Automated optimization, learning and control for cold atom platforms

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Martino Calzavara

aus Camposampiero, Italien

angenommen im Jahr 2025

List of Publications

- M. Calzavara, Y. Kuriatnikov, A. Deutschmann-Olek, F. Motzoi, S. Erne, A. Kugi, T. Calarco, J. Schmiedmayer, and M. Prüfer. Optimizing optical potentials with physics-inspired learning algorithms. Phys. Rev. Appl., 19:044090, 4, Apr. 2023. doi: 10.1103/PhysRevApplied.19.044090.

- C. Cicali, M. Calzavara, E. Cuestas, T. Calarco, R. Zeier, and F. Motzoi. Fast neutral-atom transport and transfer between optical tweezers. Phys. Rev. Appl., 24:024070, 2, Aug. 2025. doi: 10.1103/7r3w-8m61.

- M. Calzavara, T. Calarco, and F. Motzoi. Classical feature map surrogates and metrics for quantum control landscapes, 2025. arXiv: 2509.25930 [quant-ph]. url: https://arxiv.org/abs/2509.25930

  - A. Deutschmann-Olek, M. Tajik, M. Calzavara, J. Schmiedmayer, T. Calarco, and A. Kugi. Iterative shaping of optical potentials for one-dimensional bose-einstein condensates. In 2022 IEEE 61st Conference on Decision and Control (CDC), pages 5801–5806, 2022. doi: 10.1109/CDC51059.2022.9993271.
  - M. Calzavara and L. Salasnich. Dark Solitons in the Unitary Bose Gas, 2020. Symmetry, 12(6), 957. doi: 10.3390/sym12060957

Abstract

Quantum technologies based on cold atom platforms, while benefiting from highly stable and scalable atomic systems, still rely on the precise characterization, calibration, and control of complex experimental apparatuses. In this thesis, we develop automated approaches to address these critical tasks, integrating physical insight with optimization and learning methods, while complying with operational constraints.

The bulk of the work presented here consists of three research papers. In the first one, we introduce physics-inspired machine-learning models to optimize optical dipole potentials for shaping ultracold gasses, achieving order-of-magnitude improvements in speed. In the second paper, we investigate atom transport between optical tweezers, a key auxiliary operation for scalable quantum processors, and demonstrate how shortcuts to adiabaticity and quantum optimal control can minimize heating and excitations during fast transport. The third and last paper analyzes a class of optimization landscapes relevant for quantum control, deriving from first principles a set of classical surrogates. We show how time and energy constraints translate into limited bandwidth and derivatives for the landscape, with consequences for the design of regression models and for optimization, which we relate to bounds in associated metrics.

Together, these results provide concrete examples of how automated optimization, learning, and control can expand the toolset available for building quantum firmware for cold atom platforms. The methods presented not only improve experimental performance but keep implementation overhead at a minimum, thereby simplifying operations and facilitating scale-up efforts.

Contents

Chapter 1

Introduction

Long after spawning the first clues of physics beyond the classical realm which lead to the development of Quantum Mechanics, Atomic, Molecular and Optical (AMO) platforms have established themselves nowadays as strong contenders in the Second Quantum Revolution [1]. In this regard, making use of individual quantum systems (i.e. atoms) which are guaranteed to be identical by the laws of Nature proved to be not only an elegant idea, but also a quite effective one. Yet it took almost a century before this could be fully realized experimentally: from the Stern-Gerlach experiment (1922) [2] in which quantum state preparation through projective measurements was demonstrated in atomic beams for the first time, it took at least the invention of the laser (1960) [3] and of laser cooling and trapping (1980's) [4] to get to the first Bose-Einstein Condensate (1995) [5, 6]. From there on, steady improvement in experimental techniques allowed to achieve single atom addressing in Quantum Gas Microscopes using optical lattices loaded with bosons (2009) [7] and fermions (2015) [8].

In current experiments with cold atoms, it is customary to trap, individually manipulate and measure from tens to several hundreds of atoms [9, 10, 11], with recent demonstrations reaching up to thousands [12, 13], and coherence times in the scale of 10 s. Meanwhile, in experiments with ultracold gasses clouds containing $\sim 10^{3\div5}$ atoms can be created, giving up single atom addressing in exchange for control over their collective degrees of freedom [14, 15]. In these systems not only external potentials can be controlled almost arbitrarily by shaping optical dipole potentials [16] and magnetic fields [17], but even interactions can be harnessed by using Rydberg states [18] and tuned via magnetic fields by means of Feshbach resonances [19].

1.1 Applications

The high degree of controllability and long coherence times of cold atom platforms makes them very attractive analog quantum simulators [20], physically implementing the Hamiltonian of complex quantum systems which would be otherwise very hard to simulate on a classical computer or to measure directly in an experiment. This idea found application in the field of strongly correlated condensed matter systems, where atom arrays have been used to simulate the Superfluid-Mott insulator phase transition [21, 22], Bose- and Fermi-Hubbard models [23, 24], and spin-chain Hamiltonians [25, 26, 27, 18], while field theories like the sine-Gordon model can be simulated using ultracold gasses [15]. Recently a lot of emphasis has been put on topological matter [28], exploiting the electronic level structure to implement synthetic dimensions and gauge fields [29], and implementing effective Floquet Hamiltonians using periodic driving [30].

Another prospective application of cold atom platforms is gate-based quantum information processing. In this case the machine operates at a higher level of abstraction, using a universal set of unitary operations to implement quantum algorithms [31], which can perform digital quantum simulation [32] or solve more generic mathematical problems, like integer factorization [33], search [34] or linear systems [35]. Critically, the gate-based approach is compatible with Error Correction, effectively allowing to trade higher requirements in physical qubits for increased robustness against decoherence [36]. The implementations of these algorithms can benefit from the high connectivity of reconfigurable atom arrays and fast unitary gates ($\gtrsim 1\,\mu$s), while the long preparation ($\gtrsim 100\,$ms) and readout ($\gtrsim 500\,\mu$s) times are a limiting factor, giving rise to low repetition rates hardly exceeding ~ 2 Hz [37]. Overall, this translates in an estimated requirement of $\sim 10^7$ physical qubits to factor a 2048 bit integer using Shor's algorithm on a time scale of days [38]. While these number of qubits are currently out of reach, steady progress in the development of these platforms is being made, with operation of up to ~ 50 logical qubits using Rydberg atoms in optical tweezers [11], which constitute one of the most promising avenues for scaling the number of physical qubits up to $\sim 10^4$ in the near future [13].

1.2 Automated optimization, learning and control

As these systems are scaled up, the complexity related to calibrating, characterizing and controlling these systems increases too, and is to be tackled on the one side by hardware improvement but also by developing control firmware which is up to its task [39]. This consists in providing an interface between hard-

ware and software operating at a higher abstraction level, matching on one side the hardware-generated signals with physically meaningful output metrics (characterization/calibration) and on the other side user directives with appropriate hardware inputs (control). As real world hardware involves both unknown parameters and uncertainty, this matching task requires in practice solving minimization problems related to a simplified model of the system at hand. Automated optimization plays therefore an essential role in the quantum firmware, both for control/calibration purposes and for adapting the models to the real system by using the information gathered during operation (learning) [40, 41].

While in this sense cold atoms – being identical, stable and well characterized systems – are at an advantage compared to engineered quantum systems like superconducting circuits, in practice they still rely on non-ideal hardware for measures and control, whose transfer function need to be taken into account [42]. Moreover, the mentioned low repetition rates constitute a serious bottleneck for data acquisition, tightening the operational constraints for optimization and learning strategies. While several approaches to tackle these challenges are available (e.g. using pre-training and data amplification), the main tool that we employ across this thesis is to derive specialized methods and Ansätze which encode as much physical information about the system as possible [43, 44]. This typically results effectively in a dimensional reduction of the initial problem, allowing to achieve convergence with a small experimental data set or to treat the problem analytically, while preserving generalization capabilities. This approach mirrors the spirit underlying large part of physics, in which deriving simplified effective models is one of main ways in which we can make sense of Nature, both compressing the information coming from experiments and allowing to easily compute quantities of interest, guiding further investigation by providing clues as to how to achieve relevant phases of matter or physical regimes.

1.3 Structure of the thesis

The contents of this thesis are organized into two parts: the background which is needed for a self-sufficient treatment of the results is presented in Part I, giving a brief account of the operational principles of cold atom experiments in Chapter 2, and introducing the main ideas behind quantum control, both using numerical and analytical methods in Chapter 3. The three publications in Chapter 4,5 and 6, constituting the bulk of the Author's research work, are then gathered in Part II, where their contents are put in the context of the thesis. Finally, Chapter 7 discusses the relevance of the results in the development of a quantum firmware for cold atom platforms.

Part I

Background and Motivation

Chapter 2

Quantum hardware: physics and controls of cold atoms

Before approaching the core topics of this thesis, which are related to firmware development for cold atom platforms, it is helpful to first introduce some basic concepts regarding this class of experiments. While this Chapter is not intended as an exhaustive treatment of the physics of cold atoms, for which many excellent sources are available [45, 46, 47, 48, 49, 50], it should give the reader enough information to navigate the results presented in Part II. More specifically, we cover the basic atomic physics involved in the qualitative structure and nomenclature of energy levels, together with the external potentials whose control is one of the main concerns of Part II, and general operational principles and structure of cold atom platforms.

