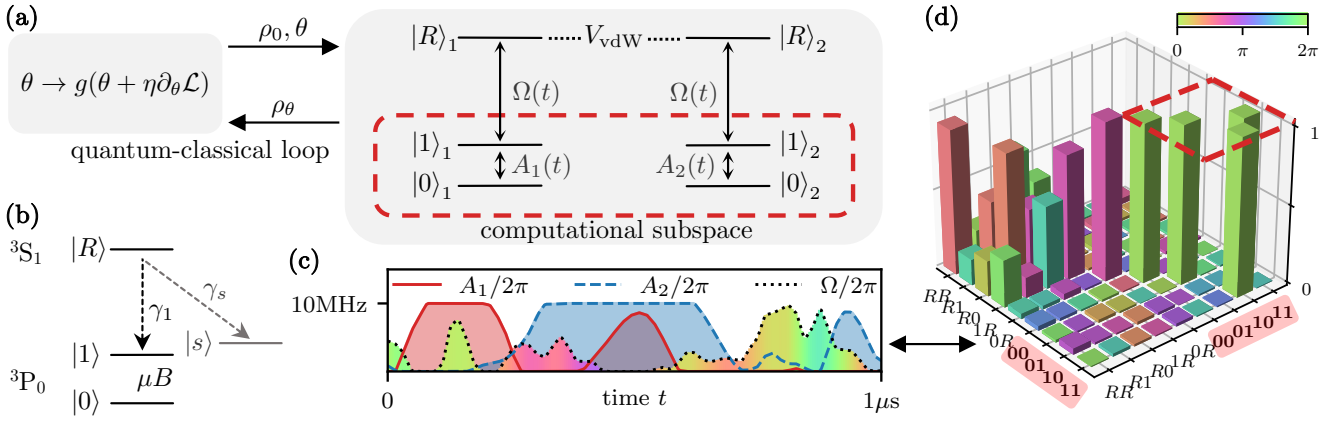


Figure 1. **Optimization platform. (a)** Hybrid quantum-classical optimization scheme. The hyperfine states $|0\rangle$ and $|1\rangle$ act as logical states, forming the computational subspace depicted as red-dashed lines, and can be manipulated individually by the control Raman protocol $A_1(t)$ and the target Raman protocol $A_2(t)$, in which qubit 1 is the control qubit and qubit 2 is the target qubit. The Rabi protocol $\Omega(t)$ controls transfer between the state $|1\rangle$ and a highly excited Rydberg state $|R\rangle$. The protocols $\Omega(t)$, $A_j(t)$ and B are parameterized by the transformation parameters $\theta = \{\theta_i\}$. The atoms are coupled via the van-der-Waals interaction $V_{\rm vdW}$ of the Rydberg states. For a transformation parameter set θ, an initial state ρ_0 is propagated yielding the final state ρ_θ. This time propagation in the quantum unit is controlled by a classical unit synthesising the loss function $\mathcal{L}$ from ρ_θ, allowing to optimize the transformation parameters θ. **(b)** Level diagram and dissipation channels, for our main example of ^{171}Yb. We consider decay from $|R\rangle$ into $|1\rangle$, as well as into $|s\rangle$ which is otherwise decoupled dynamically, and is introduced to model population loss. The magnetic field B defines the Zeeman splitting of the two hyperfine states. **(c)** A high fidelity protocol implementing the CNOT transformation. The Rabi protocol $\Omega(t)$ and the Raman protocols $A_j(t)$ are constrained by a maximal frequency of $\Omega_{\rm max} = A_{\rm max} = 2\pi \times 10$MHz. The phase $\phi(t)$ of the Rabi protocol $\Omega(t)$ is depicted by the filling color. **(d)** The transformation U corresponding to the pulse sequence shown in (c), depicted at the time $\tau = 1\mu$s. The CNOT operation is clearly visible in the computational subspace, enclosed by the red-dashed square.

manuscript. In Sect. III, we present the performance and the protocols of the hybrid quantum-classical optimizer within the weak-coupling limit. In Sect. IV, we show how spatial fluctuations affect the gate fidelity for a realistic range of interatomic distances. In Sect. V, we conclude.

II. MODEL

We consider neutral atoms trapped individually in optical tweezers. For each of the atoms we consider two long-lived, low-energy states that constitute a qubit, written as $|0\rangle$ and $|1\rangle$. Additionally we consider a highly excited Rydberg state $|R\rangle$, and a generic state $|s\rangle$ that we use to model the decay of the Rydberg state. The Rydberg state $|R\rangle$ is utilized for its strong van-der-Waals interaction between two atoms in this state, providing a non-linearity to design two-qubit gates. We consider the Hamiltonian

$$H = \sum_j H_j + \sum_{i,j} V_{\rm vdW}^{i,j} \left|R\right\rangle_i \left|R\right\rangle_j \left\langle R\right|_i \left\langle R\right|_j , \qquad (1)$$

where

$$V_{\rm vdW}^{i,j} = \frac{\hbar C_6}{|r_i - r_j|^6} \qquad (2)$$

is the van-der-Waals interaction between the ith and jth atom at the respective positions r_i and r_j. C_6 is the coefficient of the van-der-Waals interaction and depends on the specific atom species and Rydberg state. We choose $C_6 = 1\text{THz} \cdot \mu\text{m}^6$, as inspired by [15, 20], as a typical magnitude for Rydberg atoms of two-electron atoms. H_j is the local Hamiltonian of the jth atom. It is

$$H_j = \frac{\hbar}{2}\begin{pmatrix} 0 & \Omega(t) & 0 \\ \Omega^*(t) & 0 & A_j(t) \\ 0 & A_j(t) & 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & \mu B & 0 \\ 0 & 0 & -\mu B \end{pmatrix}, \qquad (3)$$

and operates on the states $\{|R\rangle, |1\rangle, |0\rangle\}$. The two logical qubit states $|0\rangle$ and $|1\rangle$ are realized as hyperfine states of the atom, and μB is the Zeeman splitting between them induced by an external magnetic field B, where 2μ is the difference of the magnetic moments. The corresponding Zeeman shift of the Rydberg state is normalized to zero in the rotating frame. $A_j(t)$ is the Raman coupling between the logical qubit states $|0\rangle_j$ and $|1\rangle_j$ of atom j, which derives from a two-photon transition, and which we consider to be real-valued. This assumption is realized by optimizing the excitation light homogeneity to select the phases of each coupling $A_j(t)$ [8] to zero. $\Omega(t) = |\Omega(t)|e^{-i\phi(t)}$ is the complex-valued Rabi coupling between the levels $|1\rangle_j$ and $|R\rangle_j$ for all j, i.e. in a global fashion. As discussed in App. D, we show that using either a global Rabi coupling and individual Raman couplings, or indvidual Rabi couplings and a global Raman coupling, is sufficient for universal quantum computing. We illustrate the hybrid quantum-classical optimizer in Fig. 1. For a given optimization task, we

limit the Rabi and Raman coupling by maximal values of $\Omega_{\max}$ and $A_{\max}$. Here, we focus on the case of $\Omega_{\max} = A_{\max} = 2\pi \times 10\text{MHz}$, but note that our approach equally applies to the general case, i.e. $\Omega_{\max} \neq A_{\max}$. In App. C, we show that a value of $\Omega_{\max} = 2\pi \times 10\text{MHz}$ is sufficient for our analysis. The maximum gradient of the Rabi phase $\phi(t)$ is $\partial_t \phi_{\max} = \pm\pi/100\text{ns}$, motivated by typical acousto-optic modulator bandwidths. The magnetic field B is assumed to be stationary, but can be chosen arbitrarily within the range $B_{\min} \leq B \leq B_{\max}$, where $B_{\min} = 100\text{G}$ and $B_{\max} = 200\text{G}$ which corresponds to $0.1\text{MHz} \leq \mu B/h \leq 0.2\text{MHz}$.

The finite lifetime $1/\gamma$ of the Rydberg state leads to decoherence. We consider two contributions to the decay, black body radiation and spontaneous decay [47]. Both black body radiation and spontaneous decay induce transitions out of the Rydberg state $|R\rangle$ to states other than $|0\rangle$, $|1\rangle$, and $|R\rangle$. We model these states with an auxiliary state $|s\rangle$. Spontaneous decay also induces transitions from the Rydberg state $|R\rangle$ to the state $|1\rangle$, i.e. the computational subspace. In Fig. 1 (b) we illustrate this effective dissipation model. The dynamics of the system are governed by the Lindblad master equation

$$\dot{\rho} = -\frac{i}{\hbar}[H, \rho] + \sum_{ij} \mathcal{D}[L_i^j]\rho, \tag{4}$$

where $\mathcal{D}[L]\rho = L\rho L^\dagger - \frac{1}{2}\{L^\dagger L, \rho\}$, with the Lindblad operators $L_j^s = \sqrt{\gamma_s}\,|s\rangle_j \langle R|_j$ and $L_j^1 = \sqrt{\gamma_1}\,|1\rangle_j \langle R|_j$ of the jth atom. The total decay rate of the Rydberg state obeys $\gamma = \gamma_s + \gamma_1$ where $\gamma_s = 20\gamma_1$ [21]. Here we choose typical values of the lifetime of the Rydberg state of $1/\gamma = 10\mu\text{s}$, $100\mu\text{s}$ and $500\mu\text{s}$ [15]. We assume magic-trapping between $|1\rangle$ and $|R\rangle$ and neglect losses arising from turning off the trap during gate operations as well as dephasing contributions [18, 48].

Gradient Ascent Pulse Engineering (GRAPE) [49] is a quantum optimal control technique to construct pulse sequences, which determine the dynamical evolution of the system, such that a desired target transformation U is realized. Note that we employ this method for non-unitary dynamics, given in Eq. 4. We consider a general Hamiltonian $H_\theta(t) = H_0 + \sum_k f_k(t;\theta)h_k + \text{h.c.}$, where $f_k(t;\theta)$ are complex-valued functions, $\theta = \{\theta_i\}$ are parameters, which are to be optimized, and the h_k are hermitian operators. The transformation parameters θ correspond to a transformation which we evaluate for a given state ρ_0 by integrating Eq. (4) over the algorithm time τ. We denote the final state a particular parameter set θ as ρ_θ. Throughout this work, we fix the algorithm time to $\tau = 1\mu\text{s}$. The optimization is performed with respect to the objective, i.e. the loss function, which in our case we define as

$$\mathcal{L} = 1 - F_\theta = 1 - \frac{1}{4}\langle |\text{Tr}(\rho_\theta^\dagger P U \rho_0 U^\dagger P)| \rangle_{\rho_0}, \tag{5}$$

where F_θ is the fidelity and $P = \sum_q |q\rangle\langle q|$ is the projector onto the computational subspace, which is $P = |00\rangle\langle 00| + |01\rangle\langle 01| + |10\rangle\langle 10| + |11\rangle\langle 11|$. $\langle\cdot\rangle_{\rho_0}$ is the average over 32 initial random product states $\rho_0 = \bigotimes_i \rho_i$ sampled from the Bloch spheres of the computational subspaces. The batch size of 32 is an empirical value that provides efficient optimization. The optimal transformation parameters $\theta_{\text{opt}} = \text{argmin}_\theta \mathcal{L}$ are inferred via stochastic gradient descent [50]. First, the loss $\mathcal{L}$ is evaluated given the transformation parameters θ. Next, the parameters are varied as $\theta_i \to \theta_i + \epsilon$ by a small amount $\epsilon = 10^{-8}$, and subsequently the modified loss $\mathcal{L}_i$ is evaluated. The first order gradient is approximated by the finite difference $\partial\mathcal{L}/\partial\theta_i = (\mathcal{L}_i - \mathcal{L})/\epsilon$ and the parameters are then updated as

$$\theta_i \to g_i\left(\theta_i + \eta_i \frac{\partial \mathcal{L}}{\partial \theta_i}\right), \tag{6}$$

where η_i are dynamically adapted learning rates according to the ADAM method [51]. The functions g_i impose constraints on the protocols. Note that these constraints do not affect the gradient. We refer to this step in the algorithm as a training epoch and illustrate this in Fig. 1 (a). Optimization occurs by iterating over the training epochs until convergence.

The central example that we apply this optimization method to, is the optimal implementation of the CNOT gate. So the number of atoms $N_a = 2$. However, we emphasize that the methodology presented here naturally applies to atom systems with larger numbers. The notion of a global Rabi coupling implies that for $N_a > 2$, any additional atom besides the two involved in the CNOT operation will also experience the global coupling $\Omega(t)$. This results in a transformation on these other qubits, which may be undesired. Our optimization method can also be utilized to learn a coupling $A_{j>2}(t)$ that implements the identity operation, in the presence of the fixed global coupling $\Omega(t)$. We emphasize that this is possible, because even in the case of arbitrarily many neutral atoms, a single global Rabi coupling is sufficient for universal quantum computing. For example, the resulting transformation on the other qubits can be mitigated efficiently by moving the other atoms sufficiently far apart such that the van-der-Waals interaction becomes negligible while additionally applying the control coupling $A_1(t)$. By construction of the CNOT gate, the control coupling $A_1(t)$ will transform the qubit states $|0\rangle$ and $|1\rangle$ into themselves, respectively. This will in general only result in a relative phase between these states which can be corrected. Alternatively, for alkaline-earth-like atoms like ^{1}71Yb, the omg qubit architecture [15] can be employed to realize local two-qubit gates despite the global coupling $\Omega(t)$. Since the Rydberg excitation originates from the meta-stable $^3P_0$ state it is straight forward to site-selectively shelve atoms in the ground state qubit $^1S_0$ if the CNOT gate is not desired. We note that universal quantum computing is equally possible in the case in which there is a global Raman coupling that is equal for all atoms, and the Rabi couplings are applied to the atoms individually. This result is both conceptually in-

teresting, as well as of experimental relevance, because it suggests an alternative, minimal set of experimental control parameters. We expand on this implementation approach elsewhere. In this work, we focus on the case of a global Rabi pulse. For details on the computational universality under these constraints, see App. D.

We optimize the transformation parameters θ, which parameterize $\Omega(t)$ and $A_j(t)$ as stepwise functions which we linearly interpolate in the dynamics, as well as B which we consider to be constant during the time evolution such that it is represented by a single parameter. We refer to these parameterizations as the Rabi protocol $\Omega(t)$, the control Raman protocol $A_1(t)$ and the target Raman protocol $A_2(t)$. We give a detailed account of the parameterization in App. A. We construct the initial protocols $|\Omega^0(t)|$ and $A_j^0(t)$ to be positive and slowly varying. The initial phase of the Rabi protocol $\phi^0(t)$ is generated via a random walk starting at $\phi^0(0) = 0$, see App. B.

In Fig. 1 (c) we illustrate the Rabi protocol $\Omega(t)$ and the Raman protocols $A_j(t)$ of a high fidelity CNOT realization. In Fig. 1 (d) we show the transformation corresponding to this high fidelity realization. The CNOT transformation is visible in the computational subspace, while the transformation on the remaining subspace is arbitrary.