2.1 Controlling atoms with electromagnetic fields

External electromagnetic fields are the main control knobs at our disposal to influence the internal and motional degrees of freedom of cold atoms. In order see how this is possible, we must first discuss the main features of the emission spectrum of alkali atoms. Because of their simple structure, featuring strong and narrow transitions in spectral regions which are accessible to industrial lasers, they figure among the most established species used in cold atom platforms.

The starting point of our discussion is the Hamiltonian H_0 of an alkali atom, i.e. belonging to the first group of the periodic table. Since the atoms in this group only have one electron in the outer shell, the many-electron system can be treated approximately by only considering the dynamics of the valence electron in the field of the (rigid) ionic core containing the closed shell electrons and the

nucleus. By using the relative and center of mass coordinates $\vec{r}, \vec{R}$ given by:

$$\vec{r} = \vec{r}_e - \vec{r}_{\text{core}}, \quad \vec{R} = \frac{m_e \vec{r}_e + m_{\text{core}} \vec{r}_{\text{core}}}{m_e + m_{\text{core}}},$$

with associated momenta $\vec{p}$ and $\vec{P}$, the Hamiltonian of the system then reads as follows:

$$H_0 \approx \underbrace{\frac{\vec{P}^2}{2M}}_{\text{Motional d.o.f.}} + \underbrace{\frac{\vec{p}^2}{2m_r} + \tilde{V}(\vec{r})}_{\text{Electronic d.o.f.}}, \tag{2.1}$$

where $M = m_{\text{core}} + m_e$ is the mass of the atom, m_r is the reduced mass of the electron

$$\frac{1}{m_r} = \frac{1}{m_{\text{core}}} + \frac{1}{m_e},$$

and $\tilde{V}(\vec{r})$ is a modified central potential that accounts for the finite-size charge distribution in the core, thus only affecting the radial part of the wavefunction. Quantum defect theory can then be used to compute the correction to the hydrogenic energy levels caused by the modified potential, obtaining the modified Rydberg formula [51]:

$$E_{nL} = -\frac{R_\infty}{(n - \delta_L)^2} \left(\frac{m_r}{m_e} \right). \tag{2.2}$$

The *quantum defect* term δ_L appearing in Eq. (2.2) depends on the atomic species and is larger for low angular momentum L, as the corresponding wavefunctions have larger overlap with the ionic core. For the $L = 0, 1$ orbitals of most alkali atoms, δ_L is comparable with the smallest available principal quantum number n for the valence electron, resulting in an energy correction in the eV scale [51]. This effect lifts the degeneracy with respect to L of the electronic Hamiltonian, leading to the $L = 0$ ground state and the so-called D *emission line* connecting it to the first excited $L = 1$ orbital, both characteristic of alkali spectra. The alkali D-line wavelengths are in the optical range $580 \div 900\,\text{nm}$ (e.g. $\sim 780\,\text{nm}$ for ^{87}Rb [52]) and thus addressable with commonly available lasers.

The leading corrections to Eq.(2.1) give rise to the *fine structure*, splitting the D-line into a doublet. An important contribution to the perturbation H_{fs} comes from the spin-orbit coupling which, in the moving frame the valence electron, can be thought of as the effect of the internal magnetic field caused by the motion of the ionic core, leading to a Hamiltonian of the form

$$H_{SO} = A_{fs} \vec{L} \cdot \vec{S},$$

where $\vec{L}$ and $\vec{S}$ represent respectively the angular and spin angular momentum of the valence electron. While H_0 commutes with $\vec{L}$ and $\vec{S}$, the same is not

true for H_{SO}, so that the azimuthal numbers m_L, m_S are not conserved anymore. If instead we consider the total electronic angular momentum $\vec{J} = \vec{L} + \vec{S}$, we have $[H_{SO}, \vec{J}] = \vec{0}$, so that J, m_J are good quantum numbers. We can then express the energy eigenstates using the quantum numbers L, S, J, m_J (instead of L, S, m_L, m_S), giving rise to the LS-coupling of angular momentum. This motivates the usage of the corresponding spectroscopic notation $^{2S+1}L_J$, where the S, P, D, F notation for L is employed (S stands for $L = 0$, P for $L = 1$, D for $L = 2$ and so on). Thus, the D_1 and D_2 components of an alkali fine structure doublet correspond respectively to the transitions $^2S_{1/2} \rightarrow {}^2L_{1/2}$ and $^2S_{1/2} \rightarrow {}^2L_{3/2}$. As an example, for ^{87}Rb the fine splitting energy is around $\Delta E_{fs} \sim 0.03\,\text{eV}$ corresponding to $\Delta\lambda_{fs} \sim 15\,\text{nm}$ [52] which is enough to be resolved by commercial lasers, that can be therefore used to address only one of the two transitions selectively.

The next order of the perturbative expansion contains the interaction between the magnetic moment of the ionic core and the internal magnetic field due to the motion of the valence electron, resulting in the *hyperfine structure*. The corresponding Hamiltonian is given by[1]

$$H_{hfs} = -\vec{B}_J(\vec{0}) \cdot \vec{M}_I \approx A_{hfs} \vec{I} \cdot \vec{J}, \tag{2.3}$$

where $\vec{I}$ is the spin of the ionic core, and A_{hfs} is the hyperfine structure constant, which depends on the energy level and on the atom. The approximation in Eq. (2.3), which holds for the states in the D-line, lies in discarding couplings between states with different J (or L), which are separated in energy by the much larger fine structure terms, and therefore contribute negligibly to the perturbed eigenstates [53]. Analogously to the previous discussion regarding the spin-orbit coupling, this term breaks the invariance of $\vec{I}$ and $\vec{J}$, and by defining the total angular momentum $\vec{F} = \vec{I} + \vec{J}$, we can obtain a new set of conserved quantum numbers as L, S, J, I, F, m_F. The ground state is thus split into a doublet, whose components in the case of ^{87}Rb ($I = 3/2$) are $F = 1, 2$, standing $\Delta E_{hfs} \sim 3 \times 10^{-5}\,\text{eV}$ apart from each other, corresponding to a radiofrequency transition of $\sim 7\,\text{GHz}$ [52].

Thanks to their long lifetimes (in the order of 10 s in realistic scenarios), hyperfine states serve as ideal qubits or pseudo-spins for quantum computing or simulation, offering a wide range of controls over their internal state (e.g. with optical Raman transitions), interactions (through Feshbach resonance) and trapping potentials. In what follows we are going to review the external potentials

[1] Here we neglect smaller terms due to higher orders of the multipole expansion and ionic core polarizability.

which are most relevant in the context of this thesis, as they will play a central role in Part II.

2.1.1 Static magnetic fields: Zeeman shift

The first class of potentials we focus on are due to the presence of a static magnetic field. Considering the uniform case $\vec{B} = B_0\hat{z}$, an atom with a non-zero magnetic moment $\vec{M}$ will couple to $\vec{B}$ through the Zeeman Hamiltonian [52]:

$$H_Z = -(\vec{M}_L + \vec{M}_S + \vec{M}_I) \cdot \vec{B} = \frac{\mu_B}{\hbar}(g_L L_z + g_S S_z - \frac{\mu_N}{\mu_B} g_I I_z)B_0, \tag{2.4}$$

where the total magnetic moment $\vec{M}$ is the sum of the orbital $\vec{M}_L$, spin $\vec{M}_S$ and nuclear $\vec{M}_I$ contributions. We then expressed the Hamiltonian in Eq. (2.4) in terms of Bohr's magneton $\mu_B = e\hbar/(2m_e)$, the nuclear magneton $\mu_N = e\hbar/(2m_p)$, the angular moments $\vec{L}, \vec{S}, \vec{I}$ and their corresponding non-dimensional Landé g-factors $g_L = 1$, $g_S \approx 2$, and g_I.

In the weak field limit relative to the fine structure $H_Z \ll H_{fs}$ the energy contribution of the Zeeman term can be computed as a perturbation, by evaluating the expectation values of H_Z over the original eigenstates of H_{fs}. The resulting expression can be derived equivalently using the vector precession formalism: as $\vec{J}$ is approximately conserved (up to the perturbation), we can think of $\vec{L}$ and $\vec{S}$ as classical vectors precessing fast around $\vec{J}$, while $\vec{J}$ precesses slowly around the external field $\vec{B}$. Because of this frequency scale separation, only the parallel components $\vec{L} \cdot \vec{J}$ and $\vec{S} \cdot \vec{J}$ contribute appreciably to the expectation value of the perturbation [49]. Since $\mu_N/\mu_B \sim 10^{-3}$, we can also neglect the small contribution due to the nuclear spin $\vec{I}$ to obtain the following:

$$H_Z \approx \frac{\mu_B}{\hbar} \frac{(\vec{L} + 2\vec{S}) \cdot \vec{J}\vec{B} \cdot \vec{J}}{\vec{J}^2} = \frac{\mu_B}{\hbar} \left(\vec{J}^2 + \frac{\vec{J}^2 - \vec{L}^2 + \vec{S}^2}{2} \right) \frac{B_0 J_z}{\vec{J}^2}$$

so that the expectation value yields

$$\langle LSJm_J | \, H_Z \, | LSJm_J \rangle = \mu_B \underbrace{\left(1 + \frac{J(J+1) - L(L+1) + S(S+1)}{2J(J+1)} \right)}_{g_J} B_0 m_J, \tag{2.5}$$

which gives us an expression for the overall electronic Landé g-factor g_J.