III. WEAK-COUPLING SOLUTIONS

In this section, we identify optimal implementations of the CNOT gate in the weak-coupling limit, i.e. based on dynamical phase accumulation. We consider a large interatomic distance $r \approx 10\mu$m, of the two atoms. For this distance, the van-der-Waals interaction is small compared to the maximal Rabi frequency $V_{\mathrm{vdW}} \ll \hbar\Omega_{\max}$ allowing for occupation of the Rydberg-Rydberg state $|R\rangle \otimes |R\rangle$, where $V_{\mathrm{vdW}} = V_{\mathrm{vdW}}^{1,2}(r)$, based on the van-der-Waals interaction in Eq. 2, with $r = |r_1 - r_2|$. Hence, the magnitude of the nonlinearity is limited by the algorithm time τ and the interaction strength V_{vdW}. We introduce the dimensionless gate action

$$\Phi = \tau V_{\mathrm{vdW}}/\hbar, \tag{7}$$

as the maximally achievable non-linear phase accumulation. Note that as the algorithm time $\tau = 1\mu$s is fixed, the gate action Φ is equivalently a measure of the interaction strength. In this section we treat the interaction strength V_{vdW} as an external parameter rather than a trainable parameter.

In Fig. 2 (a) we show the average of the infidelity $1 - \langle F\rangle_\theta$ over 15 optimized protocols [52] for the target transformation of the CNOT gate. We show this as a function of the gate action Φ and the number of training epochs in the absence of dissipation, i.e. for $\gamma = 0$. We find that for small gate actions $\Phi \lesssim 2\pi$ the optimization algorithm does not generate a high-fidelity protocol. The fidelity steadily increases with increasing gate action Φ. We fit the expression $1-\langle F\rangle_\theta = A(\Phi-\Phi_c)^2$ in the vicinity of the critical gate action Φ_c and find that $A = 0.37$ and $\Phi_c = 2.018\pi$. For values of $2\pi < \Phi < 3\pi$, the infidelity converges to approximately $1 - \langle F\rangle_\theta \approx 7 \times 10^{-3}$, which indicates sufficient gate action Φ, i.e. it indicates that sufficient time and interaction is provided to generate a two-qubit operation. For gate actions of values $\Phi > 3\pi$ the infidelity decreases further as it converges to approximately $1 - \langle F\rangle_\theta \approx 3 \times 10^{-3}$. In this regime, we observe more efficient optimization behavior that reaches values of $1 - \langle F\rangle_\theta < 10^{-2}$ after roughly 200 training epochs. In the case of no interaction, the transformation consists of single-qubit transformations that cannot implement the CNOT operation.

In Fig. 2 (b) we show representations of transformations in the computational subspace for $\Phi = \pi/2$ and $\Phi = 4\pi$. In the case of insufficient interaction strength, for $\Phi = \pi/2$, the implemented transformation is visibly distinct from a CNOT gate. In the case of sufficiently large interaction strength, for $\Phi = 4\pi$, a high fidelity implementation of the CNOT gate is visible. In Fig. 2 (c) we show the infidelity of optimized protocols as a function of the gate action Φ, with and without dissipation. We use the dissipative parameters discussed in Sect. II. For small values of Φ, the gate fidelity is independent of dissipation, as the Rydberg state is weakly occupied during the protocol. For increasing values of Φ, the protocols approach high fidelities with dissipation, but with an increased infidelity. This increase of the infidelity is also visible along the learning trajectory for $\Phi = 4\pi$ as we show in Fig. 2 (d). Here we see that dissipation results in a lower bound of the infidelity of the optimized protocol. This lower bound is reduced by minimizing the occupation time of the Rydberg states $|R\rangle_j$. The maximal Rabi frequency $\Omega_{\max}$ provides a limitation of this optimization in the case of a fixed algorithm time τ.

We find that the optimization method provides high fidelity protocols in the presence of experimentally motivated dissipation. Generally, higher fidelities than what we present can be achieved by increasing the number of training epochs. In a realistic setup, measurement noise, laser phase- and intensity noise, and spatial fluctuations are additional challenges, that can be included in our optimization approach.

IV. SPATIAL FLUCTUATIONS

To demonstrate the robustness properties of the optimal implementations that we have obtained, we include fluctuations of the distance between the two Rydberg atoms. In an experimental realization, these fluctuations might derive from thermal motion of each of the atoms in the tweezer potentials, or fluctuations of the tweezer potential itself. We consider spatial distances between the two atoms of 4µm to 11µm, which interpolates between the blockade regime and the weak-coupling limit.

We consider a high-fidelity implementation U of the

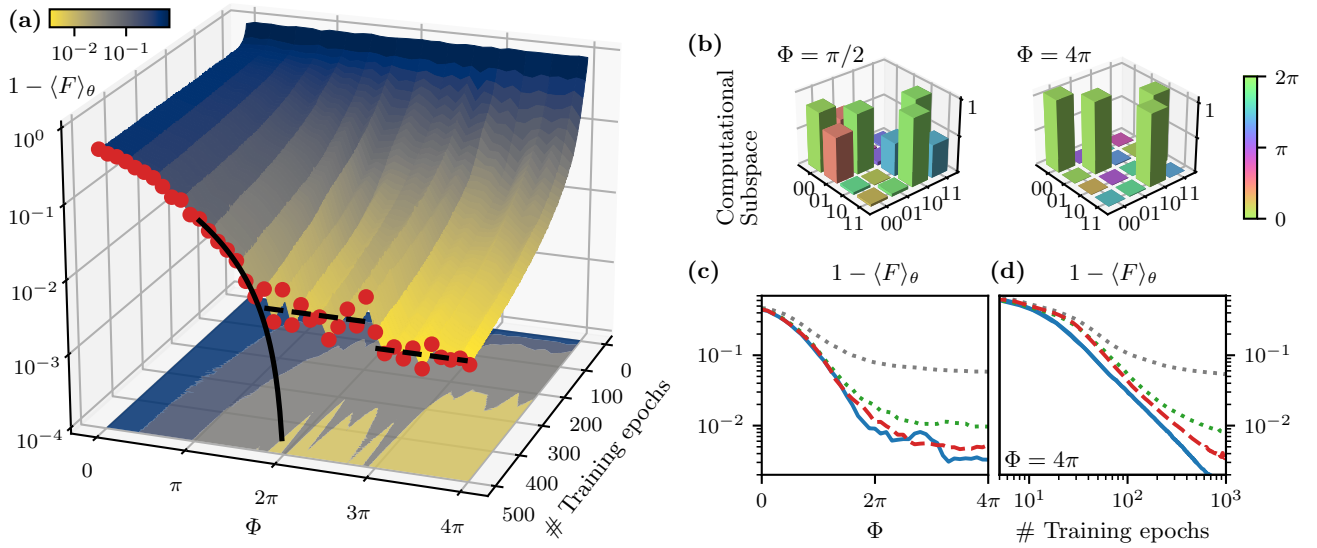


Figure 2. **Gate optimization in the weak-coupling limit. (a)** We display the infidelity $1-\langle F\rangle_\theta$ as a function of the gate action Φ and the number of training epochs. The infidelity after 500 training epochs is depicted as red dots. Near $\Phi_c \approx 2\pi$, we fit the infidelity $1-\langle F\rangle_\theta$ with the fitting function $\Phi = \sqrt{(1-\langle F\rangle_\theta)/A} + \Phi_c$, which we depict as a black solid line, while the black-dashed lines depict the converged infidelities of 7×10^{-3} for $2\pi < \Phi < 3\pi$ and 3×10^{-3} for $\Phi > 3\pi$. **(b)** Optimized gate transformations in the computational subspace. For a gate action of $\Phi = \pi/2$ the infidelity is $1-\langle F\rangle_\theta = 0.25$, indicating that the gate action is insufficient to create a high fidelity protocol. For $\Phi = 4\pi$ we show a transformation having an infidelity of $1-\langle F\rangle_\theta = 3\times10^{-4}$, indicating sufficient gate action. The phases of the matrix elements are encoded on a cyclic color map. **(c-d)** Infidelity without dissipation (blue) and with dissipation of values $1/\gamma = 500$µs (red-dashed), $1/\gamma = 100$µs (green-dotted) and $1/\gamma = 10$µs (grey-dotted). **(c)** Infidelity after 500 training epochs as a function of the gate action Φ. **(d)** The infidelity during training for a fixed gate action $\Phi = 4\pi$ as a function of the number of training epochs. Dissipation results in slower reduction of the infidelity with the number of training epochs and determines the lower bound of the infidelity that is visible for large dissipation, i.e. for large γ.

CNOT gate, which has been optimized for a specific interaction strength $V_{\rm vdW}$ and zero dissipation $\gamma = 0$. We now introduce fluctuations of the atom distance, i.e. $r \to r + \delta r(t)$, in which the spatial fluctuations $\delta r(t)$ are sampled from a normal distribution $N(0, \sigma_r)$ at a frequency of 512MHz. Based on a single, stochastic time series $r + \delta r(t)$, we now determine the modified, time-dependent interaction strength

$$V_{\rm vdW}(t) = \frac{\hbar C_6}{|r + \delta r(t)|^6}. \tag{8}$$

We use this interaction strength to generate the time-evolution $\tilde{U}(\sigma_r)$, while keeping all other features of the protocol unchanged, i.e. we use the same $\Omega(t)$, $A_j(t)$ and B protocol. To quantify to what degree the fidelity is reduced due to the spatial fluctuations, we define the average transformation error as

$$\epsilon(\sigma_r) = 1 - \frac{1}{4}\langle|{\rm Tr}(\tilde{U}(\sigma_r)^\dagger {\rm CNOT})|\rangle_{\delta r(t),\theta}. \tag{9}$$

Here we take the statistical average of the implementation error over 50 sampled trajectories of $\delta r(t)$ and 10 high fidelity protocols provided by transformation parameters θ optimized from different initial values.

In Fig. 3 we show the transformation error $\epsilon(\sigma_r)$ as a function of the interatomic distance r for various values of the standard deviation σ_r. At large distances of about $r > 9$µm, the protocols are only weakly susceptible to spatial fluctuations. Since the van-der-Waals interaction in Eq. 2 scales with r^{-6}, the gradient falls off rapidly as the mean distance r increases. Because of this rapid fall-off, fluctuations of r result in a smaller and smaller increase of the error $\epsilon(\sigma_r)$ with increasing r. On the other hand, in the blockade regime at distances of $r < 5$µm, the interaction strength $V_{\rm vdW}$ dominates the maximal Rabi frequency $\Omega_{\rm max}$. In this limit, transitions into the Rydberg-Rydberg state $|R\rangle_1 |R\rangle_2$ are highly suppressed. Therefore, the spatial fluctuations do not induce large errors in this limit either but are more noticeable than in the weak-coupling limit, in this example. However, in the intermediate regime of 5µm $< r < 8$µm, the transformation error $\epsilon(\sigma_r)$ is highly susceptible to spatial fluctuations. At these interatomic distances, the van-der-Waals interaction and the the maximal Rabi frequency are of the same order, i.e. $V_{\rm vdW} \sim \hbar\Omega_{\rm max}$. Hence, the optimized protocols are highly susceptible to spatial fluctuations, making these intermediate interatomic distances undesirable in any realization. The robustness with respect to spatial fluctuations is one of the key features that makes the Rydberg blockade regime attractive for quantum computing purposes [10]. However, we emphasize that in the weak-coupling limit, the system is equally

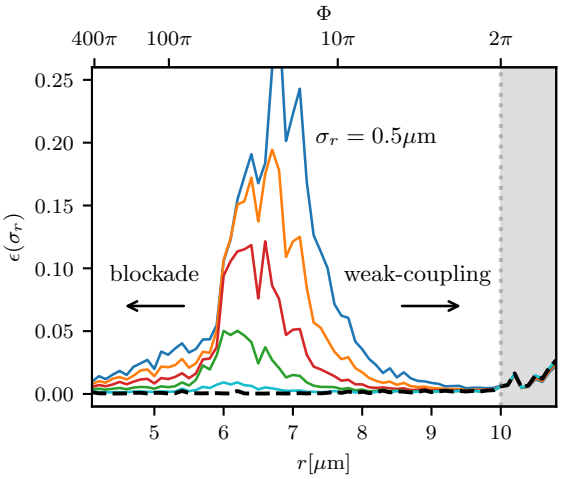


Figure 3. **Robustness against spatial fluctuations.** The transformation error $\epsilon(\sigma_r)$ in the presence of spatial fluctuations, for the noise parameter $\sigma_r = 0.5, 0.4, 0.3, 0.2, 0.1, 0\mu$m (blue, orange, red, green, cyan, black dashed), and as a function of the interatomic distance r. For small ($< 6\mu$m) and large ($> 8\mu$m) distances the optimal implementations are robust against spatial fluctuations. For intermediate distances of $r = 6 - 8\mu$m, the system is strongly susceptible to spatial fluctuations and the relative error grows quickly as a function of σ_r. The vertical dashed line depicts the minimum distance required to implement the CNOT gate, see Fig. 2 for reference.

robust against spatial fluctuations [53].

V. CONCLUSION

In conclusion, we have demonstrated quantum gate optimization of a CNOT gate in a Rydberg architecture under experimentally motivated constraints, via machine learning. The two qubit states are two long-lived hyperfine states of each of the two atoms. Additionally, we include a Rydberg state in each atom in our model, as an auxilliary state to provide a van-der-Waals interaction to create a two-qubit gate. These atoms are held in optical tweezers, at fixed distance. The model and parameter choices are based on 171-Yb, such as the dissipative properties of the Rydberg state. However, we emphasize that our approach is universally applicable to Rydberg architectures. We assume that the long-lived hyperfine states can be driven by Raman protocols, and the transition from one of the hyperfine states to the Rydberg state by a single global Rabi protocol. We show that utilizing either individual Raman protocols for each atom and a global Rabi protocol for both atoms, or individual Rabi protocols for each atom and a global Raman protocol for both atoms, is sufficient for universal quantum computing. Focusing on the case of individual Raman protocols for each atom and a global Rabi protocol, we utilize a hybrid quantum-classical optimization approach, based on gradient ascent pulse engineering (GRAPE), to determine protocols that implement a high fidelity CNOT gate. Keeping the total algorithm time of the protocols fixed at 1µs, we scan the optimal implementations as a function of the interaction strength. Finally, we map out the robustness of the optimal protocols against spatial fluctuations of the interatomic distance. We find that both for the weak-coupling limit and for the blockade regime, the implementations are robust. However, the intermediate regime, at which the maximal Rabi frequency is comparable to the van-der-Waals interaction, is not robust and thus undesirable. Additional imperfections, such as doppler shifts or imperfections in the laser intensities, will be explored elsewhere. We also note that the weak-coupling regime enables gate implementations in tweezer arrays with strongly suppressed next-nearest interactions, resulting in more straight-forward implementations. With these results, we have demonstrated the immediate and significant impact that hybrid quantum-classical optimization, or machine learning inspired methods in general, have on quantum gate design. Going forward, this suggests systematic, large-scale, and in-depth utilization of quantum machine learning methods.

ACKNOWLEDGMENTS

This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - SFB-925 - project 170620586 and the Cluster of Excellence 'Advanced Imaging of Matter' (EXC 2056) project 390715994.

[1] M. Saffman, Quantum computing with atomic qubits and Rydberg interactions: progress and challenges, Journal of Physics B: Atomic, Molecular and Optical Physics **49**, 202001 (2016).

[2] S. R. Cohen and J. D. Thompson, Quantum computing with circular rydberg atoms, PRX Quantum **2**, 030322 (2021).

[3] T. M. Graham, Y. Song, J. Scott, C. Poole, L. Phuttitarn, K. Jooya, P. Eichler, X. Jiang, A. Marra, B. Grinkemeyer, M. Kwon, M. Ebert, J. Cherek, M. T. Lichtman, M. Gillette, J. Gilbert, D. Bowman, T. Ballance, C. Campbell, E. D. Dahl, O. Crawford, N. S. Blunt, B. Rogers, T. Noel, and M. Saffman, Multi-qubit entanglement and algorithms on a neutral-atom quantum computer, Nature **604**, 457 (2022).

[4] M. Morgado and S. Whitlock, Quantum simulation and computing with rydberg-interacting qubits, AVS Quantum Science **3**, 023501 (2021).

[5] A. Browaeys and T. Lahaye, Many-body physics with individually controlled Rydberg atoms, Nature Physics **16**, 132 (2020).

[6] K. Schymik, B. Ximenez, E. Bloch, D. Dreon, A. Signoles, F. Nogrette, D. Barredo, A. Browaeys, and T. Lahaye, In situ equalization of single-atom loading in large-scale optical tweezer arrays, Phys. Rev. A **106**, 022611 (2022).

[7] D. Barredo, S. d. Léséleuc, V. Lienhard, T. Lahaye, and A. Browaeys, An atom-by-atom assembler of defect-free arbitrary two-dimensional atomic arrays, Science **354**, 1021 (2016).

[8] M. Endres, H. Bernien, A. Keesling, H. Levine, E. R. Anschuetz, A. Krajenbrink, C. Senko, V. Vuletic, M. Greiner, and M. D. Lukin, Atom-by-atom assembly of defect-free one-dimensional cold atom arrays, Science **354**, 1024 (2016).

[9] I. S. Madjarov, J. P. Covey, A. L. Shaw, J. Choi, A. Kale, A. Cooper, H. Pichler, V. Schkolnik, J. R. Williams, and M. Endres, High-fidelity entanglement and detection of alkaline-earth rydberg atoms, Nature Physics **16**, 857 (2020).

[10] D. Jaksch, J. I. Cirac, P. Zoller, S. L. Rolston, R. Côté, and M. D. Lukin, Fast Quantum Gates for Neutral Atoms, Physical Review Letters **85**, 2208 (2000).

[11] M. D. Lukin, M. Fleischhauer, R. Cote, L. M. Duan, D. Jaksch, J. I. Cirac, and P. Zoller, Dipole Blockade and Quantum Information Processing in Mesoscopic Atomic Ensembles, Physical Review Letters **87**, 037901 (2001).

[12] M. Saffman and T. G. Walker, Analysis of a quantum logic device based on dipole-dipole interactions of optically trapped Rydberg atoms, Physical Review A **72**, 022347 (2005).

[13] E. Urban, T. A. Johnson, T. Henage, L. Isenhower, D. D. Yavuz, T. G. Walker, and M. Saffman, Observation of Rydberg blockade between two atoms, Nature Physics **5**, 110 (2009).

[14] T. Wilk, A. Gaëtan, C. Evellin, J. Wolters, Y. Miroshnychenko, P. Grangier, and A. Browaeys, Entanglement of Two Individual Neutral Atoms Using Rydberg Blockade, Physical Review Letters **104**, 010502 (2010).

[15] N. Chen, L. Li, W. Huie, M. Zhao, I. Vetter, C. H. Greene, and J. P. Covey, Analyzing the rydberg-based optical-metastable-ground architecture for ^{171}Yb nuclear spins, Phys. Rev. A **105**, 052438 (2022).

[16] A. Cooper, J. P. Covey, I. S. Madjarov, S. G. Porsev, M. S. Safronova, and M. Endres, Alkaline-Earth Atoms in Optical Tweezers, Physical Review X **8**, 041055 (2018).

[17] M. A. Norcia, A. W. Young, and A. M. Kaufman, Microscopic Control and Detection of Ultracold Strontium in Optical-Tweezer Arrays, Physical Review X **8**, 041054 (2018).

[18] J. T. Wilson, S. Saskin, Y. Meng, S. Ma, R. Dilip, A. P. Burgers, and J. D. Thompson, Trapping alkaline earth rydberg atoms optical tweezer arrays, Phys. Rev. Lett. **128**, 033201 (2022).

[19] A. Jenkins, J. W. Lis, A. Senoo, W. F. McGrew, and A. M. Kaufman, Ytterbium nuclear-spin qubits in an optical tweezer array, Phys. Rev. X **12**, 021027 (2022).

[20] S. Ma, A. P. Burgers, G. Liu, J. Wilson, B. Zhang, and J. D. Thompson, Universal Gate Operations on Nuclear Spin Qubits in an Optical Tweezer Array of Yb 171 Atoms, Physical Review X **12**, 021028 (2022).

[21] Y. Wu, S. Kolkowitz, S. Puri, and J. Thompson, Erasure conversion for fault-tolerant quantum computing in alkaline earth rydberg atom arrays, Nature Communications **13**, 4657 (2022).

[22] T. Topcu and A. Derevianko, Possibility of triple magic trapping of clock and rydberg states of divalent atoms in optical lattices, Journal of Physics B: Atomic, Molecular and Optical Physics **49**, 144004 (2016).

[23] A. P. Burgers, S. Ma, S. Saskin, J. Wilson, M. A. Alarcón, C. H. Greene, and J. D. Thompson, Controlling rydberg excitations using ion-core transitions in alkaline-earth atom-tweezer arrays, PRX Quantum **3**, 020326 (2022).

[24] M. Cerezo, A. Arrasmith, R. Babbush, S. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. Coles, Variational quantum algorithms, Nature Reviews Physics (2021).

[25] A. Choquette, A. Di Paolo, P. K. Barkoutsos, D. Sénéchal, I. Tavernelli, and A. Blais, Quantum-optimal-control-inspired ansatz for variational quantum algorithms, Phys. Rev. Res. **3**, 023092 (2021).

[26] Z. Yang, A. Rahmani, A. Shabani, H. Neven, and C. Chamon, Optimizing variational quantum algorithms using pontryagin's minimum principle, Phys. Rev. X **7**, 021027 (2017).

[27] C. Brif, R. Chakrabarti, and H. Rabitz, Control of quantum phenomena: past, present and future, New Journal of Physics **12**, 075008 (2010).

[28] T. Caneva, T. Calarco, and S. Montangero, Chopped random-basis quantum optimization, Physical Review A **84**, 022326 (2011).

[29] A. B. Magann, C. Arenz, M. D. Grace, T. Ho, R. L. Kosut, J. R. McClean, H. A. Rabitz, and M. Sarovar, From pulses to circuits and back again: A quantum optimal control perspective on variational quantum algorithms, PRX Quantum **2**, 010101 (2021).

[30] C. P. Koch, U. Boscain, T. Calarco, G. Dirr, S. Filipp, S. Glaser, R. Kosloff, S. Montangero, T. Schulte-Herbrueggen, D. Sugny, and F. Wilhelm, Quantum optimal control in quantum technologies. strategic report on current status, visions and goals for research in europe, EPJ Quantum Technology **9** (2022).

[31] L. Broers and L. Mathey, Reducing barren plateaus in quantum algorithm protocols (2021), arXiv:2111.08085.

[32] J. Preskill, Quantum Computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[33] M. Kang, Q. Liang, B. Zhang, S. Huang, Y. Wang, C. Fang, J. Kim, and K. R. Brown, Batch optimization of frequency-modulated pulses for robust two-qubit gates in ion chains, Phys. Rev. Appl. **16**, 024039 (2021).

[34] C. Figgatt, A. Ostrander, N. M. Linke, K. A. Landsman, D. Zhu, D. Maslov, and C. Monroe, Parallel entangling operations on a universal ion-trap quantum computer, Nature **572**, 368 (2019).

[35] T. Choi, S. Debnath, T. A. Manning, C. Figgatt, Z.-X. Gong, L.-M. Duan, and C. Monroe, Optimal quantum control of multimode couplings between trapped ion qubits for scalable entanglement, Phys. Rev. Lett. **112**, 190502 (2014).

[36] V. Nebendahl, H. Häffner, and C. F. Roos, Optimal control of entangling operations for trapped-ion quantum computing, Phys. Rev. A **79**, 012312 (2009).

[37] M. Werninghaus, D. J. Egger, F. Roy, S. Machnes, F. K. Wilhelm, and S. Filipp, Leakage reduction in fast superconducting qubit gates via optimal control, npj Quantum

Information **7**, 14 (2021).
[38] S. Huang and H. Goan, Optimal control for fast and high-fidelity quantum gates in coupled superconducting flux qubits, Phys. Rev. A **90**, 012318 (2014).
[39] D. J. Egger and F. K. Wilhelm, Optimized controlled-z gates for two superconducting qubits coupled through a resonator, Superconductor Science and Technology **27**, 014001 (2013).
[40] P. Rebentrost and F. K. Wilhelm, Optimal control of a leaking qubit, Phys. Rev. B **79**, 060507 (2009).
[41] S. Jandura and G. Pupillo, Time-Optimal Two- and Three-Qubit Gates for Rydberg Atoms, Quantum **6**, 712 (2022).
[42] M. Mohan, R. de Keijzer, and S. Kokkelmans, Robust control and optimal rydberg states for neutral atom two-qubit gates (2022), arXiv:2212.10159.
[43] S. Jandura, J. D. Thompson, and G. Pupillo, Optimizing rydberg gates for logical-qubit performance, PRX Quantum **4** (2023).
[44] M. H. Goerz, E. J. Halperin, J. M. Aytac, C. P. Koch, and K. B. Whaley, Robustness of high-fidelity rydberg gates with single-site addressability, Phys. Rev. A **90**, 032329 (2014).
[45] M. M. Müller, D. M. Reich, M. Murphy, H. Yuan, J. Vala, K. B. Whaley, T. Calarco, and C. P. Koch, Optimizing entangling quantum gates for physical systems, Phys. Rev. A **84**, 042315 (2011).
[46] S. J. Evered, D. Bluvstein, M. Kalinowski, S. Ebadi, T. Manovitz, H. Zhou, S. H. Li, A. A. Geim, T. T. Wang, N. Maskara, H. Levine, G. Semeghini, M. Greiner, V. Vuletic, and M. D. Lukin, High-fidelity parallel entangling gates on a neutral atom quantum computer (2023), arXiv:2304.05420.
[47] I. Cong, H. Levine, A. Keesling, D. Bluvstein, S. Wang, and M. D. Lukin, Hardware-efficient, fault-tolerant quantum computation with rydberg atoms, Phys. Rev. X **12**, 021049 (2022).
[48] S. Zhang, F. Robicheaux, and M. Saffman, Magic-wavelength optical traps for rydberg atoms, Phys. Rev. A **84**, 043408 (2011).
[49] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, Optimal control of coupled spin dynamics: design of NMR pulse sequences by gradient ascent algorithms, Journal of Magnetic Resonance **172**, 296 (2005).
[50] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016).
[51] D. Kingma and J. Ba, Adam: A method for stochastic optimization, International Conference on Learning Representations (2014).
[52] We choose the number of optimized protocols, with randomly sampled initial protocols, empirically.
[53] We note that if we expand this optimization to include the optimization of interatomic distances or spatial trajectories of the tweezer locations, this effect may be important to consider. The susceptibility to errors at intermediate distances will create learning pressure away from protocols that cross from the weak-coupling limit at large distances to the Rydberg blockade regime at small distances potentially inhibiting the convergence of the optimization.

Appendix A: Protocols and parameterization

In the following we detail the parameterization of the protocols in our optimization method. We denote the concatenated parameters of the protocol as $\theta = \{\theta_\Omega, \theta_{\partial\phi}, \theta_{A_1}, \theta_{A_2}, \theta_B\}$. We represent the protocols $|\Omega(t)|$ and $A_j(t)$ in a step-wise discretized manner such that the elements of $\vartheta \in \theta_\Omega, \theta_{A_j}$ represent the amplitudes of respective protocols at m discrete points in time. We linearly interpolate these step-wise representations $\vartheta \in \mathbb{R}^m$ on the temporal lattice with the step-size $\Delta t = \tau/(m-1)$. The interpolation is

$$s(\vartheta, t) = (1 - p_i)\vartheta_i + p_i\vartheta_{i+1}, \tag{A1}$$

where $i = \lfloor t/\Delta t \rfloor$ is the latest index corresponding to the time t and $p_i = t/\Delta t - i$ is an interpolation weight. The amplitudes of the Rabi protocol $|\Omega(t)|$, the control Raman protocol $A_1(t)$ and the target Raman protocol $A_2(t)$ are then given by

$$|\Omega(t)| = s(\theta_\Omega, t) \tag{A2}$$
$$A_1(t) = s(\theta_{A_1}, t) \tag{A3}$$
$$A_2(t) = s(\theta_{A_2}, t). \tag{A4}$$

The phase $\phi(t)$ of the Rabi protocol $\Omega(t)$ is given by the stepwise differential parameterization

$$\phi(t) = s(\theta_\phi, t), \tag{A5}$$

where $\theta_{\phi,i} = \theta_{\phi,i-1} + \theta_{\partial\phi,i}$ and $\theta_{\phi,0} = \theta_{\partial\phi,0}$. This construction creates slowly varying phase protocols and avoids sudden phase-jumps. The magnetic field is given by the constant parameterization

$$B = \theta_B. \tag{A6}$$

In the presented analysis we consider a total number of $4 \times 64 + 1$ parameters, that is $\theta \in \mathbb{R}^{257}$. Further, as mentioned in the main text, we constrain the parameters in between minimal and maximal values. The constraints $g_i(\theta_i)$ in Eq. 6 are defined as follows

$$g_\Omega(\theta_{\Omega,i}) = \max(0, \min(\Omega_{\max}, \theta_{\Omega,i})) \tag{A7}$$
$$g_{A_1}(\theta_{A_1,i}) = \max(0, \min(A_{\max}, \theta_{A_1,i})) \tag{A8}$$
$$g_{A_2}(\theta_{A_2,i}) = \max(0, \min(A_{\max}, \theta_{A_2,i})) \tag{A9}$$
$$g_{\partial\phi}(\theta_{\partial_t\phi,i}) = \max(\partial_t\phi_{\min}, \min(\partial_t\phi_{\max}, \theta_{\partial\phi,i})) \tag{A10}$$
$$g_B(\theta_B) = \max(B_{\min}, \min(B_{\max}, \theta_B)), \tag{A11}$$

with $\Omega_{\max} = 2\pi \times 10\text{MHz}$, $A_{\max} = 2\pi \times 10\text{MHz}$, $\partial_t\phi_{\min} = -\pi/100\text{ns}$, $\partial\phi_{\min} = +\pi/100\text{ns}$, $B_{\min} = 100\text{G}$ and $B_{\max} = 200\text{G}$.