In the weak field limit for the hyperfine structure $H_Z \ll H_{hfs}$ a similar reasoning can be applied, but now the vectors $\vec{J}$ and $\vec{I}$ are precessing around the (approximately) conserved total momentum $\vec{F}$, which in turn slowly precesses

around $\vec{B}$. By projecting all vector quantities in Eq. (2.5) along $\vec{F}$ we obtain

$$H_Z \approx \frac{\mu_B g_J}{\hbar} \frac{(\vec{J}\cdot\vec{F})(\vec{B}\cdot\vec{F})}{\vec{F}^2} = \frac{\mu_B g_J}{\hbar}\left(\frac{\vec{F}^2+\vec{J}^2-\vec{I}^2}{2}\right)\frac{B_0 F_z}{\vec{F}^2},$$

whose expectation value on the unperturbed hyperfine eigenstates is given by:

$$\langle LSJIFm_F|\, H_Z\, |LSJIFm_F\rangle = \mu_B \underbrace{g_J\left(\frac{F(F+1)+J(J+1)-I(I+1)}{2F(F+1)}\right)}_{g_F} B_0 m_F. \tag{2.6}$$

Let us now suppose that the magnetic field $\vec{B}(\vec{r})$ is inhomogeneous. Provided that the atoms are following an adiabatic evolution in the external field [17], they will experience a conservative potential given by

$$U_{\mathrm{mag}}(\vec{r}) = \mu_B g m |\vec{B}(\vec{r})|,$$

where the azimuthal quantum number m and Landé factor g are referred to the approximately conserved angular momentum (which as we saw before depends on the context). The sign of gm is of particular importance, distinguishing between high- ($gm<0$) and low-field ($gm>0$) seeking states. In the case of the hyperfine ground state $^2\mathrm{S}_{1/2}$ of ^{87}Rb, Eq. (2.6) leads to $g_F=-1/2$ for $F=1$ and $g_F=1/2$ for $F=2$, so that $|F=2,m_F=1,2\rangle$ and $|F=1,m_F=-1\rangle$ are low-field seeking states, while $|F=2,m_F=-1,-2\rangle$ and $|F=1,m_F=1\rangle$ are high-field seeking.

2.1.2 Oscillating electric fields: Optical dipole potentials

The other class of potentials that is most relevant for this thesis is generated by shining light on the atoms. If the light frequency is far detuned from a D-line transition, the effects on the ground state can be described strikingly well by means of a semiclassical model. In fact, since in this regime the rate of photon absorption and hence the populations of excited states are very low, saturation phenomena are avoided, effectively allowing us to treat the atom as a classical harmonic oscillator and giving rise to a conservative potential.

In presence of an oscillating electric field $\vec{\epsilon}(t)=\epsilon\hat{z}\cos(\omega t)$[2], the free Hamiltonian will be modified by a dipole term

$$H_d(t) = -\vec{\epsilon}(t)\cdot\vec{D} = \epsilon e z\cos(\omega t). \tag{2.7}$$

[2]For simplicity, we only consider the linearly polarized case. Most notably, the polarization determines which angular momentum sub-levels are connected by transitions through selection rules.

Even though the expectation value of the electric dipole $\langle \vec{D} \rangle \propto \langle z \rangle$ is zero in the spherically symmetric ground state ${}^2S_{1/2}$ ($L = 0$), the electric field induces a non-zero polarization $\vec{D}(t)$ that we can compute using time dependent perturbation theory [46]. In the general setting, we have a system described by the Hamiltonian

$$H(t) = H_0 + \lambda W(t),$$

where $|H_0| \sim |W|$ and $\lambda \ll 1$. We expand the solution to the time-dependent Schrödinger equation in the eigenbasis of the unperturbed Hamiltonian $H_0 |n\rangle = E_n |n\rangle, n = 0, 1, 2, \dots$ in its moving frame, obtaining

$$|\psi(t)\rangle = \sum_n e^{-i\frac{E_n}{\hbar}t} b_n(t) |n\rangle ,$$

where the slowly changing coefficients $b_n(t)$ satisfy

$$b_n(t) = \frac{\lambda}{i\hbar} \sum_k \int_0^t dt' W_{nk}(t') b_k(t') e^{i\omega_{nk}t'}, \tag{2.8}$$

and we defined the transition frequency $\omega_{nk} = (E_n - E_k)\hbar^{-1}$ and the matrix element $W_{nk}(t) = \langle n| W(t) |k\rangle$. Focusing on the sinusoidal case $\lambda W(t) = \lambda W \cos(\omega t)$, we can then fix the initial condition $|\psi(t = 0)\rangle = |k\rangle$ at zeroth order $b_n^{(0)} = \delta_{nk}$ and then obtain the higher orders recursively using Eq. (2.8). At first order we obtain the general solution

$$\begin{aligned} |\psi(t)\rangle = &\left(1 + \frac{\lambda}{2i\hbar} W_{kk} \left[\frac{e^{i\omega t} - e^{-i\omega t}}{i\omega}\right]\right) e^{-i\frac{E_k}{\hbar}t} |k\rangle + \\ &+ \frac{\lambda}{2i\hbar} \sum_{n\neq k} W_{nk} \left[\frac{e^{i(\omega+\omega_{nk})t} - 1}{i(\omega + \omega_{nk})} + \frac{e^{i(-\omega+\omega_{nk})t} - 1}{i(-\omega + \omega_{nk})}\right] e^{-i\frac{E_n}{\hbar}t} |n\rangle + \mathcal{O}(\lambda^2), \end{aligned}$$

which we can use to compute the expectation value of the (time independent) perturbation operator $\langle W(t)\rangle = \langle \psi(t)| W |\psi(t)\rangle$

$$\langle W(t)\rangle \quad = \quad \left(W_{kk} + \frac{2\lambda}{\hbar} \sum_{n\neq k} \frac{|W_{nk}|^2 \omega_{nk}}{\omega_{nk}^2 - \omega^2} (\underbrace{\cos(\omega_{nk} t)}_{\text{Transient term}} - \cos(\omega t)) \right) \; + \; \mathcal{O}(\lambda^2).$$

The expression we obtained features the sum of a term that comes from H_0 and as such oscillate at the resonant frequency ω_{nk}, and a term that oscillate at the driving frequency ω. As in practice physical systems involve dissipation and therefore a finite damping rate Γ, the long-term response of the system is dominated by the driven term ω, allowing us to neglect the transient term over time scales much longer than $2\pi\Gamma^{-1}$ [49].

We can finally obtain the relevant expression in the specific case of the electric dipole operator in Eq. (2.7) and the ground state $k = 0$ by substituting $\lambda \to -\epsilon, W \to -ez$:

$$\langle D_z(t)\rangle = \frac{2e^2}{\hbar}\sum_{n\neq 0}\frac{|\langle n|\, z\,|0\rangle|^2\omega_{n0}}{\omega_{n0}^2-\omega^2}\epsilon\cos(\omega t) = \chi(\omega)\epsilon(t),$$

obtaining a linear response relation featuring the (real) polarizability $\chi(\omega)$. We can further simplify the expression by introducing the non-dimensional *oscillator strengths* f_{n0} [48]

$$f_{n0} = \frac{2m}{\hbar}\omega_{n0}\left|\langle n|\, z\,|0\rangle\right|^2,$$

which allows us to write the polarizability as a weighted average of harmonic oscillators with frequencies ω_{n0}

$$\chi(\omega) = \sum_n f_{n0}\frac{e^2}{m(\omega_{n0}^2-\omega^2)} = \sum_n f_{n0}\chi^{HO}_{\omega_{n0},0}(\omega).$$

All quantum mechanical effects are encoded in the factors f_{n0}, while the remaining physics is the same as for a charged classical harmonic oscillator. In fact, the polarizabilities $\chi^{HO}_{\omega_{n0},\Gamma_{n0}}(\omega)$ can be derived from the linear response of the system obeying the following law of motion:

$$\ddot{z}(t) + \omega_{n0}^2 z(t) + \Gamma_{n0}\dot{z}(t) = -\frac{e}{m}\epsilon e^{-i\omega t}$$

by setting $\Gamma_{n0} = 0$. Damping effects due to spontaneous emission can be added phenomenologically in the form of a friction term $\Gamma_{n0}\dot{z}$, leading to the (complex) polarizability:

$$\chi^{HO}_{\omega_{n0},\Gamma_{n0}}(\omega) = \frac{e^2}{m(\omega_{n0}^2-\omega^2-i\Gamma_{n0}\omega)},$$

where Γ_{n0} is given by the spontaneous decay rate for a two-levels system [48]

$$\Gamma_{n0} = \frac{e^2\omega_{n0}^3}{3\pi\varepsilon_0\hbar c}\left|\langle n|\, z\,|0\rangle\right|^2.$$

The overall polarizability is then given by

$$\chi(\omega) = \sum_n f_{n0}\frac{e^2}{m(\omega_{n0}^2-\omega^2-i\Gamma_{n0}\omega)} = \sum_n f_{n0}\chi^{HO}_{\omega_{n0},\Gamma_{n0}}(\omega) \tag{2.9}$$

and correspondingly we have the following electric dipole for a real field $\vec{\epsilon}(t) = \vec{\epsilon}\cos(\omega t)$:

$$\vec{D}(t) = \vec{\epsilon}\left(\mathrm{Re}\,\chi(\omega)\cos(\omega t) + \mathrm{Im}\,\chi(\omega)\sin(\omega t)\right).$$