Appendix B: Protocol initialization

We construct the initial parameters such that the resulting protocols vary slowly, start and end at zero, and

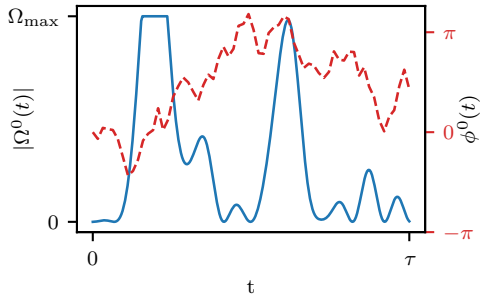


Figure 4. **Protocol initialization.** The initial Rabi protocol $\Omega(t)$ of algorithm time $\tau = 1\mu$s. The amplitude of a sample initial Rabi protocol $|\Omega^0(t)|$ (blue) contains 16 sinosiodal modes and has a maximal Rabi frequency of $\Omega_{\max}$. The initial Raman protocols $A_1^0(t)$ and $A_2^0(t)$ are initialized by the same strategy. The phase of the initial Rabi protocol $\phi^0(t)$ (red-dashed) is given by a random walk starting at $\phi^0(0) = 0$.

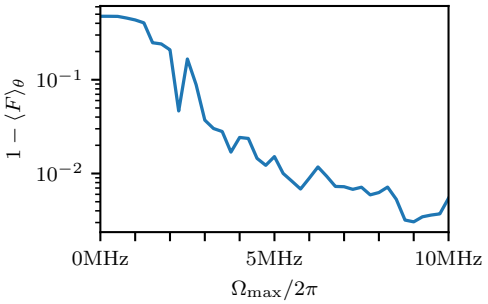


Figure 5. **Maximal Rabi frequency.** High fidelity protocols are realized for $\Omega_{\max} = 2\pi \times 10$MHz, with a gate action of $\Phi = 4\pi$ after 400 training epochs. Lower infidelities can be achieved with more training epochs.

are fairly well-behaved. We first consider a distribution of initial parameterizations

$$S = \left\{ \sum_k \phi_k \sin(k\pi t/\tau) \right\}, \tag{B1}$$

where ϕ_k are random numbers from the uniform distributions $[-1/\sqrt{k}, +1/\sqrt{k}]$. We introduce this dependence on k to emphasize slow modes. We initialize the transformation parameters θ_Ω, θ_{A_1} and θ_{A_2} such that the following initial protocols are realized

$$|\Omega^0(t)| = \Omega_{\text{init}} \min(1, s_\Omega^2(t)) \tag{B2}$$
$$A_1^0(t) = A_{\text{init}} \min(1, s_{A_1}^2(t)) \tag{B3}$$
$$A_2^0(t) = A_{\text{init}} \min(1, s_{A_2}^2(t)), \tag{B4}$$

where we sample s_Ω, s_{A_1} and s_{A_2} from the distribution S, Ω_{init} is the maximal frequency of the initial Rabi protocol and A_{init} is the maximal frequency of the initial Raman protocols. Through out this work we scale the initial protocols by the corresponding maximal frequencies, i.e. $\Omega_{\text{init}} = \Omega_{\max}$ and $A_{\text{init}} = A_{\max}$. The parameters of the phase are initialized as $\theta^0_{\partial_\phi,0} = 0$ and $\theta^0_{\partial_\phi,i}$ is uniformly sampled from $[-\delta, \delta]$. where $\delta \in \mathbb{R}^+$. Here we choose $\delta = 1.5$. Fig. 4 shows an example of a random initial Rabi protocol $\Omega^0(t)$. The initial magnetic field B^0 is sampled from the distribution $[B_{\min}, B_{\max}]$.

Appendix C: Maximal Rabi frequency

The maximal Rabi frequency $\Omega_{\max}$ provides a limitation on the minimally achievable infidelity $1 - F$ of the gate protocol. If $\Omega_{\max}$ is not sufficiently large to complete a Rabi oscillation of one of the states $|1\rangle_1$ or $|1\rangle_2$ to one of the Rydberg states $|R\rangle_1$ or $|R\rangle_2$ during the protocol of algorithm time τ, then no protocols with satisfactory fidelity can be constructed. In Fig. 5 we show the average infidelity $1 - \langle F\rangle_\theta$ over several optimized protocols as a function of the Rabi frequency $\Omega_{\max}$ for a fixed maximal Raman frequency of $A_{\max} = 2\pi \times 10$MHz and a gate action of $\Phi = 4\pi$. For small values of the maximal Rabi frequency $\Omega_{\max} < 2\pi \times 2$MHz, the infidelity displays a plateau of large values of roughly $1 - \langle F\rangle_\theta \approx 0.5$. With increasing $\Omega_{\max} > 2\pi \times 2$MHz, the infidelity decreases. For $\Omega_{\max} > 2\pi \times 8$MHz, the infidelity begins to saturate at values $1 - \langle F\rangle_\theta < 10^{-2}$. Note that the protocols are not necessarily fully converged and lower infidelities can be achieved with more training epochs. Therefore, a maximal Rabi frequency $\Omega_{\max} = 2\pi \times 10$MHz is sufficient for our analysis.

Appendix D: Universal Quantum Computing with Global Pulses

In the following we demonstrate that a single global Rabi coupling for N_a neutral atoms is capable of universal quantum computing. We also show this for the case of a single global Raman coupling and invidiual Rabi couplings. We consider N_a three-level systems consisting of the states $\{|R\rangle, |1\rangle, |0\rangle\}$, and the Hamiltonian

$$H = \mu B \sum_{j=1}^{n} \sigma_z^j + \sum_{j=1}^{n} A_j(t)\sigma_x^j + \sum_{j=1}^{n} (\Omega_x^j(t)\tau_x^j + \Omega_y^j(t)\tau_y^j) + H_I, \tag{D1}$$

where

$$\sigma_x^j = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \sigma_y^j = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \quad \sigma_z^j = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \tag{D2}$$

act on the subspace $\{|1\rangle, |0\rangle\}$ of the jth system.

$$\tau_x^j = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \tau_y^j = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \tau_z^j = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{D3}$$

act on the subspace $\{|R\rangle, |1\rangle\}$ of the jth atom. Analogously, we also define the matrices on the subspace $\{|R\rangle, |0\rangle\}$ of the jth atom as

$$\nu_x^j = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad \nu_y^j = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}. \tag{D4}$$

The interaction term is

$$H_I = \sum_{\langle i,j \rangle} \frac{C_6}{(r_i(t) - r_j(t))^6} |R\rangle_i \otimes |R\rangle_j \langle R|_i \otimes \langle R|_j, \tag{D5}$$

where $r_j(t)$ is the real-space position of the jth atom. μB is the Zeeman splitting due to a constant and global magnetic field B. $A_j(t)$ is the amplitude of the jth Raman coupling of the jth atom that control the transition between $|1\rangle_j$ and $|0\rangle_j$. $\Omega_{x,y}^j(t)$ are the Rabi coupling components of the jth atom that control the transition between $|1\rangle_j$ and $|R\rangle_j$. For convenience we denote the global sums of local operators as $S_z = \sum_{j=1}^n \sigma_z^j$, $T_x = \sum_{j=1}^n \tau_x^j$, $T_y = \sum_{j=1}^n \tau_y^j$ and $V = \sum_{i,j} V_{i,j}$. We consider the base set of generators contained in Eq. D1,

$$\mathcal{H}_0' = \{\sigma_x^1, \ldots, \sigma_x^n, S_z, \tau_x^1, \ldots, \tau_x^n, \tau_y^1, \ldots, \tau_y^n, V\}. \tag{D6}$$

We consider that $B > 0$ is always on, which makes controlling individual local rotations more difficult. We consider the rotating frame given by $U = \exp\{i\frac{1}{2}\mu B t S_z\}$, such that we reduce the base set of generators to

$$\mathcal{H}_0'' = \{\sigma_x^1, \ldots, \sigma_x^n, \tau_x^1, \ldots, \tau_x^n, \tau_y^1, \ldots, \tau_y^n, V\}, \tag{D7}$$

with all generators now being controllable individually from each other.

First, we consider the case in which $\Omega_{x,y}^j = \Omega_{x,y}$, such that the local operators $\tau_{x,y}^i$ are no longer individually controllable. The base set of generators becomes

$$\mathcal{H}_0 = \{\sigma_x^1, \ldots, \sigma_x^n, T_x, T_y, V\}. \tag{D8}$$

From these base generators we find the commutators

$$[\sigma_x^i, T_x] = \sum_{j=1}^n [\sigma_x^i, \tau_x^j] = -i\nu_y^i \tag{D9}$$

$$[\nu_y^i, \sigma_x^i] = -i\tau_x^i \tag{D10}$$

$$[\nu_y^i, T_y] = \sum_{j=1}^n [\nu_y^i, \tau_y^j] = -i\sigma_y^i. \tag{D11}$$

Eqs. D9 and D10 can be repeated analogously to obtain ν_x^i and τ_y^i. This means that despite the global Rabi term that determines the transition between $|1\rangle_j$ and $|R\rangle_j$ for all $1 \leq j \leq N_a$, the local generators $\tau_{x,y}^i$ are part of the dynamical Lie algebra and therefore controllable individually. From Eq. D11 we see that that σ_y^i is accessible, and therefore σ_z^i is accessible as well. This allows full access to local single-qubit operations. Note that this means the magnetic field B was not necessary for computational purposes to begin with. However, in experimental setups it serves the purpose of providing non-degenerate levels $|0\rangle$ and $|1\rangle$ at all times. Since the transformation U into the comoving frame does not affect the interaction term, i.e. $UVU^\dagger = V$, we find that the base set of generators in Eq. D8 is computationally universal on the logical space $\otimes_{j=1}^n \{|0\rangle^j, |1\rangle^j\}$. We demonstrate this by constructing specific examples of rotations around τ_x^i, τ_y, σ_y^i and σ_z^i by an arbitrary angle α as

$$e^{i\alpha\tau_x^i} = e^{i\frac{3\pi}{2}\sigma_x^i} e^{i\frac{3\pi}{2}T_x} e^{i\alpha\sigma_x^i} e^{i\frac{\pi}{2}T_x} e^{i\frac{\pi}{2}\sigma_x^i} \tag{D12}$$

$$e^{i\alpha\tau_y^i} = e^{i\frac{3\pi}{2}\sigma_x^i} e^{i\frac{3\pi}{2}T_y} e^{i\alpha\sigma_x^i} e^{i\frac{\pi}{2}T_y} e^{i\frac{\pi}{2}\sigma_x^i} \tag{D13}$$

$$e^{i\alpha\sigma_y^i} = e^{i\frac{3\pi}{2}\sigma_x^i} e^{i\frac{3\pi}{2}T_y} e^{i\frac{3\pi}{2}T_x} e^{i\alpha\sigma_x^i} e^{i\frac{\pi}{2}T_x} e^{i\frac{\pi}{2}T_y} e^{i\frac{\pi}{2}\sigma_x^i} \tag{D14}$$

$$e^{i\alpha\sigma_z^i} = e^{i\frac{7\pi}{4}\sigma_x^i} e^{i\frac{3\pi}{2}T_y} e^{i\frac{3\pi}{2}T_x} e^{i\alpha\sigma_x^i} e^{i\frac{\pi}{2}T_x} e^{i\frac{\pi}{2}T_y} e^{i\frac{\pi}{4}\sigma_x^i}. \tag{D15}$$

From these rotations, entanglement between qubits can be achieved utilizing V in the canonical manner of Rydberg architectures.

Second, we consider the case of individual Rabi couplings $\Omega_{x,y}^j(t)$, but a single global Raman coupling $A_j(t) = A(t)$. Analogously to the previous case, the base set of generators then becomes

$$\mathcal{H}_0 = \{S_x, \tau_x^1, \ldots, \tau_x^n, \tau_y^1, \ldots, \tau_y^n, V\}. \tag{D16}$$

The argument follows analogously and we find that

$$[S_x, \tau_x^j] = \sum_{i=1}^n [\sigma_x^i, \tau_x^j] = -i\nu_y^j \tag{D17}$$

$$[\nu_y^i, \tau_x^i] = i\sigma_x^i \tag{D18}$$

$$[\nu_y^i, \tau_y^i] = -i\sigma_y^i, \tag{D19}$$

and therefore arbitrary single-qubit rotations can be constructed as

$$e^{i\alpha\sigma_x^i} = e^{i\frac{3\pi}{2}\tau_x^i} e^{i\frac{3\pi}{2}S_x} e^{i\alpha\tau_x^i} e^{i\frac{\pi}{2}S_x} e^{i\frac{\pi}{2}\tau_x^i} \tag{D20}$$

$$e^{i\alpha\sigma_y^i} = e^{i\frac{3\pi}{2}\tau_x^i} e^{i\frac{3\pi}{2}S_x} e^{i\alpha\tau_y^i} e^{i\frac{\pi}{2}S_x} e^{i\frac{\pi}{2}\tau_x^i} \tag{D21}$$

$$e^{i\alpha\sigma_z^i} = e^{i\frac{3\pi}{2}\tau_x^i} e^{i\frac{3\pi}{2}S_x} e^{i\frac{\pi}{4}\tau_y^i} e^{i\alpha\tau_x^i} e^{i\frac{7\pi}{4}\tau_y^i} e^{i\frac{\pi}{2}S_x} e^{i\frac{\pi}{2}\tau_x^i}, \tag{D22}$$

despite only global control over S_x. It again follows that the dynamical Lie algebra is capable of all necessary operations for universal quantum computing. Note in particular that the generator S_z associated with the magnetic field was again not necessary for constructing arbitrary single-qubit rotations.