Now that we have the expression for the polarizability we can compute the energy contribution to the system in the presence of a light field. When treating light as a classical field, the dipole term in Eq. (2.7) is the leading term of the resulting Hamiltonian, as in the optical range $a_0/\lambda \sim 10^{-4}$, so the atom will experience a very homogeneous field. Then, the atom will fill the average potential energy of an induced dipole in the electric field [54, 55], that is:

$$U_{\text{dip}} = -\frac{1}{2}\langle \vec{\epsilon} \cdot \vec{D} \rangle = -\frac{\epsilon^2}{4} \operatorname{Re} \chi(\omega) = -\frac{\operatorname{Re} \chi(\omega)}{2\varepsilon_0 c} I,$$

where $\langle \cdot \rangle$ denotes a time average, that creates an extra 1/2 factor on top of the one coming from the fact that we are dealing with an induced dipole. The light intensity I is defined in terms of the Poynting vector $\vec{S} = \mu_0^{-1} \vec{\epsilon} \times \vec{B}$. For a (real) plane wave $\vec{\epsilon}(t, \vec{r}) = \epsilon \cos\Big(\omega t - \vec{k} \cdot \vec{r}\Big)$, $\vec{B}(t, \vec{r}) = c^{-1}(\hat{k} \times \vec{\epsilon}) \cos\Big(\omega t - \vec{k} \cdot \vec{r}\Big)$, we have

$$I = \langle |\vec{S}(t, \vec{0})| \rangle = \frac{\epsilon^2}{2} \varepsilon_0 c.$$

This result shows how the reactive part of atom-light interaction determines the appearance of a conservative potential. In the case of a spatially inhomogeneous light field[3], it will be proportional to the local light intensity, thus exerting a force given by

$$\vec{F}_{\text{dip}} = -\vec{\nabla} U_{\text{dip}} = \frac{\operatorname{Re} \chi(\omega)}{2\varepsilon_0 c} \vec{\nabla} I(\vec{r}).$$

The atom will also absorb on average a power given by [54, 55]

$$P_{\text{abs}} = \langle \vec{\epsilon} \cdot \dot{\vec{D}} \rangle = \frac{\epsilon^2}{2} \omega \operatorname{Im} \chi(\omega) = \frac{\omega}{\varepsilon_0 c} \operatorname{Im} \chi(\omega) I.$$

If we think at the laser beam as a stream of photons each with energy $\hbar\omega$, the absorbed power corresponds to a scattering rate of the incoming photons Γ_{sc}

$$\Gamma_{\text{sc}} = \frac{P_{\text{abs}}}{\hbar\omega} = \frac{1}{\hbar \varepsilon_0 c} \operatorname{Im} \chi(\omega) I.$$

At the steady state this will be counterbalanced by an equivalent emitted power accounted for by the scattered photons[4]. The cross section $\sigma(\omega)$ for this process is readily obtained as

$$\sigma(\omega) = \Gamma_{sc} \frac{\hbar\omega}{I} = \frac{\omega}{\varepsilon_0 c} \operatorname{Im} \chi(\omega),$$

[3] Assuming that the atom moves slowly enough to treat the electron dynamics adiabatically.

[4] For the Lorenz polarizability in Eq. (2.9), this is only true for a two level system, strictly speaking [48]. In general, the Lorenz model does not satisfy the optical theorem we used implicitly, but in practice if $\omega_{0n} \gg f_{0n}\Gamma_{n0}$ the deviation is not so large.

which when evaluated at resonance ω_0 for a two level system yields the same result as it can be obtained from a fully quantum mechanical treatment [48]

$$\sigma_0 = 3\frac{\lambda_0^2}{2\pi}.$$

As the atom emits photons isotropically, while the incoming photons all have momentum $\hbar\vec{k}$, there is a net rate of momentum transfer to the atom. This effect is known as *radiation pressure*, corresponding to a force $\vec{F}_{\text{sc}}$

$$\vec{F}_{\text{sc}} = \frac{d\vec{P}}{dt} = \Gamma_{\text{sc}}\hbar\vec{k} = \frac{\vec{k}}{\varepsilon_0 c}\,\text{Im}\,\chi(\omega)I.$$

A useful quantity that is typically introduced in the fully quantum mechanical treatment of the problem is the *saturation intensity* I_{sat}, which we define here as

$$I_{\text{sat},n} = \frac{\hbar\omega_{n0}m\varepsilon_0 c\Gamma_{n0}^2}{2e^2 f_{n0}}.$$

Since for $I \sim I_{\text{sat}}$ the saturation effects become important [48], the semiclassical model we are discussing is only valid as long as the light intensity is much smaller than this scale $I \ll I_{\text{sat}}$ or if we are far away from any resonance $|\omega - \omega_{n0}| \gg \Gamma_{n0}$. We can write simplified expressions for U_{dip} and Γ_{sc} assuming that we are close to a single resonance ω_0 so that we can ignore all the others. By introducing the detuning $\Delta = \omega - \omega_0$, which we assume to satisfy $\omega_0 \gg \Delta \gg \Gamma$ to also neglect dissipation and counter-rotating terms, we can write

$$U_{\text{dip}}(\vec{r}) \approx \frac{\hbar\Gamma^2}{8\Delta}\frac{I(\vec{r})}{I_{\text{sat}}}, \quad \Gamma_{\text{sc}} \approx \frac{\Gamma^3}{8\Delta^2}\frac{I}{I_{\text{sat}}} = \frac{\Gamma}{\hbar\Delta}U_{\text{dip}}.$$

These expressions allows us to gain several insights: first of all the sign of the detuning Δ determines the sign of the potential. This leads to repulsive potentials for blue detuned light $\Delta > 0$ and attractive potentials for red-detuned light $\Delta < 0$. Moreover, by choosing appropriately large detunings, it is typically possible to achieve very low scattering rates $\Gamma_{\text{sc}} \ll \Gamma$, so that atom heating is negligible. As an example, let us consider ^{87}Rb and far-detuned laser with $\lambda \sim 660$nm (so that $\Delta \sim 2\pi \times 70$ THz with respect to the D-line) and a spectral bandwidth of 30 GHz FWHM, corresponding to $\Gamma \sim 2\pi \times 13$ GHz [14]. Then for $U_{\text{dip}}/\hbar \sim 2\pi \times 1$ kHz scattering rates around $\Gamma_{\text{sc}} \sim 2\pi \times 0.2$ mHz are expected, leading to only 2 scattering events per second in a cloud of 10^4 atoms.

2.2 Experiments with cold atoms

Now that the basic physical concepts concerning alkali atoms and how they interact with electromagnetic fields have been covered, we can move on to comment on the structure and operation of cold atom platforms, also focusing on the setups which are most relevant for this thesis. Cold atoms experiments vastly differ in design depending on the physical models and regimes that are meant to explore, which translates in varying characteristics in terms of atom species and number, temperature, geometries, control and measurement fields. Nevertheless, they typically share the same experimental cycle, namely:

1. emission,
2. cooling and trapping,
3. manipulation,
4. measurement.

During the first phase, a hot beam of atoms is usually created by heating up a macroscopic sample of alkali metal, leading the most energetic tail of the atoms to evaporate. By enclosing the sample in an oven with an opening that acts as a collimator, a hot beam is created, where the kinetic energy is mostly due to the directed motion of the atoms along the beam axis.

In the second phase, the atoms in the beam are slowed and cooled down until they can be trapped in an external potential. This can be achieved by means of the absorptive forces generated by resonant light, using a wide range of techniques collectively known as *laser cooling* [4]. In the same experiment there are typically several cooling stages successively dealing with lower and lower energies, which can bring the atoms down to roughly the single photon *recoil temperature*:

$$k_B T_R = \frac{(\hbar k)^2}{2m},$$

in the range of $T \sim 0.1 \div 1\,\mu\mathrm{K}$ for most alkali atoms. In the case of ultracold gases, a final stage of *evaporative cooling* [56] is required to reach the range of $T \sim 10 \div 100\,\mathrm{nK}$, which allows to observe macroscopic population of the ground state or Bose-Einstein Condensation for most alkali species. Atom capturing and cooling is the most time-consuming stage, taking up to $\sim 10\,\mathrm{s}$ starting from hot atoms [57]. Nevertheless, state-of-the-art experiments can load pre-cooled atoms from a magneto-optical trap in around $\sim 100\,\mathrm{ms}$ [37], while evaporative cooling stages, whenever necessary, can last several seconds.

Once the atoms have reached the desired temperature, resonant driving for cooling is turned off, leaving only conservative potentials for confinement, the most typical choices being optical dipole potentials and magnetic traps (see Sec. 2.1). By controlling these traps it is possible to manipulate the motional state of the atoms, while operations on the electronic states are performed using e.g. Raman pulses or interactions between the atoms. The latter can be controlled by using Feshbach resonances to tune the scattering length [19] or by exploiting the long-range Van der Waals couplings typical of Rydberg states [58, 27]. While typical time scales for the application of individual gates are in the μs range [11], manipulation sequences can take up a consistent fraction of overall cycle time, lasting up to 500 ms [57].