In this case of global Raman coupling, the Rabi couplings can be performed individually which means that there is no undesirable population in Rydberg states as overhead of unrelated transformations. The overhead transformation occurs only on the local σ_x which is easily circumvented. We can perform a CNOT gate on the 1st

and 2nd qubit while in the Rydberg blockade radius as

$$\mathrm{CNOT} = e^{i\frac{\pi}{2}\tau_x^2} R e^{i\pi\tau_x^1} R e^{i\frac{\pi}{2}\tau_x^2} \tag{D23}$$

$$R = e^{i\frac{\pi}{2}S_x} e^{i\frac{\pi}{4}\tau_x^2} e^{i\frac{\pi}{\sqrt{8}}(\tau_x^2+\tau_y^2)} e^{i\frac{\pi}{4}\tau_x^2} e^{i\frac{3\pi}{2}S_x}, \tag{D24}$$

in the presence of any amount of other qubits which by construction also experience the rotations generated by S_x. This transformation will in particular act on all these additional qubits as the identity, as desired. Finally, we note that the controlled-Y gate is implemented more naturally in this case as

$$\mathrm{C(Y)} = e^{-i\frac{\pi}{2}\tau_x^2} e^{-i\frac{\pi}{4}S_x} e^{i\pi\tau_x^1} e^{i\frac{\pi}{4}S_x} e^{i\frac{\pi}{2}\tau_x^2}. \tag{D25}$$

Here it is implied that every transformation also contains the presence of the interaction term V which is large enough to consider the system to be in the Rydberg blockade regime. We emphasize that the Rydberg blockade is not necessary for the arguments we make about universal quantum computing. We only consider the Rydberg blockade for the analytically constructed examples in Eqs. D23, D24 and D25.

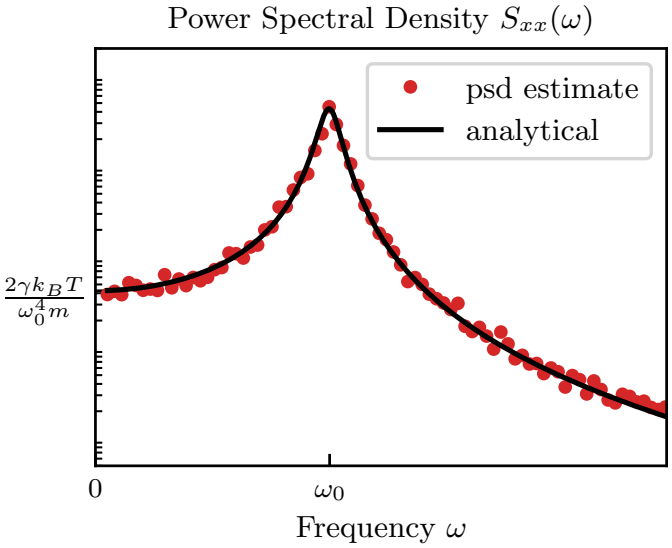


Figure 5.2: **Trapped atoms in the presence of thermal fluctuations**. The power spectral density in Eq. 5.11 (black line) and the numerical estimate of the power spectral density (red dots) are in good agreement.

5.5 Revised robustness analysis in the presence of thermal fluctuations

In this thesis, I have included the arXiv preprint [2] of the manuscript at the date of writing this thesis. In the revised version of the manuscript, we have extended the noise model and our analysis regarding the robustness of our implementations to spatial fluctuations. In this section, we will elaborate these results.

We include fluctuations of the interatomic distance between the Rydberg atoms. In an experimental realization, these fluctuations derive from the thermal motion of the atoms in the tweezer potentials, or fluctuations of the tweezer potential itself. We consider the Langevin equation

$$\partial_t v_j(t) = -\gamma_{\text{trap}} v_j(t) - \omega_0^2 x_j(t) + A_j(t). \tag{5.7}$$

$x_j(t)$ and $v_j(t)$ are the displacement from equilibrium and the velocity of the j-th atom, respectively. $\omega_0 = 2\pi \times 50\text{kHz}$ is the trap frequency, $\gamma_{\text{trap}} = 5\text{kHz}$ is a classical damping rate, and $A_j(t)$ is a fluctuating force such that

$$\langle A_i(t) A_j(t') \rangle = \frac{2\gamma k_B T}{m} \delta_{ij} \delta(t - t'), \tag{5.8}$$

for an effective temperature T, and the atomic mass $m = 171u$ of ^{171}Yb, where u is the unit atomic mass. Based on the equipartition theorem [192], we initialize each atom in

a thermal state by sampling from the zero-mean normal distributions with

$$\langle x_i^2 \rangle = \frac{k_B T}{m\omega_0^2}, \tag{5.9}$$

$$\langle v_i^2 \rangle = \frac{k_B T}{m}. \tag{5.10}$$

The power spectral density is

$$S_{xx}(\omega) = \langle x(\omega) x^*(\omega) \rangle = \frac{2\gamma k_B T}{m} \frac{1}{(\omega_0^2 - \omega^2)^2 + \gamma^2 \omega^2} \tag{5.11}$$

In Fig. 5.2 we compare the numerically obtained estimate of the power spectral density to the analytical form Eq. 5.11.

We analyze the robustness of our optimized protocols to spatial fluctuations, in the presence of non-zero dissipation, for interatomic distances of ranging from the blockade regime to the weak-coupling limit. Based on two fluctuating trajectories $x_1(t)$ and $x_2(t)$ according to Eq. 5.7, we determine the time-dependent van-der-Waals interaction strength

$$V_{\mathrm{vdW}}(t) = \frac{\hbar C_6}{|r + x_1(t) - x_2(t)|^6}, \tag{5.12}$$

where r is the interatomic distance in equilibrium at zero temperature. To further study the robustness of our implementations, we consider the following three cases:

(i) We optimize the transformation parameters in the absence of thermal fluctuations for an ensemble of random initial transformation parameters. We refer to the corresponding average infidelity as the zero-temperature infidelity $1 - F(0)$.

(ii) We calculate the infidelity for the protocols obtained in (i), but in the presence of spatial fluctuations for an effective temperature T. We refer to the corresponding average infidelity as the post-facto infidelity $1 - F_T(0)$.

(iii) We optimize the transformation parameters in the presence of thermal fluctuations for an ensemble of random initial parameters. We refer to the corresponding average infidelity as the in-situ infidelity $1 - F(T)$.

Specifically, in (i) and (iii), we generate $N_\theta = 100$ optimized protocols for an effective temperature of T, corresponding to the set of transformation parameters $\theta_i(T)$ optimized from different initial values $\theta_i^0(T)$, for each value of the interatomic distance r, over 2000 training epochs.

In Fig. 5.3 we show the resulting infidelities $1 - F(0)$, $1 - F_T(0)$ and $1 - F_T(0)$. The zero-temperature protocols are highly susceptible to thermal fluctuation post-facto, for interatomic distances $r \approx r_B$, rendering the intermediate regime unfavorable. However, we find that the in-situ optimized protocols are less susceptible to spatial fluctuations in all considered dynamical regimes. The robustness with respect to spatial fluctuations is one of the key features that makes the Rydberg blockade regime attractive for quantum

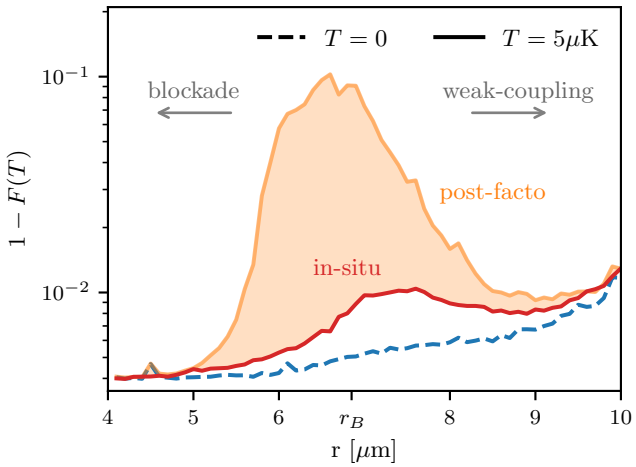


Figure 5.3: **Robustness against thermal fluctuations.** The average infidelity $1 - F$, as a function of the interatomic distance r, for zero temperature (dashed), an effective temperature of $T = 5\mu$K (solid lines), and a dissipation rate of $1/\gamma = 100\mu$s. The zero-temperature infidelity $1-F(0)$ is depicted in blue, the post-facto infidelity $1 - F_T(0)$ is depicted in orange and the in-situ infidelity $1 - F(T)$ is depicted in red. The zero-temperature protocols are highly susceptible to thermal fluctuation post-facto in the intermediate regime. The in-situ optimized protocols are less susceptible to spatial fluctuations for all considered interatomic distances.

computing purposes. However, we emphasize that for the case of in-situ optimization the system is robust against spatial fluctuations in the intermediate regime and the weak-coupling limit as well. We also note that the intermediate regime and weak-coupling limit enable gate implementations in tweezer arrays with strongly suppressed next-nearest interactions, resulting in more straight-forward implementations.

6 Conclusion

In this cumulative dissertation, I have presented my work on numerically optimized control in the context of gravitational-wave astronomy, pulse engineering for quantum computers, and quantum gate optimization for Rydberg-atom based quantum computing. During my research, I developed time-series forecasting models based on artificial neural networks, and numerical quantum computing models for the design of optimized quantum gates in the presence of experimentally motivated constraints, noise and decoherence. These numerical platforms may serve as a basis for future research on these topics.

In publication [1], we studied machine learning-assisted motion prediction in a proof-of-principle model inspired by pendulum suspension systems in gravitational wave observatories. This approach includes the utilization of witnessed seismic noise that affects the displacement of the pendulum suspended mass. Our model provides a flexible way to include witness sensors allowing for the utilization of large seismic sensing networks. We have found that our trained neural network learned both the natural behavior of the pendulum and the transfer function from the witnessed seismic noise to the displacement of the pendulum suspended mass. Additionally, our model captures instrumentation artifacts such as non-linearities of the used lateral effect sensor, which linear transfer functions are incapable of. Finally, our work supports the idea that machine learning-based corrective forward stabilization is a promising approach for next generation gravitational-wave detectors.

In publication [2] and publication [3], we implemented numerical hybrid quantum-classical platforms for the purpose of quantum gate optimization in the presence of experimentally motivated constraints, noise and decoherence. The underlying parameterizations of the protocols are adjusted via gradient-based optimal control techniques. We include dissipation via the Lindblad master equation and learn optimized implementations in an online learning approach based on the dynamical evolutions of randomly prepared states. In publication [2], we utilized GRAPE combined with mini-batch learning and an adaptive learning rate according to the ADAM method. Thus, gradients are approximated via finite differences in the parameter space. By utilizing both global and local coherent control fields, we demonstrate optimized implementations of the CNOT gate for Rydberg-atom based quantum computing platforms. Furthermore, we studied the susceptibility of our implementations to spatial fluctuations of the trapped atoms. We find that in-situ optimization leads to protocols that are more robust against such fluctuations, not only in the Blockade regime, but also ranging into the weak-coupling limit. This potentially enables gate implementations with strongly suppressed next-nearest neighbour interactions, resulting in more straight-forward implementations. In publication [3], we extended gradient-based quantum optimal control to the framework

of linear response theory, enabling gradient estimation through the linear response of the system to time-local perturbations of the control fields. This approach gives rise to finite differences in the space spanned by control operators of the control fields. This has motivated us to refer to our method as pulse engineering via projection of response functions (PEPR). We provide a multi-parameter update rule for the trainable parameters of the perturbed control field based on only a single dynamical evaluation, and in an ϵ-free manner. Thus, the numerical complexity to adjust the parameterization does not depend its the expressibility. Furthermore, we have implemented GRAPE as a comparative benchmark for the example of the optimization of the CNOT gate. Here, we found that PEPR is consistently more efficient in a direct comparison to GRAPE, as it produces high-fidelity implementations more reliably. Our results suggest a reduced demand in online utilization, which benefits hybrid quantum-classical approaches directly. We also demonstrated PEPR in the presence of experimentally motivated constraints and dissipation. Finally, with our results [2, 3] we have demonstrated the immediate and significant impact of optimized control for quantum gate design toward fault-tolerant quantum computing.

Acknowledgements

I would like to thank my esteemed colleagues and friends at the Institute for Quantum Physics for the wonderful and unforgettable time we spent together. Also, I want to thank my collaborators for the great research projects we have carried out together.

This work would not have been possible without the unconditional support from my Family, my dear partner Aylin and my daughter Nuri.

Bibliography

[1] N. Heimann, J. Petermann, D. Hartwig, R. Schnabel, and L. Mathey, "Predicting the motion of a high-q pendulum subject to seismic perturbations using machine learning", *Applied Physics Letters*, vol. 122, no. 25, p. 254 101, 2023, ISSN: 0003-6951. DOI: 10.1063/5.0144593.

[2] N. Heimann *et al.*, *Quantum gate optimization for rydberg architectures in the weak-coupling limit*, arXiv: 2306.08691, 2023.

[3] N. Heimann, L. Broers, and L. Mathey, "Pulse engineering via projection of response functions", *Phys. Rev. Res.*, vol. 7, no. 1, p. 013 101, 2025. DOI: 10.1103/PhysRevResearch.7.013101.

[4] Flanagan and S. A. Hughes, "The basics of gravitational wave theory", *New Journal of Physics*, vol. 7, no. 1, p. 204, 2005. DOI: 10.1088/1367-2630/7/1/204.

[5] B. P. Abbott *et al.*, "LIGO: The Laser Interferometer Gravitational-Wave Observatory", *Reports on Progress in Physics*, vol. 72, no. 7, p. 076 901, 2009, ISSN: 0034-4885, 1361-6633. DOI: 10.1088/0034-4885/72/7/076901.

[6] G. M. Harry and (forthe LIGO Scientific Collaboration), "Advanced ligo: The next generation of gravitational wave detectors", *Classical and Quantum Gravity*, vol. 27, no. 8, p. 084 006, 2010. DOI: 10.1088/0264-9381/27/8/084006.

[7] T. Accadia *et al.*, "Virgo: A laser interferometer to detect gravitational waves", *Journal of Instrumentation*, vol. 7, no. 03, P03012, 2012. DOI: 10.1088/1748-0221/7/03/P03012.

[8] F Acernese *et al.*, "Advanced virgo: A second-generation interferometric gravitational wave detector", *Classical and Quantum Gravity*, vol. 32, no. 2, p. 024 001, 2014. DOI: 10.1088/0264-9381/32/2/024001.

[9] B Willke *et al.*, "The geo 600 gravitational wave detector", *Classical and Quantum Gravity*, vol. 19, no. 7, p. 1377, 2002. DOI: 10.1088/0264-9381/19/7/321.

[10] K. L. Dooley and (for the LIGO Scientific Collaboration), "Status of geo 600", *Journal of Physics: Conference Series*, vol. 610, no. 1, p. 012 015, 2015. DOI: 10.1088/1742-6596/610/1/012015.

[11] C. J. Moore, R. H. Cole, and C. P. L. Berry, "Gravitational-wave sensitivity curves", *Classical and Quantum Gravity*, vol. 32, no. 1, p. 015 014, 2014. DOI: 10.1088/0264-9381/32/1/015014.