Finally, measurements are performed on the system, using a selection of readout techniques — which predominantly involve the imaging of the atom distribution — which mainly differ in whether atoms are lost in the process or not. For instance, during time of flight measurements the atoms are released from the trap and let evolve ballistically which inevitably leads to the loss of the entire cloud [57, 14]. Absorption measurements, since they rely on measuring the attenuation of strongly resonant light as it passes through the atom cloud, typically lead to atom heating and losses because of photon recoil [59]. On the other hand, less destructive techniques such fluorescence imaging — where atoms are cooled by resonant light while the scattered photons are collected [7, 24]— and phase-contrast dispersion imaging — which instead avoids losses by employing off-resonant light— can be applied in-situ to obtain information about the system with smaller loss or heating rates, thus avoiding the necessity of reloading and cooling new atoms at each cycle. In any case, the readout phase takes much longer than individual gates, lasting 500 μs in optimized settings [11, 38].

2.2.1 Trapping setups

In Chapters 4 and 5 of this thesis we will be mainly concerned with the manipulation and control of the motional state of cold atoms by means of conservative potentials. It is then useful to shortly review the trapping setups which are most relevant for our purposes, which can be categorized based on whether they employ magnetic or optical confinement.

Magnetic traps employ the Zeeman shift caused by an inhomogeneous static magnetic field (see Sec. 2.1.1 for details) that results in a position-dependent energy shift. For instance, in Ioffe-Pritchard traps, the field intensity displays a minimum around which it increases quadratically, giving rise to a harmonic potential that can confine low-field seeking states, which can be prepared e.g.

using optical pumping. The confining magnetic fields can be generated by coils or by an integrated circuit, in a setup known as atom chip [60, 17]. The methods we discuss in Chapter 4 are designed for this kind of experiment, where an ultracold gas is obtained by forced RF evaporative cooling in a magnetic trap created by atom chip, and then further manipulated using optical dipole potentials [14].

This brings us to optical traps, which instead employ the AC Stark shift induced by laser light (see Sec. 2.1.2 for details) to create a potential which is proportional to the local light intensity. Moreover, the sign of the detuning allows to realize both repulsive ($\Delta > 0$) and attractive ($\Delta < 0$) potentials. Among the most common setups for creating optical dipole traps we find *optical lattices*, where the interference between split and/or reflected laser beams creates a standing wave, resulting in a periodic potential. By selecting the number and orientation of the different beams, it is possible to create a wide range of lattice geometries, confining the atoms in 1,2 or 3-dimensional arrays [23, 24, 30]. Another widely adopted setup consists in creating individual microtraps by focusing a red detuned gaussian beam, a technique known as *optical tweezers*. Arrays of such microtraps can be generated and controlled by acousto-optic modulators (AOM) and spatial light modulators (SLM), giving rise to a higher degree of controllability, even in real time, of the position of the individual atoms [61, 9, 10, 18]. In Chapter 5 we will tackle the atom transport problem for one of such setups, where a static optical tweezers array is combined with sorting tweezers.

While these two techniques focus on the realization of ordered arrays of atoms, it is also possible to use SLM to shape arbitrary potentials. These have been employed in experiments with ultracold gasses, where they can be used to create uniform box potentials by smoothing out trap inhomogeneities or to generate complicated patterns [16, 62]. Prospectively, they can also be used to generate site-dependent potentials to address single sites in atom lattices. In Chapter 4, we consider a setup that employs a Digital Micromirror Device, which is an array of micromirrors whose orientation can be individually programmed, acting as a reflective binary SLM. We will then see how to design and control a DMD optical setup in order to shape arbitrary optical potentials for an atom chip experiment.

2.2.2 Simulating spin chain Hamiltonians

While until now we mostly dealt with the motional degree of freedom of trapped atoms, the results and methods we develop in Chapter 6 will be tested on control problems concerning spin chain Hamiltonians, whose implementation within cold atom platforms typically involves the electronic degrees of freedom. Given the importance that this kind of Hamiltonians play in the study of quantum

magnetism, a wide variety of implementations have been proposed and realized experimentally. Without the ambition of reviewing in depth such a vast field, we deem it useful to briefly comment on the most prominent ones.

For instance, in the context of Rydberg atoms experiments, the spin state can be encoded in the ground $|\downarrow\rangle = |g\rangle$ and highly excited Rydberg states $|\uparrow\rangle = |r\rangle$. Then, resonant laser driving with Rabi frequency Ω and detuning Δ, together with long range Van der Waals interactions, gives rise to an Ising-type Hamiltonian [27, 18] of the form:

$$H_{Ryd} = \frac{\hbar\Omega}{2}\sum_i \sigma_x^{(i)} - \hbar\Delta\sum_i n^{(i)} + \sum_{i<j}\frac{C_6(\theta)}{|\vec{r}_i - \vec{r}_j|^6} n^{(i)}n^{(j)},$$

where $n^{(i)} = (1+\sigma_z^{(i)})/2$ and θ represents the angle between $\vec{r}_i$ and $\vec{r}_j$. Thus, the interactions can be tuned by rearranging the geometry of the lattice, which can be done conveniently using reconfigurable arrays of optical tweezers.

In the case of optical lattices, the mapping is more involved, as these setups naturally realize Bose- or Fermi-Hubbard Hamiltonians [21, 23, 63] involving nearest neighbour hopping t and on-site interaction U:

$$H_{Hub} = -\sum_\sigma t_\sigma \sum_{<i,j>}(c_\sigma^{(i)\dagger}c_\sigma^{(j)} + h.c.) + \sum_{\sigma,\sigma'} U_{\sigma\sigma'}\sum_i n_\sigma^{(i)}n_{\sigma'}^{(i)},$$

where $\sigma = \uparrow,\downarrow$ is an electronic state index mapped to a pseudo-spin (e.g. encoded in the sublevels of a hyperfine groundstate), $c_\sigma^{(i)}$ is the (bosonic or fermionic) destruction operator for an atom in lattice site i and state σ and $n_\sigma^{(i)} = c_\sigma^{(i)\dagger}c_\sigma^{(i)}$ the corresponding number operator. Nearest neighbours interactions can then be implemented by means of the superexchange interaction, which arises thanks to second order tunneling and on-site interaction, giving rise to a term with energy $\sim t^2/U$ [64]. In combination with inhomogeneous external electromagnetic fields and resonant driving between the electronic levels, a wide range of effective spin chain models can be realized, using bosons [25, 24] and fermions [65]. Alternative schemes involve encoding the pseudo-spins in occupation number imbalance between different sites in tilted optical lattices [66, 26].

Chapter 3

Quantum firmware: characterization, calibration and control

In this chapter, we set the stage for the main results that will follow in Part II of this thesis, by establishing a common context for them as steps towards the construction of a *quantum firmware* [39] for cold atom platforms – a set of methods forming a layer that interfaces the hardware described in the previous Chapter with the quantum software designed to be run on it. The main aim of the quantum firmware is therefore to implement in a scalable and reproducible fashion the basic operations that build up a program, and to do it as efficiently and precisely as possible. In order to achieve this, the quantum firmware must typically fulfill three tasks, namely, the characterization, calibration and control of the quantum system [40, 41].

In abstract terms, we can think at the experiment or platform as a physical system realizing some idealized model that we want to simulate or to implement some computation with. The first task of the firmware is then to characterize the system, that is to match system and model as close as possible by properly choosing the parameters of the latter as to minimize systematic errors. Moreover, the inputs and outputs of the hardware are typically not equivalent to the ones of the model, but are instead related through some (possibly non-linear) transfer function, which also needs to be characterized. Another task of the firmware is then to calibrate the system by establishing this relation between hardware and model I/O. Finally, by using this information, the firmware must be able to translate basic operations on the model into (typically time dependent) hardware inputs, also called control pulses, that realize the desired operation efficiently.

The development of this kind of firmware has to deal with several challenges,

some of which are general, while others are specific to AMO platforms and cold atoms. For instance, all quantum systems share physical limits to observability due to measurement back-action. Because of this fact, continuous feedback methods, which are widely adopted in the control of classical systems [67], are typically of limited applicability in the context of quantum systems. Here, we have to resort to either feed-forward methods, where the target quantities are computed offline using the previously determined model of the system and of the transfer functions [68], or to iterative feedback methods where the measurements at the end of the experimental cycle are used to correct the parameters of the next iteration, until convergence is reached [69]. In the context of cold atom platforms, these challenges come together with typically low repetition rates, which limit the amount of data that can be acquired from the experiment in a convenient amount of time. It follows that minimizing the data requirements for fulfilling the three tasks described above will be one of our main concerns. Bearing these considerations in mind, we can move on to introduce some of the tools and concepts that are going to be used throughout Part II.

3.1 Regression

As pointed out before, the characterization and calibration of quantum hardware requires matching a model with the physical system at hand, including the transfer functions, on the basis of experimental data and prior information about the system. Moreover, it is essential for feed-forward control to have access to such a model of the system in order to compute the control pulses without experimental feedback. This matching problem can be conveniently set in the language of regression [70]: let $\vec{f} : \mathbb{R}^m \mapsto \mathbb{R}^n$ be the function representing the relation between inputs $\vec{x} \in \mathbb{R}^m$ and outputs $\mathbb{R}^n \ni \vec{y} = \vec{f}(\vec{x})$. As an example, $\vec{f}$ could be the map between the controls of the quantum system and the fidelity (so that $n = 1$) of the evolved state with respect to some target state as in Chapter 6, or the transfer function mapping hardware inputs to the inputs of the quantum system, as we will see in Chapter 4. By operating the platform, we gather a training data set $\mathcal{S}_{\text{train}} = \{(\vec{x}_i, \vec{f}(\vec{x}_i))\}_{i=1}^{N_{\text{train}}}$ composed of I/O pairs. We then choose a parametrized family of functions $\vec{f}_{\vec{\alpha}} : \mathbb{R}^m \mapsto \mathbb{R}^n$, with parameters $\vec{\alpha} \in \mathbb{R}^p$ that constitute our model of the system. In order to find the parameter value $\vec{\alpha}^*$ that best fits the data set, we define a loss function L, that measures the deviation of the predicted outputs $\vec{f}_{\vec{\alpha}}(\vec{x}_i)$ from the measured ones $\vec{y}_i = \vec{f}(\vec{x}_i)$ and minimize it with respect

to $\vec{\alpha}$. A common choice is a loss of the form

$$L_{\beta,\gamma}(\vec{\alpha}, \mathcal{S}_{\text{train}}) = \sum_{i=1}^{N_{\text{train}}} ||\vec{f}_{\vec{\alpha}}(\vec{x}_i) - \vec{y}_i||_\gamma^\beta$$

which for $n = 1, \gamma = \beta = 2$ gives rise to the sum of square residuals, or L_2 loss, and for $n = 1, \gamma = \beta = 1$ to the sum of absolute residuals, or L_1 loss. Extra terms can be added to the loss, for instance to regularize the problem or to enforce sparsity.

Except for special cases such as linear models, loss minimization is performed numerically by means of gradient-based methods, such as L-BFGS [71], Adam or SGD to name just a few [72]. These methods offer far superior performance to gradient-free methods, and the computation of gradients can be implemented with minimal coding overhead thanks to automatic differentiation, which is nowadays available in software libraries for the most popular languages [73]. Moreover, methods like Adam and SGD are designed to deal with stochastic or noisy gradients, which allows us to compute the loss gradient over random subsets of the overall training data set. This can be very helpful in instances in which computing the full gradient would be too computationally expensive. In order to monitor and test the prediction performance of the model during and after training, it is necessary to employ data points which are independent of the training data set. Hence, we typically define a validation $\mathcal{S}_{\text{val}}$ and test $\mathcal{S}_{\text{test}}$ data sets alongside $\mathcal{S}_{\text{train}}$, over which we can evaluate error metrics (e.g. RMS error). For instance, validation loss or error can be used to trigger stopping conditions for gradient descent to avoid overfitting to the training data set.

Concerning the design of regression models for physical systems, we primarily face a trade-off between using widely employed models which benefit from a vast literature and thorough testing by the scientific community and custom models which are tailored to the specifics of the system at hand. The latter typically require more work for finding loss functions, hyperparameter values and initial conditions that lead to successful training, but can potentially lead to lower requirements in experimental data [43, 44], which is one of our main concerns. In Part II we will also explore this trade-off, first using a specialized model for transfer function learning in Chapter 4 to decrease training data requirements and then adapting well understood linear models to the study of quantum cost landscapes in Chapter 6, which allows us to simplify training and gain more insight into the mathematical structure of the problem.

We note that in this very brief discussion of the topic we are leaving many conceptual problems out of the picture, such as the possible presence of multiple local minima in parameter space or the bias of $\vec{f}_{\vec{\alpha}^*}$ as an estimator of $\vec{f}$. While

these are important details that can only be discussed in more specific cases, in realistic scenarios it can be very hard or impossible to have mathematical assurances in this regard because of the complexity of the models and loss functions or of the non-ideal statistical properties of the data set. Therefore, during our investigations, and especially in the case of custom regression models, we mostly follow a trial-and-error approach, tweaking the design of the regression problem until we obtain a result which satisfies our practical requirements.

3.2 Quantum Optimal Control

Once the hardware has been characterized and calibrated, we are left with the problem of implementing the elementary operations on the quantum system that serve as building blocks for the simulation or computation. These can be for instance quantum state preparation routines, unitary gates belonging to a universal gate set for digital quantum computing, time evolution under a given Hamiltonian for quantum simulation or other auxiliary operations. In Part II we are going to deal with two examples of these tasks, namely transporting atoms between different lattice sites in Chapter 5 and quantum state transfer in Chapter 6. Quantum Optimal Control (QOC) is a framework in which the implementation of a quantum routine is set as an optimization problem [74, 75]. This establishes a close parallel with our previous discussion on regression in Sec. 3.1, where a matching problem is also set in these terms. In this case the objective is to match an operation on the hardware, parametrized by the control pulse $u(t) : \mathbb{R} \mapsto \mathbb{R}^c$ to a target operation on the quantum system. Given a Hamiltonian depending on time through the control pulse $H[u(t)]$, its associated time evolution is given by the unitary operator

$$U[u(t)] = \mathcal{T} \exp\left(-\frac{i}{\hbar}\int_0^t d\tau H[u(\tau)]\right),$$

where $\mathcal{T}$ denotes time ordering in the exponential [76]. Once again, we first define a figure of merit that quantifies the deviation from a perfect match and then minimize it with respect to the control pulse. Often in QOC the figure of merit is chosen as the state infidelity I

$$I = 1 - |\langle\chi| U[u(t)] |\psi\rangle|^2,$$

which is appropriate for state transfer or preparation involving a target $|\chi\rangle$. If instead we want to quantify the deviation from a target unitary V we can use for

instance the average gate infidelity [77]

$$I_{\text{avg}} = 1 - \int d\psi \, |\langle\psi|V^\dagger U|\psi\rangle|^2 = 1 - \frac{|\text{Tr}(U^\dagger V)|^2 + d}{d(d+1)}.$$

Thus, the resulting functional $I[u(t)]$ is a quantum cost landscape that we minimize to find the optimal pulse $u^*(t)$ realizing the task.

Depending on whether the figure of merit is evaluated through experimental measurements or computed using a model, we can distinguish between closed- and open-loop control. The latter critically benefits from efficient gradient-based methods [68, 78], which can also be applied to parametrized pulses [79, 80] and even employ second order derivatives to speed up convergence [81, 82]. On the other hand, closed-loop optimal control allows us to treat the system as a black-box, avoiding the necessity of a precise model of the system and its transfer functions. This comes at the price of using gradient-free methods, which can only deal with a relatively small parameter search space and display slower convergence. Nevertheless, random parametrizations can be used to explore bandwidth limited pulse spaces while avoiding local traps in such settings [69, 83].

3.3 Shortcuts to Adiabaticity

Belonging to the category of feed-forward control methods, Shortcuts to Adiabaticity (STA) [84] take a different perspective on the problem. Instead of formulating quantum control as an optimization task, the aim of STA methods is to speed up adiabatic dynamics, which ensures the system, once prepared in an eigenstate $|n(0)\rangle$ at $t = 0$, to remain in the instantaneous eigenstate $|n(t)\rangle$ at all times thanks to the adiabatic theorem [76]. This theorem, which holds exactly only in the limit of infinitely slow temporal change of the Hamiltonian, can in practice be only applied provided that the following condition is satisfied:

$$\forall m, n \quad |R_{mn}(t)| := \left| \frac{\langle m(t)| \, \dot{H}(t) \, |n(t)\rangle}{E_m(t) - E_n(t)} \right| \ll 1. \tag{3.1}$$

In other terms, the Hamiltonian must change very little in the time scale defined by the inverse gap. Such adiabatic approach can as of itself be very useful, as it provides an experimentally simple way of preparing ground states of target Hamiltonians. Its main limitation is the potentially long time scales over which this process has to be applied once the gap becomes small, and its breakdown in case of gap closing [22, 69].

STA methods offer a way out of this shortcomings by decoupling the Hamil-

tonian driving the time evolution $\tilde{H}(t)$ from the one $H(t)$ whose instantaneous eigenstates we want the dynamics to follow. As the adiabatic condition Eq. (3.1) is not invoked, the quantity $R_{mn}(t)$ is not small anymore, and appears instead in the expression of $\tilde{H}(t)$ [85]:

$$\tilde{H}(t) = H(t) + \underbrace{i\hbar \sum_{m \neq n} \sum_{n} R_{mn}(t) \left|m(t)\right\rangle \left\langle n(t)\right|}_{H_{CD}(t)}. \tag{3.2}$$

The correction $H_{CD}(t)$ is known as counterdiabatic term, and has the precise purpose of canceling the deviations from adiabatic dynamics following $H(t)$. For this reason this method belonging to the STA family is known as *counterdiabatic driving* (or transitionless quantum driving) [86, 87]. An important drawback of this method is that it requires explicit knowledge of all the eigenstates of $H(t)$, which can limit its applicability in the context of many-body systems.