[12] P Nguyen *et al.*, "Environmental noise in advanced ligo detectors", *Classical and Quantum Gravity*, vol. 38, no. 14, p. 145 001, 2021. DOI: 10.1088/1361-6382/ac011a.

[13] F Matichard *et al.*, "Seismic isolation of advanced ligo: Review of strategy, instrumentation and performance", *Classical and Quantum Gravity*, vol. 32, no. 18, p. 185 003, 2015. DOI: 10.1088/0264-9381/32/18/185003.

[14] S. M. Aston *et al.*, "Update on quadruple suspension design for Advanced LIGO", *Classical and Quantum Gravity*, vol. 29, no. 23, p. 235 004, 2012, ISSN: 0264-9381, 1361-6382. DOI: 10.1088/0264-9381/29/23/235004.

[15] R. DeRosa *et al.*, "Global feed-forward vibration isolation in a km scale interferometer", *Classical and Quantum Gravity*, vol. 29, no. 21, p. 215 008, 2012, ISSN: 0264-9381, 1361-6382. DOI: 10.1088/0264-9381/29/21/215008.

[16] R. P. Feynman, "Simulating physics with computers", *International Journal of Theoretical Physics*, vol. 21, no. 6, pp. 467–488, 1982. DOI: 10.1007/BF02650179.

[17] D. E. Deutsch and R. Penrose, "Quantum computational networks", *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, vol. 425, no. 1868, pp. 73–90, 1989. DOI: 10.1098/rspa.1989.0099.

[18] K. Bharti *et al.*, "Noisy intermediate-scale quantum (NISQ) algorithms", *Reviews of Modern Physics*, vol. 94, no. 1, p. 015 004, 2022, ISSN: 0034-6861, 1539-0756. DOI: 10.1103/RevModPhys.94.015004.

[19] M. Cerezo *et al.*, "Variational quantum algorithms", *Nature Reviews Physics*, vol. 3, no. 9, pp. 625–644, 2021. DOI: 10.1038/s42254-021-00348-9.

[20] C. P. Koch *et al.*, "Quantum optimal control in quantum technologies. strategic report on current status, visions and goals for research in europe", *EPJ Quantum Technology*, vol. 9, no. 1, 2022.

[21] V. Ramakrishna and H. Rabitz, "Relation between quantum computing and quantum controllability", *Phys. Rev. A*, vol. 54, no. 2, pp. 1715–1716, 1996. DOI: 10.1103/PhysRevA.54.1715.

[22] A. B. Magann *et al.*, "From pulses to circuits and back again: A quantum optimal control perspective on variational quantum algorithms", *PRX Quantum*, vol. 2, no. 1, p. 010 101, 2021. DOI: 10.1103/PRXQuantum.2.010101.

[23] A. Castro and I. V. Tokatly, "Quantum optimal control theory in the linear response formalism", *Phys. Rev. A*, vol. 84, no. 3, p. 033 410, 2011. DOI: 10.1103/PhysRevA.84.033410.

[24] M. Saffman and T. G. Walker, "Analysis of a quantum logic device based on dipole-dipole interactions of optically trapped Rydberg atoms", *Physical Review A*, vol. 72, no. 2, p. 022 347, 2005, ISSN: 1050-2947, 1094-1622. DOI: 10.1103/PhysRevA.72.022347.

[25] M. Saffman, T. G. Walker, and K. Mølmer, "Quantum information with rydberg atoms", *Rev. Mod. Phys.*, vol. 82, no. 3, pp. 2313–2363, 2010. DOI: 10.1103/RevModPhys.82.2313.

[26] M. Saffman, "Quantum computing with atomic qubits and Rydberg interactions: Progress and challenges", *Journal of Physics B: Atomic, Molecular and Optical Physics*, vol. 49, no. 20, p. 202 001, 2016, ISSN: 0953-4075, 1361-6455. DOI: 10.1088/0953-4075/49/20/202001.

[27] M. Morgado and S. Whitlock, "Quantum simulation and computing with rydberg-interacting qubits", *AVS Quantum Science*, vol. 3, no. 2, p. 023 501, 2021. DOI: 10.1116/5.0036562.

[28] A. Browaeys and T. Lahaye, "Many-body physics with individually controlled Rydberg atoms", *Nature Physics*, vol. 16, no. 2, pp. 132–142, 2020, ISSN: 1745-2473, 1745-2481. DOI: 10.1038/s41567-019-0733-z.

[29] P. Scholl *et al.*, "Quantum simulation of 2d antiferromagnets with hundreds of rydberg atoms", *Nature*, vol. 595, no. 7866, pp. 233–238, 2021. DOI: 10.1038/s41586-021-03585-1.

[30] S. Ebadi *et al.*, "Quantum phases of matter on a 256-atom programmable quantum simulator", *Nature*, vol. 595, no. 7866, pp. 227–232, 2021. DOI: 10.1038/s41586-021-03582-4.

[31] T. F. Gallagher, "Rydberg atoms", *Reports on Progress in Physics*, vol. 51, no. 2, p. 143, 1988. DOI: 10.1088/0034-4885/51/2/001.

[32] X. Wu *et al.*, "A concise review of rydberg atom based quantum computation and quantum simulation*", *Chinese Physics B*, vol. 30, no. 2, p. 020 305, 2021. DOI: 10.1088/1674-1056/abd76f.

[33] E. Urban *et al.*, "Observation of Rydberg blockade between two atoms", *Nature Physics*, vol. 5, no. 2, pp. 110–114, 2009, ISSN: 1745-2473, 1745-2481. DOI: 10.1038/nphys1178.

[34] T. Wilk *et al.*, "Entanglement of Two Individual Neutral Atoms Using Rydberg Blockade", *Physical Review Letters*, vol. 104, no. 1, p. 010 502, 2010, ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.104.010502.

[35] D. Jaksch, J. I. Cirac, P. Zoller, S. L. Rolston, R. Côté, and M. D. Lukin, "Fast quantum gates for neutral atoms", *Phys. Rev. Lett.*, vol. 85, no. 10, pp. 2208–2211, 2000. DOI: 10.1103/PhysRevLett.85.2208.

[36] X. L. Zhang, L. Isenhower, A. T. Gill, T. G. Walker, and M. Saffman, "Deterministic entanglement of two neutral atoms via rydberg blockade", *Phys. Rev. A*, vol. 82, no. 3, p. 030 306, 2010. DOI: 10.1103/PhysRevA.82.030306.

[37] L. Isenhower *et al.*, "Demonstration of a neutral atom controlled-not quantum gate", *Phys. Rev. Lett.*, vol. 104, no. 1, p. 010 503, 2010. DOI: 10.1103/PhysRevLett.104.010503.

[38] K. M. Maller *et al.*, "Rydberg-blockade controlled-not gate and entanglement in a two-dimensional array of neutral-atom qubits", *Phys. Rev. A*, vol. 92, no. 2, p. 022 336, 2015. DOI: 10.1103/PhysRevA.92.022336.

[39] Y. Zeng *et al.*, "Entangling two individual atoms of different isotopes via rydberg blockade", *Phys. Rev. Lett.*, vol. 119, no. 16, p. 160 502, 2017. DOI: 10.1103/PhysRevLett.119.160502.

[40] H. Levine *et al.*, "Parallel implementation of high-fidelity multiqubit gates with neutral atoms", *Phys. Rev. Lett.*, vol. 123, no. 17, p. 170 503, 2019. DOI: 10.1103/PhysRevLett.123.170503.

[41] N. Chen *et al.*, "Analyzing the rydberg-based optical-metastable-ground architecture for ^{171}Yb nuclear spins", *Phys. Rev. A*, vol. 105, no. 5, p. 052 438, 2022. DOI: 10.1103/PhysRevA.105.052438.

[42] A. Jenkins, J. W. Lis, A. Senoo, W. F. McGrew, and A. M. Kaufman, "Ytterbium nuclear-spin qubits in an optical tweezer array", *Phys. Rev. X*, vol. 12, no. 2, p. 021 027, 2022. DOI: 10.1103/PhysRevX.12.021027.

[43] J. T. Wilson *et al.*, "Trapping alkaline earth rydberg atoms optical tweezer arrays", *Phys. Rev. Lett.*, vol. 128, no. 3, p. 033 201, 2022. DOI: 10.1103/PhysRevLett.128.033201.

[44] A. Fox, *Quantum Optics: An Introduction* (Oxford Master Series in Physics). OUP Oxford, 2006, ISBN: 9780198566724.

[45] S. Chu, "Atom interferometry", in *Coherent atomic matter waves*, R. Kaiser, C. Westbrook, and F. David, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 317–370, ISBN: 978-3-540-45338-3.

[46] C. Brif, B. P. Ruzic, and G. W. Biedermann, "Characterization of errors in interferometry with entangled atoms", *PRX Quantum*, vol. 1, p. 010 306, 1 2020. DOI: 10.1103/PRXQuantum.1.010306.

[47] A. Einstein, "Die grundlage der allgemeinen relativitätstheorie", *Annalen der Physik*, vol. 354, no. 7, pp. 769–822, 1916. DOI: https://doi.org/10.1002/andp.19163540702.

[48] A. Einstein, "Die Feldgleichungen der Gravitation", *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, pp. 844–847, 1915.

[49] A. Einstein, "Über Gravitationswellen", *Sitzungsberichte der Kñiglich Preussischen Akademie der Wissenschaften*, pp. 154–167, 1918.

[50] R. A. Hulse and J. H. Taylor, "Discovery of a pulsar in a binary system.", *Astrophysical Journal*, vol. 195, pp. L51–L53, 1975. DOI: 10.1086/181708.

[51] J. M. Weisberg, J. H. Taylor, and L. A. Fowler, "Gravitational waves from an orbiting pulsar", *Scientific American*, vol. 245, pp. 74–82, 1981. DOI: 10.1038/scientificamerican1081-74.

[52] J. M. Weisberg and J. H. Taylor, "The Relativistic Binary Pulsar B1913+16", in *Radio Pulsars*, M. Bailes, D. J. Nice, and S. E. Thorsett, Eds., ser. Astronomical Society of the Pacific Conference Series, vol. 302, 2003, p. 93. DOI: 10.48550/arXiv.astro-ph/0211217.

[53] R. A. Hulse and J. H. Taylor Jr., *Nobel prize in physics 1993*, https://www.nobelprize.org/prizes/physics/1993/summary/ [Zugriff am: 2025-07-08], 1993.

[54] G. Huang *et al.*, "A review of optical interferometry for high-precision length measurement", *Micromachines*, vol. 16, no. 1, 2025, ISSN: 2072-666X.

[55] B. P. Abbott *et al.*, "Observation of gravitational waves from a binary black hole merger", *Phys. Rev. Lett.*, vol. 116, no. 6, p. 061 102, 2016. DOI: 10.1103/PhysRevLett.116.061102.

[56] R. Weiss, B. C. Barish, and S. T. Kip, *Nobel prize in physics 2017*, https://www.nobelprize.org/prizes/physics/2017/summary/ [Zugriff am: 2025-07-08], 2017.

[57] B. P. Abbott *et al.*, "Gwtc-1: A gravitational-wave transient catalog of compact binary mergers observed by ligo and virgo during the first and second observing runs", *Phys. Rev. X*, vol. 9, no. 3, p. 031 040, 2019. DOI: 10.1103/PhysRevX.9.031040.

[58] R. Abbott *et al.*, "Gwtc-2: Compact binary coalescences observed by ligo and virgo during the first half of the third observing run", *Phys. Rev. X*, vol. 11, no. 2, p. 021 053, 2021. DOI: 10.1103/PhysRevX.11.021053.

[59] F. Frasconi and P. Rapagnani, "An introduction to the virgo suspension system", in *Advanced Interferometers and the Search for Gravitational Waves: Lectures from the First VESF School on Advanced Detectors for Gravitational Waves*. Cham: Springer International Publishing, 2014, pp. 193–223, ISBN: 978-3-319-03792-9. DOI: 10.1007/978-3-319-03792-9_7.

[60] T. L. S. Collaboration *et al.*, "Advanced ligo", *Classical and Quantum Gravity*, vol. 32, no. 7, p. 074 001, 2015. DOI: 10.1088/0264-9381/32/7/074001.

[61] M. Tse *et al.*, "Quantum-enhanced advanced ligo detectors in the era of gravitational-wave astronomy", *Phys. Rev. Lett.*, vol. 123, no. 23, p. 231 107, 2019. DOI: 10.1103/PhysRevLett.123.231107.

[62] S. Dwyer, D. Sigg, S. W. Ballmer, L. Barsotti, N. Mavalvala, and M. Evans, "Gravitational wave detector with cosmological reach", *Phys. Rev. D*, vol. 91, no. 8, p. 082 001, 2015. DOI: 10.1103/PhysRevD.91.082001.

[63] B. P. Abbott *et al.*, "Exploring the sensitivity of next generation gravitational wave detectors", *Classical and Quantum Gravity*, vol. 34, no. 4, p. 044 001, 2017. DOI: 10.1088/1361-6382/aa51f4.

[64] M. Branchesi *et al.*, "Science with the einstein telescope: A comparison of different designs", *Journal of Cosmology and Astroparticle Physics*, vol. 2023, no. 07, p. 068, 2023. DOI: 10.1088/1475-7516/2023/07/068.

[65] C. M. Caves, "Quantum-mechanical radiation-pressure fluctuations in an interferometer", *Phys. Rev. Lett.*, vol. 45, no. 2, pp. 75–79, 1980. DOI: 10.1103/PhysRevLett.45.75.

[66] J. Abadie *et al.*, "A gravitational wave observatory operating beyond the quantum shot-noise limit", *Nature Physics*, vol. 7, no. 12, pp. 962–965, 2011. DOI: 10.1038/nphys2083.

[67] D. McClelland, N. Mavalvala, Y. Chen, and R. Schnabel, "Advanced interferometry, quantum optics and optomechanics in gravitational wave detectors", *Laser & Photonics Reviews*, vol. 5, no. 5, pp. 677–696, 2011. DOI: https://doi.org/10.1002/lpor.201000034.

[68] J. Harms, "Terrestrial gravity fluctuations", *Living Reviews in Relativity*, vol. 22, no. 1, p. 6, 2019, ISSN: 2367-3613, 1433-8351. DOI: 10.1007/s41114-019-0022-2.

[69] J. C. Driggers, J. Harms, and R. X. Adhikari, "Subtraction of Newtonian noise using optimized sensor arrays", *Physical Review D*, vol. 86, no. 10, p. 102 001, 2012, ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.86.102001.

[70] V. Dattilo, "The virgo suspensions: Design and recent performance measurements", *Physics Letters A*, vol. 318, no. 3, pp. 192–198, 2003, ISSN: 0375-9601. DOI: https://doi.org/10.1016/j.physleta.2003.07.012.