While counterdiabatic driving is a widely employed STA method, it is by no means the only way to realize fast adiabatic dynamics. Another strategy, typically referred to as *invariant-based engineering* is to exploit the existence of dynamical invariants of $H(t)$, which allows to reverse engineer quantum dynamics in such a way to realize state transfer tasks [88, 89]. More specifically, this technique requires that a (typically, hermitian) operator $I(t)$ exists which is constant in the Heisenberg picture:

$$\frac{dI}{dt} = \frac{\partial I}{\partial t} + \frac{1}{i\hbar}[I, H] = 0. \tag{3.3}$$

This in turn implies that the dynamics generated by $H(t)$ can be decomposed into dynamical modes, which can be easily computed from the eigenstates of $I(t)$. If moreover we have that

$$[I(0), H(0)] = [I(T), H(T)] = 0, \tag{3.4}$$

these dynamical modes will coincide with the energy eigenstates at the beginning and end of time evolution, ensuring transitionless transport between them [90]. One of the advantages of this technique is to not rely on explicit eigenstate knowledge, but on the other hand requiring the existence of a computationally simple invariant satisfying Eq. (3.3,3.4) limits its applicability to only specific Hamiltonians. In the context of this thesis, we are employing this technique in Chapter 5 to derive analytical solutions for the problem of transporting an atom between different lattice sites using an optical tweezer. This is possible thanks to a well known family of dynamical invariants for single particle systems [91].

Part II

Results and Publications

Chapter 4

Optimizing optical potentials with physics-inspired learning algorithms

---

Authors:	**Martino Calzavara**, Yevhenii Kuriatnikov, Andreas Deutschmann-Olek, Felix Motzoi, Sebastian Erne, Andreas Kugi, Tommaso Calarco, Jörg Schmiedmayer, and Maximilian Prüfer
Year:	2023
Journal:	Physical Review Applied
Volume/Issue:	19/4
DOI:	10.1103/PhysRevApplied.19.044090

4.1 Summary and context

The scope of this paper [92] falls within the characterization, calibration and control of the transfer function of an experiment with ultracold gasses. More specifically, we consider an experimental setup that can shape arbitrary one-dimensional optical dipole potentials (see Sec. 2.1.2) by means of a Digital Micromirror Device [14]. Thus, the non-linear transfer function that we are studying takes as input the configuration of the DMD and outputs the optical dipole potential generated by the setup. Shaping a potential is an output matching problem that requires finding an appropriate input DMD configuration by minimizing deviation from the target output.

While DMD-assisted beam shaping is an established technique in cold atom platforms [16, 93, 94], our target application is the precise shaping of arbitrary time-dependent potentials that would allow the implementation of open-loop optimal control or STA schemes (see Sec. 3.2 and Sec. 3.3) to one-dimensional ul-

tracold gasses [95, 87]. Since pulse sequences are realized by concatenating many static potential frames, it is essential for our methods to minimize the amount of online iterations required for each frame.

In order to achieve this, we develop a physics-inspired model of the transfer function, that we train by regression on experimental data. The optimization can therefore happen offline using the trained model, preventing the number of online iterations to scale with the number of frames. Here, we choose to develop a custom model that can be trained with rather small datasets, taking less than an hour of experimental time to be acquired. Moreover, we also develop an optimizer based on Iterative Learning Control (ILC) [96], that employs both physical information on the transfer function and iterative feedback to speed up convergence by an order of magnitude, dramatically cutting down the time and data requirements of the procedure.

In the context of this thesis, this paper shows how automated optimization and physical insight can be combined into efficient characterization, calibration and control methods for experimental transfer functions.

4.2 Author contribution

The Author is responsible for the ideation and development of the machine learning model of the optical system and contributed to the development of the ILC algorithm. He is also responsible for the development of most of the software concerned in this paper (including calibration, characterization, open- and closed-loop optimization) and of its deployment and tuning on the experiment. He contributed in generating the data for Fig. 3.b,c,d and Fig. 5, in the design of the Figures, and in the writing of all Sections.

Optimizing optical potentials with physics-inspired learning algorithms

M. Calzavara[1,4,*], Y. Kuriatnikov[2,*], A. Deutschmann-Olek[3], F. Motzoi[1], S. Erne[2], A. Kugi[3], T. Calarco[1,4], J. Schmiedmayer[2], M. Prüfer[2,§]

[1] *Forschungszentrum Jülich GmbH, Peter Grünberg Institute, Quantum Control (PGI-8), 52425 Jülich, Germany*

[2] *Vienna Center for Quantum Science and Technology, Atominstitut, TU Wien, Stadionallee 2, 1020 Vienna, Austria*

[3] *Automation and Control Institute, TU Wien, Gußhausstraße 27-29, 1040 Vienna, Austria*

[4] *Institute for Theoretical Physics, Universität zu Köln, 50937 Cologne, Germany*

We present our experimental and theoretical framework which combines a broadband superluminescent diode (SLD/SLED) with fast learning algorithms to provide speed and accuracy improvements for the optimization of 1D optical dipole potentials, here generated with a Digital Micromirror Device (DMD). To characterize the setup and potential speckle patterns arising from coherence, we compare the superluminescent diode to a single-mode laser by investigating interference properties. We employ Machine Learning (ML) tools to train a physics-inspired model acting as a digital twin of the optical system predicting the behavior of the optical apparatus including all its imperfections. Implementing an iterative algorithm based on Iterative Learning Control (ILC) we optimize optical potentials an order of magnitude faster than heuristic optimization methods. We compare iterative model-based "offline" optimization and experimental feedback-based "online" optimization. Our methods provide a route to fast optimization of optical potentials which is relevant for the dynamical manipulation of ultracold gases.

I. INTRODUCTION

The precise control and manipulation of light fields are required for many diverse areas of research ranging from microscopy [1] to quantum simulators [2]. In particular, optical beam shaping constitutes a common task, for which wavefront manipulating devices, such as Spatial Light Modulators (SLM) or Digital Micromirror Devices (DMD), are especially suited. The beam shaping is important for experiments with ultracold gases, where optical dipole potentials have proven to be a versatile tool to provide the demanded level of control. In combination with a DMD, almost arbitrary shaping of the optical potential is possible, both in 1D [3–6] and 2D [7–11] settings. These potentials can, as an example, be used for generating homogeneous box potentials in ultracold gases experiments [9]. In addition to static potentials, dynamical perturbations and time-averaged potentials can also be created, by projecting sequences of patterns onto the DMD [4, 12].

There exist two main approaches to the optimization of optical potentials: precalculating DMD patterns based on physical assumptions [13, 14] and models [7] or iteratively updating DMD patterns based on experimental feedback [6, 7]. The first approach avoids the need for feedback measurement but is limited by model precision and thus requires detailed system characterization, while the second gives the most accurate results but typically requires a large number of experimental iterations. Here we implement a purely data-driven approach that combines different learning algorithms to get the best of both worlds.

Using a digital twin of the system makes it possible to shape different types of target potentials without the need for experimental feedback. Yet because of speckles caused by imperfections, a model featuring just a few known experimental parameters (parametric model) can only predict its behavior up to limited precision and might not always be suitable for precise potential shaping. In some cases, though, such as in "clean" systems with pin-hole filtering, simulations combined with input beam characterization deliver very low errors on precalculated DMD patterns [7]. We improve the prediction performance compared to parametric models by employing data-driven learning techniques. Learning methods are already used for estimating the transfer matrix of complex optical systems [15] and provide good results. As their main disadvantage, they generally require a large amount of data for sufficient training. In our approach, we develop a physics-inspired model that requires a smaller amount of training data and thus saves experimental time.

Despite any improvement in model precision, the effect of residual error sources can only be mitigated by using experimental feedback [6, 7]. Therefore, we introduce a feedback optimization method based on Iterative Learning Control (ILC) [16, 17]. This method is directly applicable to various types of experiments with wavefront manipulating devices. Since system knowledge is directly employed in the update law that adjusts the DMD setting based on feedback, the number of required experimental iterations is significantly reduced compared to heuristic methods [6].

Because of the learning-based nature of both the ML model and ILC method, they benefit from high pre-

* These authors contributed equally to this work

§ maximilian.pruefer@tuwien.ac.at

dictability of the system behavior, which can be achieved by using light sources with low coherence. White light was used for Bose-Einstein condensates (BEC) trapping in order to minimize the impact of speckles on density fluctuations [18], and the advantages of using a superluminescent diode (SLD) in comparison with a single-frequency laser were shown also in combination with a DMD [12]. Indeed, we find that using the SLD improves model prediction results, while the feedback-based methods give errors comparable to measurement errors for both light sources.

In perspective, the ability to efficiently shape a large number of potentials using a DMD will provide a high level of control for the dynamical manipulation of quantum gases. The realization of non-harmonic optimal protocols [19] and quantum thermal machines [20] are but two examples among countless possible applications.

The paper is organized as follows: in Sec.II we describe our experimental setup and test the coherence properties of the light sources by measuring interferences of a SLD in comparison with a single-frequency laser; in Sec.III we introduce the physics-inspired learning model used for representing the optical system including its imperfections; in Sec.IV we use iterative learning control algorithms for optimizing optical dipole potentials; in Sec.V we summarize our results.