[71] Q. Li *et al.*, "Measurements of the gravitational constant using two independent methods", *Nature*, vol. 560, no. 7720, pp. 582–588, 2018. DOI: 10.1038/s41586-018-0431-5.

[72] N. A. Robertson *et al.*, "Quadruple suspension design for advanced ligo", *Classical and Quantum Gravity*, vol. 19, no. 15, p. 4043, 2002. DOI: 10.1088/0264-9381/19/15/311.

[73] G. González, "Suspensions thermal noise in the ligo gravitational wave detector", *Classical and Quantum Gravity*, vol. 17, no. 21, p. 4409, 2000. DOI: 10.1088/0264-9381/17/21/305.

[74] K. J. Åström and R. M. Murray, *Feedback Systems*. Princeton: Princeton University Press, 2009, ISBN: 978-0-691-13576-2.

[75] K. A. Strain and B. N. Shapiro, "Damping and local control of mirror suspensions for laser interferometric gravitational wave detectors", *Review of Scientific Instruments*, vol. 83, no. 4, p. 044 501, 2012, ISSN: 0034-6748, 1089-7623. DOI: 10.1063/1.4704459.

[76] J. C. Driggers, M. Evans, K. Pepper, and R. Adhikari, "Active noise cancellation in a suspended interferometer", *Review of Scientific Instruments*, vol. 83, no. 2, p. 024 501, 2012, ISSN: 0034-6748, 1089-7623. DOI: 10.1063/1.3675891.

[77] C Di Fronzo *et al.*, "A control strategy for seismic noise reduction on advanced ligo gravitational-wave detector", *Classical and Quantum Gravity*, vol. 42, no. 4, p. 045 019, 2025. DOI: 10.1088/1361-6382/adab5f.

[78] F Badaracco *et al.*, "Machine learning for gravitational-wave detection: Surrogate wiener filtering for the prediction and optimized cancellation of newtonian noise at virgo", *Classical and Quantum Gravity*, vol. 37, no. 19, p. 195 016, 2020. DOI: 10.1088/1361-6382/abab64.

[79] E. Cuoco *et al.*, "Enhancing gravitational-wave science with machine learning", *Machine Learning: Science and Technology*, vol. 2, no. 1, p. 011 002, 2020. DOI: 10.1088/2632-2153/abb93a.

[80] R. Ormiston, T. Nguyen, M. Coughlin, R. X. Adhikari, and E. Katsavounidis, "Noise reduction in gravitational-wave data via deep learning", *Phys. Rev. Research*, vol. 2, no. 3, p. 033 066, 2020. DOI: 10.1103/PhysRevResearch.2.033066.

[81] G. Vajente *et al.*, "Machine-learning nonstationary noise out of gravitational-wave detectors", *Phys. Rev. D*, vol. 101, no. 4, p. 042 003, 2020. DOI: 10.1103/PhysRevD.101.042003.

[82] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications*. The MIT Press, 1949, ISBN: 9780262257190. DOI: 10.7551/mitpress/2946.001.0001.

[83] M. Razzano and E. Cuoco, "Image-based deep learning for classification of noise transients in gravitational wave detectors", *Classical and Quantum Gravity*, vol. 35, no. 9, p. 095 016, 2018, ISSN: 0264-9381, 1361-6382. DOI: 10.1088/1361-6382/aab793.

[84] P. T. Baker, S. Caudill, K. A. Hodge, D. Talukder, C. Capano, and N. J. Cornish, "Multivariate classification with random forests for gravitational wave searches of black hole binary coalescence", *Physical Review D*, vol. 91, no. 6, p. 062 004, 2015, ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.91.062004.

[85] P. Graff, F. Feroz, M. P. Hobson, and A. Lasenby, "BAMBI: Blind accelerated multimodal Bayesian inference: BAMBI", *Monthly Notices of the Royal Astronomical Society*, vol. 421, pp. 169–180, 2012, ISSN: 00358711. DOI: 10.1111/j.1365-2966.2011.20288.x.

[86] Z. Doctor, B. Farr, D. E. Holz, and M. Pürrer, "Statistical gravitational waveform models: What to simulate next?", *Physical Review D*, vol. 96, no. 12, p. 123 011, 2017, ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.96.123011.

[87] D. Hartwig, "Mitigating anthropogenic seismic noise for precision experiments in urban environments", https://ediss.sub.uni-hamburg.de/handle/ediss/11439, Ph.D. dissertation, University of Hamburg, 2025.

[88] J. Petermann, "Suspended mirror pendulum with radiation pressure compensated laser-interferometric position readout", https://ediss.sub.uni-hamburg.de/handle/ediss/9800, Ph.D. dissertation, University of Hamburg, 2022.

[89] J. Shi, M. Jain, and G. Narasimhan, "Time series forecasting using various deep learning models", *International Journal of Computer and Systems Engineering*, vol. 16, no. 6, pp. 224–232, 2022, ISSN: eISSN: 1307-6892.

[90] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[91] B. D. Ripley, *Pattern Recognition and Neural Networks*. Cambridge University Press, 1996.

[92] F. Sohil, M. U. Sohali, and J. Shabbir, "An introduction to statistical learning with applications in r", *Statistical Theory and Related Fields*, vol. 6, no. 1, pp. 87–87, 2, 2022, ISSN: 2475-4269, 2475-4277. DOI: 10.1080/24754269.2021.1980261.

[93] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, arXiv: 1412.6980, 2017. DOI: https://doi.org/10.48550/arXiv.1412.6980.

[94] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors", *Nature*, vol. 323, no. 6088, pp. 533–536, 1986. DOI: 10.1038/323533a0.

[95] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators", *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989, ISSN: 0893-6080. DOI: https://doi.org/10.1016/0893-6080(89)90020-8.

[96] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series", in *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA, USA: MIT Press, 1998, 255–258, ISBN: 0262511029.

[97] X. Ying, "An overview of overfitting and its solutions", *Journal of Physics: Conference Series*, vol. 1168, no. 2, p. 022 022, 2019. DOI: 10.1088/1742-6596/1168/2/022022.

[98] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting", *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014.

[99] M. Abadi *et al.*, *TensorFlow: Large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015.

[100] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications: With R Examples* (Springer Texts in Statistics). Cham: Springer International Publishing, 2017, ISBN: 978-3-319-52451-1 978-3-319-52452-8. DOI: 10.1007/978-3-319-52452-8.

[101] B. Lim and S. Zohren, "Time-series forecasting with deep learning: A survey", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, no. 2194, p. 20 200 209, 2021. DOI: 10.1098/rsta.2020.0209.

[102] D. M. Durairaj and B. H. K. Mohan, "A convolutional neural network based approach to financial time series prediction", *Neural Computing and Applications*, vol. 34, no. 16, pp. 13 319–13 337, 2022, ISSN: 0941-0643, 1433-3058. DOI: 10.1007/s00521-022-07143-2.

[103] M. Markova, "Convolutional neural networks for forex time series forecasting", *AIP Conference Proceedings*, vol. 2459, no. 1, p. 030 024, 2022, ISSN: 0094-243X. DOI: 10.1063/5.0083533.

[104] R. P. Feynman, "Quantum mechanical computers", *Foundations of Physics*, vol. 16, no. 6, pp. 507–531, 1986. DOI: 10.1007/BF01886518.

[105] D. Deutsch and R. Penrose, "Quantum theory, the church–turing principle and the universal quantum computer", *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, vol. 400, no. 1818, pp. 97–117, 1985. DOI: 10.1098/rspa.1985.0070.

[106] D. Deutsch and R. Jozsa, "Rapid solution of problems by quantum computation", *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, vol. 439, no. 1907, pp. 553–558, 1992. DOI: 10.1098/rspa.1992.0167.

[107] P. Shor, "Algorithms for quantum computation: Discrete logarithms and factoring", in *Proceedings 35th Annual Symposium on Foundations of Computer Science*, 1994, pp. 124–134. DOI: 10.1109/SFCS.1994.365700.

[108] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer", *SIAM Journal on Computing*, vol. 26, no. 5, pp. 1484–1509, 1997. DOI: 10.1137/S0097539795293172.

[109] L. K. Grover, "A fast quantum mechanical algorithm for database search", in *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, ser. STOC '96, Philadelphia, Pennsylvania, USA: Association for Computing Machinery, 1996, 212–219, ISBN: 0897917855. DOI: 10.1145/237814.237866.

[110] D. P. DiVincenzo, "Quantum computation", *Science*, vol. 270, no. 5234, pp. 255–261, 1995. DOI: 10.1126/science.270.5234.255.

[111] J. I. Cirac and P. Zoller, "Quantum computations with cold trapped ions", *Phys. Rev. Lett.*, vol. 74, no. 20, pp. 4091–4094, 1995. DOI: 10.1103/PhysRevLett.74.4091.

[112] C. Monroe, D. M. Meekhof, B. E. King, W. M. Itano, and D. J. Wineland, "Demonstration of a fundamental quantum logic gate", *Phys. Rev. Lett.*, vol. 75, no. 25, pp. 4714–4717, 1995. DOI: 10.1103/PhysRevLett.75.4714.

[113] Q. A. Turchette *et al.*, "Deterministic entanglement of two trapped ions", *Phys. Rev. Lett.*, vol. 81, no. 17, pp. 3631–3634, 1998. DOI: 10.1103/PhysRevLett.81.3631.

[114] F. Schmidt-Kaler *et al.*, "Realization of the cirac–zoller controlled-not quantum gate", *Nature*, vol. 422, no. 6930, pp. 408–411, 2003. DOI: 10.1038/nature01494.

[115] D. Jaksch, H.-J. Briegel, J. I. Cirac, C. W. Gardiner, and P. Zoller, "Entanglement of atoms via cold controlled collisions", *Phys. Rev. Lett.*, vol. 82, no. 9, pp. 1975–1978, 1999. DOI: 10.1103/PhysRevLett.82.1975.

[116] M. Kjaergaard *et al.*, "Superconducting qubits: Current state of play", *Annual Review of Condensed Matter Physics*, vol. 11, no. Volume 11, 2020, pp. 369–395, 2020, ISSN: 1947-5462. DOI: https://doi.org/10.1146/annurev-conmatphys-031119-050605.

[117] E. Knill, R. Laflamme, and G. J. Milburn, "A scheme for efficient quantum computation with linear optics", *Nature*, vol. 409, no. 6816, pp. 46–52, 2001. DOI: 10.1038/35051009.

[118] P. Kok, W. J. Munro, K. Nemoto, T. C. Ralph, J. P. Dowling, and G. J. Milburn, "Linear optical quantum computing with photonic qubits", *Rev. Mod. Phys.*, vol. 79, no. 1, pp. 135–174, 2007. DOI: 10.1103/RevModPhys.79.135.

[119] N. A. Gershenfeld and I. L. Chuang, "Bulk spin-resonance quantum computation", *Science*, vol. 275, no. 5298, pp. 350–356, 1997. DOI: 10.1126/science.275.5298.350.

[120] D. G. Cory, M. D. Price, and T. F. Havel, "Nuclear magneticgr resonance spectroscopy: An experimentally accessible paradigm for quantum computing", *Physica D: Nonlinear Phenomena*, vol. 120, no. 1, pp. 82–101, 1998, Proceedings of the Fourth Workshop on Physics and Consumption, ISSN: 0167-2789. DOI: https://doi.org/10.1016/S0167-2789(98)00046-3.

[121] P. W. Shor, "Scheme for reducing decoherence in quantum computer memory", *Phys. Rev. A*, vol. 52, no. 4, R2493–R2496, 1995. DOI: 10.1103/PhysRevA.52.R2493.

[122] A. M. Steane, "Error correcting codes in quantum theory", *Phys. Rev. Lett.*, vol. 77, no. 5, pp. 793–797, 1996. DOI: 10.1103/PhysRevLett.77.793.

[123] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, "Surface codes: Towards practical large-scale quantum computation", *Phys. Rev. A*, vol. 86, no. 3, p. 032 324, 2012. DOI: 10.1103/PhysRevA.86.032324.

[124] B. M. Terhal, "Quantum error correction for quantum memories", *Rev. Mod. Phys.*, vol. 87, no. 2, pp. 307–346, 2015. DOI: 10.1103/RevModPhys.87.307.

[125] R. Acharya *et al.*, "Quantum error correction below the surface code threshold", *Nature*, vol. 638, no. 8052, pp. 920–926, 2025. DOI: 10.1038/s41586-024-08449-y.

[126] J. Preskill, "Quantum Computing in the NISQ era and beyond", *Quantum*, vol. 2, p. 79, 2018, ISSN: 2521-327X. DOI: 10.22331/q-2018-08-06-79.

[127] D. Bluvstein *et al.*, "Logical quantum processor based on reconfigurable atom arrays", *Nature*, vol. 626, no. 7997, pp. 58–65, 2024. DOI: 10.1038/s41586-023-06927-3.

[128] D. Gao *et al.*, "Establishing a new benchmark in quantum computational advantage with 105-qubit zuchongzhi 3.0 processor", *Phys. Rev. Lett.*, vol. 134, no. 9, p. 090 601, 2025. DOI: 10.1103/PhysRevLett.134.090601.

[129] Z. C. Yang, A. Rahmani, A. Shabani, H. Neven, and C. Chamon, "Optimizing variational quantum algorithms using pontryagin's minimum principle", *Phys. Rev. X*, vol. 7, no. 2, p. 021 027, 2017. DOI: 10.1103/PhysRevX.7.021027.

[130] A. Choquette, A. Di Paolo, P. K. Barkoutsos, D. Sénéchal, I. Tavernelli, and A. Blais, "Quantum-optimal-control-inspired ansatz for variational quantum algorithms", *Phys. Rev. Res.*, vol. 3, no. 2, p. 023 092, 2021. DOI: 10.1103/PhysRevResearch.3.023092.

[131] D. Lu *et al.*, "Enhancing quantum control by bootstrapping a quantum processor of 12 qubits", *npj Quantum Information*, vol. 3, no. 1, p. 45, 2017. DOI: 10.1038/s41534-017-0045-z.

[132] C. Brif, R. Chakrabarti, and H. Rabitz, "Control of quantum phenomena: Past, present and future", *New Journal of Physics*, vol. 12, no. 7, p. 075 008, 2010. DOI: 10.1088/1367-2630/12/7/075008.

[133] M. Kang *et al.*, "Batch optimization of frequency-modulated pulses for robust two-qubit gates in ion chains", *Phys. Rev. Appl.*, vol. 16, no. 2, p. 024 039, 2021. DOI: 10.1103/PhysRevApplied.16.024039.

[134] C. Figgatt *et al.*, "Parallel entangling operations on a universal ion-trap quantum computer", *Nature*, vol. 572, no. 7769, pp. 368–372, 2019, ISSN: 1476-4687. DOI: 10.1038/s41586-019-1427-5.

[135] T. Choi *et al.*, "Optimal quantum control of multimode couplings between trapped ion qubits for scalable entanglement", *Phys. Rev. Lett.*, vol. 112, no. 19, p. 190 502, 2014. DOI: 10.1103/PhysRevLett.112.190502.

[136] V. Nebendahl, H. Häffner, and C. F. Roos, "Optimal control of entangling operations for trapped-ion quantum computing", *Phys. Rev. A*, vol. 79, no. 1, p. 012 312, 2009. DOI: 10.1103/PhysRevA.79.012312.

[137] M. Werninghaus, D. J. Egger, F. Roy, S. Machnes, F. K. Wilhelm, and S. Filipp, "Leakage reduction in fast superconducting qubit gates via optimal control", *npj Quantum Information*, vol. 7, no. 1, p. 14, 2021, ISSN: 2056-6387. DOI: 10.1038/s41534-020-00346-2.

[138] S. Y. Huang and H. S. Goan, "Optimal control for fast and high-fidelity quantum gates in coupled superconducting flux qubits", *Phys. Rev. A*, vol. 90, no. 1, p. 012 318, 2014. DOI: 10.1103/PhysRevA.90.012318.

[139] D. J. Egger and F. K Wilhelm, "Optimized controlled-z gates for two superconducting qubits coupled through a resonator", *Superconductor Science and Technology*, vol. 27, no. 1, p. 014 001, 2013. DOI: 10.1088/0953-2048/27/1/014001.

[140] P. Rebentrost and F. K. Wilhelm, "Optimal control of a leaking qubit", *Phys. Rev. B*, vol. 79, no. 6, p. 060 507, 2009. DOI: 10.1103/PhysRevB.79.060507.

[141] S. Jandura and G. Pupillo, "Time-Optimal Two- and Three-Qubit Gates for Rydberg Atoms", *Quantum*, vol. 6, p. 712, 2022, ISSN: 2521-327X. DOI: 10.22331/q-2022-05-13-712.

[142] M. Mohan, R. de Keijzer, and S. Kokkelmans, "Robust control and optimal rydberg states for neutral atom two-qubit gates", *Phys. Rev. Res.*, vol. 5, no. 3, p. 033 052, 2023. DOI: 10.1103/PhysRevResearch.5.033052.

[143] S. Jandura, J. D. Thompson, and G. Pupillo, "Optimizing rydberg gates for logical-qubit performance", *PRX Quantum*, vol. 4, no. 2, 2023.

[144] M. H. Goerz, E. J. Halperin, J. M. Aytac, C. P. Koch, and K. B. Whaley, "Robustness of high-fidelity rydberg gates with single-site addressability", *Phys. Rev. A*, vol. 90, no. 3, p. 032 329, 2014. DOI: 10.1103/PhysRevA.90.032329.

[145] M. M. Müller *et al.*, "Optimizing entangling quantum gates for physical systems", *Phys. Rev. A*, vol. 84, no. 4, p. 042 315, 2011. DOI: 10.1103/PhysRevA.84.042315.

[146] A. Steane, "Quantum computing", *Reports on Progress in Physics*, vol. 61, no. 2, p. 117, 1998. DOI: 10.1088/0034-4885/61/2/002.

[147] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.

[148] I. Buluta and F. Nori, "Quantum simulators", *Science*, vol. 326, no. 5949, pp. 108–111, 2009. DOI: 10.1126/science.1177838.

[149] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, "Measurement of qubits", *Phys. Rev. A*, vol. 64, no. 5, p. 052 312, 2001. DOI: 10.1103/PhysRevA.64.052312.

[150] J. B. Altepeter, D. F. James, and P. G. Kwiat, "4 qubit quantum state tomography", in *Quantum State Estimation*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 113–145, ISBN: 978-3-540-44481-7. DOI: 10.1007/978-3-540-44481-7_4.

[151] J. L. Park, "The concept of transition in quantum mechanics", *Foundations of Physics*, vol. 1, no. 1, pp. 23–33, 1970. DOI: 10.1007/BF00708652.

[152] D. E. Deutsch, A. Barenco, and A. Ekert, "Universality in quantum computation", *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, vol. 449, no. 1937, pp. 669–677, 1995. DOI: 10.1098/rspa.1995.0065.

[153] T. Sleator and H. Weinfurter, "Realizable universal quantum logic gates", *Phys. Rev. Lett.*, vol. 74, no. 20, pp. 4087–4090, 1995. DOI: 10.1103/PhysRevLett.74.4087.

[154] D. P. DiVincenzo, "Two-bit gates are universal for quantum computation", *Phys. Rev. A*, vol. 51, no. 2, pp. 1015–1022, 1995. DOI: 10.1103/PhysRevA.51.1015.

[155] A. Barenco *et al.*, "Elementary gates for quantum computation", *Phys. Rev. A*, vol. 52, no. 5, pp. 3457–3467, 1995. DOI: 10.1103/PhysRevA.52.3457.

[156] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes", *Nature Communications*, vol. 9, no. 1, p. 4812, 2018. DOI: 10.1038/s41467-018-07090-4.

[157] L. Broers and L. Mathey, "Mitigated barren plateaus in the time-nonlocal optimization of analog quantum-algorithm protocols", *Phys. Rev. Res.*, vol. 6, no. 1, p. 013 076, 2024. DOI: 10.1103/PhysRevResearch.6.013076.

[158] S. G. Schirmer, H. Fu, and A. I. Solomon, "Complete controllability of quantum systems", *Phys. Rev. A*, vol. 63, no. 6, p. 063 410, 2001. DOI: 10.1103/PhysRevA.63.063410.

[159] J. Li, X. Yang, X. Peng, and C.-P. Sun, "Hybrid quantum-classical approach to quantum optimal control", *Phys. Rev. Lett.*, vol. 118, no. 15, p. 150 503, 2017. DOI: 10.1103/PhysRevLett.118.150503.

[160] R. S. Judson and H. Rabitz, "Teaching lasers to control molecules", *Phys. Rev. Lett.*, vol. 68, no. 10, pp. 1500–1503, 1992. DOI: 10.1103/PhysRevLett.68.1500.

[161] S. Ebadi *et al.*, "Quantum optimization of maximum independent set using rydberg atom arrays", *Science*, vol. 376, no. 6598, pp. 1209–1215, 2022. DOI: 10.1126/science.abo6587.

[162] H.-P. Breuer and F. Petruccione, *The theory of open quantum systems*. Oxford ; New York: Oxford University Press, 2002, 625 pp., ISBN: 978-0-19-852063-4.

[163] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, "Optimal control of coupled spin dynamics: Design of NMR pulse sequences by gradient ascent algorithms", *Journal of Magnetic Resonance*, vol. 172, no. 2, pp. 296–305, 2005, ISSN: 10907807. DOI: 10.1016/j.jmr.2004.11.004.

[164] L. Kley, N. Heimann, A. Parvej, L. Broers, and L. Mathey, *Optimal recoil-free state preparation in an optical atom tweezer*, 2024.

[165] D. Barredo, S. d. Léséleuc, V. Lienhard, T. Lahaye, and A. Browaeys, "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.

[166] 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, ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aah3752.

[167] K.-N. Schymik *et al.*, "Enhanced atom-by-atom assembly of arbitrary tweezer arrays", *Phys. Rev. A*, vol. 102, no. 6, p. 063 107, 2020. DOI: 10.1103/PhysRevA.102.063107.

[168] K. Barnes *et al.*, "Assembly and coherent control of a register of nuclear spin qubits", *Nature Communications*, vol. 13, no. 1, p. 2779, 2022. DOI: 10.1038/s41467-022-29977-z.

[169] K. Schymik *et al.*, "In situ equalization of single-atom loading in large-scale optical tweezer arrays", *Phys. Rev. A*, vol. 106, no. 2, p. 022 611, 2022. DOI: 10.1103/PhysRevA.106.022611.

[170] S. Ma, A. P. Burgers, G. Liu, J. Wilson, B. Zhang, and J. D. Thompson, "Universal Gate Operations on Nuclear Spin Qubits in an Optical Tweezer Array of Yb 171 Atoms", *Physical Review X*, vol. 12, no. 2, p. 021 028, 2022, ISSN: 2160-3308. DOI: 10.1103/PhysRevX.12.021028.

[171] T. Xia *et al.*, "Randomized benchmarking of single-qubit gates in a 2d array of neutral-atom qubits", *Phys. Rev. Lett.*, vol. 114, no. 10, p. 100 503, 2015. DOI: 10.1103/PhysRevLett.114.100503.

[172] T. M. Graham *et al.*, "Multi-qubit entanglement and algorithms on a neutral-atom quantum computer", *Nature*, vol. 604, no. 7906, pp. 457–462, 2022, ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-022-04603-6.

[173] S. J. Evered *et al.*, "High-fidelity parallel entangling gates on a neutral-atom quantum computer", *Nature*, vol. 622, no. 7982, pp. 268–272, 2023. DOI: 10.1038/s41586-023-06481-y.

[174] L. Henriet *et al.*, "Quantum computing with neutral atoms", *Quantum*, vol. 4, p. 327, 2020, ISSN: 2521-327X. DOI: 10.22331/q-2020-09-21-327.

[175] I. Cong, H. Levine, A. Keesling, D. Bluvstein, S. T. Wang, and M. D. Lukin, "Hardware-efficient, fault-tolerant quantum computation with rydberg atoms", *Phys. Rev. X*, vol. 12, no. 2, p. 021 049, 2022. DOI: 10.1103/PhysRevX.12.021049.

[176] D. Bluvstein *et al.*, *Architectural mechanisms of a universal fault-tolerant quantum computer*, arXiv: 2506.20661, 2025.

[177] Y. Wu, S. Kolkowitz, S. Puri, and J. Thompson, "Erasure conversion for fault-tolerant quantum computing in alkaline earth rydberg atom arrays", *Nature Communications*, vol. 13, p. 4657, 2022. DOI: 10.1038/s41467-022-32094-6.

[178] T. Caneva, T. Calarco, and S. Montangero, "Chopped random-basis quantum optimization", *Physical Review A*, vol. 84, no. 2, p. 022 326, 2011, ISSN: 1050-2947, 1094-1622. DOI: 10.1103/PhysRevA.84.022326.

[179] M. Saffman, I. I. Beterov, A. Dalal, E. J. Páez, and B. C. Sanders, "Symmetric Rydberg controlled- Z gates with adiabatic pulses", *Physical Review A*, vol. 101, no. 6, p. 062 309, 2020, ISSN: 2469-9926, 2469-9934. DOI: 10.1103/PhysRevA.101.062309.

[180] S. Ravets, H. Labuhn, D. Barredo, L. Béguin, T. Lahaye, and A. Browaeys, "Coherent dipole–dipole coupling between two single rydberg atoms at an electrically-tuned förster resonance", *Nature Physics*, vol. 10, no. 12, pp. 914–917, 2014. DOI: 10.1038/nphys3119.

[181] L. Henriet, "Robustness to spontaneous emission of a variational quantum algorithm", *Phys. Rev. A*, vol. 101, no. 1, p. 012 335, 2020. DOI: 10.1103/PhysRevA.101.012335.

[182] C. Dalyac, L.-P. Henry, M. Kim, J. Ahn, and L. Henriet, "Exploring the impact of graph locality for the resolution of the maximum- independent-set problem with neutral atom devices", *Phys. Rev. A*, vol. 108, no. 5, p. 052 423, 2023. DOI: 10.1103/PhysRevA.108.052423.

[183] A. D. Ludlow, M. M. Boyd, J. Ye, E. Peik, and P. O. Schmidt, "Optical atomic clocks", *Rev. Mod. Phys.*, vol. 87, no. 2, pp. 637–701, 2015. DOI: 10.1103/RevModPhys.87.637.

[184] A. J. Daley, M. M. Boyd, J. Ye, and P. Zoller, "Quantum computing with alkaline-earth-metal atoms", *Phys. Rev. Lett.*, vol. 101, no. 17, p. 170 504, 2008. DOI: 10.1103/PhysRevLett.101.170504.

[185] A. V. Gorshkov *et al.*, "Alkaline-earth-metal atoms as few-qubit quantum registers", *Phys. Rev. Lett.*, vol. 102, no. 11, p. 110 503, 2009. DOI: 10.1103/PhysRevLett.102.110503.

[186] K. Gosar *et al.*, "Preparation of ultracold atomic-ensemble arrays using time-multiplexed optical tweezers", *Phys. Rev. A*, vol. 106, no. 2, p. 022 604, 2022. DOI: 10.1103/PhysRevA.106.022604.

[187] A. Cooper, J. P. Covey, I. S. Madjarov, S. G. Porsev, M. S. Safronova, and M. Endres, "Alkaline-Earth Atoms in Optical Tweezers", *Physical Review X*, vol. 8, no. 4, p. 041 055, 2018, ISSN: 2160-3308. DOI: 10.1103/PhysRevX.8.041055.

[188] M. A. Norcia, A. W. Young, and A. M. Kaufman, "Microscopic Control and Detection of Ultracold Strontium in Optical-Tweezer Arrays", *Physical Review X*, vol. 8, no. 4, p. 041 054, 2018, ISSN: 2160-3308. DOI: 10.1103/PhysRevX.8.041054.

[189] I. S. Madjarov *et al.*, "High-fidelity entanglement and detection of alkaline-earth rydberg atoms", *Nature Physics*, vol. 16, no. 8, pp. 857–861, 2020. DOI: 10.1038/s41567-020-0903-z.

[190] T. Topcu and A. Derevianko, "Possibility of triple magic trapping of clock and rydberg states of divalent atoms in optical lattices", *Journal of Physics B: Atomic, Molecular and Optical Physics*, vol. 49, no. 14, p. 144 004, 2016. DOI: 10.1088/0953-4075/49/14/144004.

[191] S. Zhang, F. Robicheaux, and M. Saffman, "Magic-wavelength optical traps for rydberg atoms", *Phys. Rev. A*, vol. 84, no. 4, p. 043 408, 2011. DOI: 10.1103/PhysRevA.84.043408.

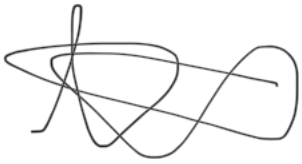
[192] K. Huang, *Statistical Mechanics, 2nd Edition*. 1987.

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben.

Sofern im Zuge der Erstellung der vorliegenden Dissertationsschrift generative Künstliche Intelligenz (gKI) basierte elektronische Hilfsmittel verwendet wurden, versichere ich, dass meine eigene Leistung im Vordergrund stand und dass eine vollständige Dokumentation aller verwendeten Hilfsmittel gemäß der Guten wissenschaftlichen Praxis vorliegt. Ich trage die Verantwortung für eventuell durch die gKI generierte fehlerhafte oder verzerrte Inhalte, fehlerhafte Referenzen, Verstöße gegen das Datenschutz- und Urheberrecht oder Plagiate.

Hamburg, den 16.10.2025



Unterschrift der Doktorandin / des Doktoranden