II. EXPERIMENTAL SETUP

In this section, we describe our experimental setup and compare the optical coherence properties of a single-frequency laser and broadband SLD. The optical apparatus was designed and optimized for manipulating 1D optical dipole potentials in our atom chip experiment [21]. The simplified optical setup is shown in Fig.1 (a). In the setup, the light source is easily interchangeable: we here use either a superluminescent 110 mW fiber-coupled diode with a spectral width of 12.7 nm and a central wavelength of 837 nm or a single-frequency 780 nm laser. The fiber is connected to a collimator and illuminates the DMD with collimated light. The DMD is placed in the focus of the first lens. The first two lenses together form a 4-f telescope. The focal point between the two lenses is a Fourier plane with respect to the DMD's image plane. The slit, which is adjustable in the transverse direction and is parallel to the observable 1D optical potential, is placed in the Fourier plane and closed to 0.625 mm. Being placed in the Fourier plane, the slit acts as a low-pass filter for the transverse spatial frequencies of the light field. This means it cuts off high-frequency k-modes of the DMD pattern effectively leading to a lowering of the resolution. The other 4 lenses form an objective designed to correct chromatic aberrations for the broadband SLD. The system acts with a demagnification of 17.5 and resolution $\sigma_{\parallel} = 3.3\,\mu$m in the direction parallel to the slit (the resolution is given by our atom chip experiment [21] for which the optical apparatus was designed and optimized)

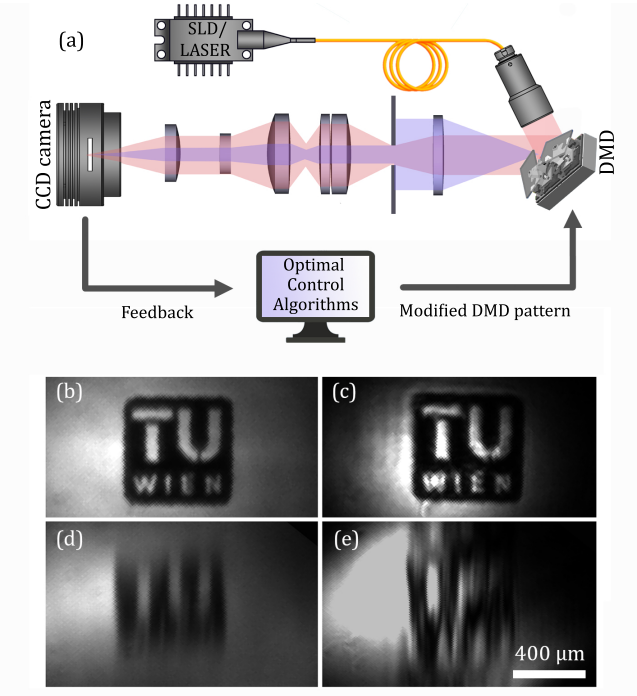

FIG. 1. **(a)** Schematic of the experimental setup. The light source is an interchangeable SLD/laser connected with a fiber to the collimator and projected onto the DMD. The DMD is an array of individually controllable micromirrors which allows for projecting arbitrary patterns. The system is shown as simplified schematic omitting mirrors. The shown lens system constitutes an imaging of the DMD pixels into the CCD camera (image) plane with a demagnification of 17.5 and minimizing chromatic aberrations. The slit is placed in an effective Fourier plane of the imaging system. The red path corresponds to the beam and the blue to a point source in the DMD plane. **(b)-(e)** Images taken with the CCD camera in the image plane of the DMD with the same pattern projected for **(b)** SLD without slit **(c)** laser without slit **(d)** SLD with slit **(e)** laser with slit.

and $\sigma_{\perp} = 25\ \mu m$ in the transverse direction (orthogonal to the slit). In the end, an image of the optical potential is acquired with a CCD camera sensor with 2.4 μm pixel size.

We use a 10.8 μm pitch near-infrared DMD, which is a 2D array of 1280x800 micromirrors. In the image plane (on the CCD camera) 5 DMD pixels correspond to $\sigma_{\parallel}$ in the longitudinal direction and 40 DMD pixels correspond to $\sigma_{\perp}$ in the transverse direction. Since many pixels in the transverse direction contribute to the local intensity (close to the center pixels contribute almost equally to the optical potential while far off-center pixels are contributing less), we can perform very smooth gray scaling, which is very important for high-precision optimization of 1D optical dipole potentials via DMD. Yet, while a narrower slit allows for better gray scaling and with that lower discretization error and higher accuracy, it leads to significantly reduced light intensity.

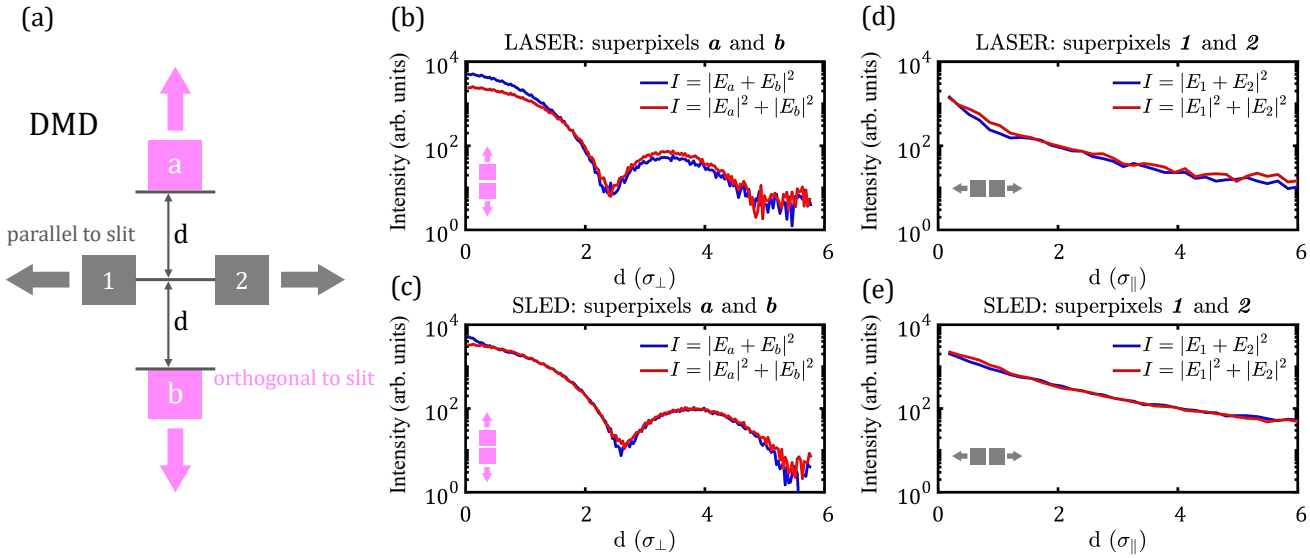

FIG. 2. **(a)** Schematics of test measurement of coherence properties. Two pairs of 3x3 DMD superpixels **a/b** and **1/2** are translated in a direction orthogonal and parallel to the slit, respectively. The resolution of the optical system is $\sigma_\parallel = 3.3\,\mu$m in the longitudinal (parallel to slit) and $\sigma_\perp = 25\,\mu$m in the transverse (orthogonal to slit) direction, with respect to the slit orientation. **(b)-(e)** Intensity measured between two superpixels as a function of distance d between them; the blue solid line shows the intensity measured for both superpixels simultaneously "on" (electric fields from superpixels contribute in measured intensity as $|E_{a/1} + E_{b/2}|^2$), the red solid line is the sum of intensities measured with only one of two superpixels "on" (contribution is independent $|E_{a/1}|^2 + |E_{b/2}|^2$). The distance d is given in units of $\sigma_\parallel$ and $\sigma_\perp$. When the laser-illuminated superpixels are moved orthogonal to the slit **(b)**, positive interference is visible in the 0^{th} slit diffraction maximum and negative interference in the 1^{st}. When SLD-illuminated superpixels are moved orthogonal to the slit **(c)** a slight interference effect is visible only in the 0^{th} slit diffraction maximum for very small distances between the superpixels. Panel **(d)** shows laser-illuminated superpixels moving parallel to the slit, and panel **(e)** shows SLD-illuminated superpixels moving parallel to the slit. There are no noticeable interference effects in the longitudinal direction on the scales larger than the camera pixel size for both light sources.

The main advantage of using the SLD compared to the single-frequency laser as a light source is its reduced temporal coherence [22, 23] which is why random interferences (speckles) are reduced (see Fig.1 (b,c)). Here, we characterize coherence effects for both light sources as we want to understand whether we can consider the individual pixels as coherent or incoherent sources.

We prepare two series of patterns, each of them containing two superpixels moving away from each other (by superpixel, we understand a square pattern of 3x3 micromirrors). In the image plane on the CCD camera, the size of the superpixel is below the resolution of the optical apparatus and the superpixel can be considered as a point source, at the same time minimizing diffraction from the edges of individual DMD mirrors. We measure the intensity in the center between two superpixels. The imaged intensity is in general given by the squared sum of the electric fields of the individual pixels. In the incoherent case, it reduces to the sum of intensities from the individual pixels such that the system behaves linearly with respect to intensity.

First, the pair of superpixels is moving orthogonal to the slit orientation (transverse direction), where the slit is closed to 0.625 mm and resolution is lowered to $\sigma_\perp$=25 μm. For both SLD and laser (see Fig.2 (b,c)), the slit diffraction maxima are clearly distinguishable. For the laser, positive interference can be observed in the 0^{th} maximum and negative in the 1^{st} one. For the SLD, interference is only observable on a very small scale (below $\sigma_\perp$). Second, the superpixels are moving in the longitudinal direction, where the slit is fully opened to 13 mm and the resolution $\sigma_\parallel$=3.3 μm is very close to the camera pixel size 2.4 μm. We observe only small interference effects for the laser, most likely because they are hardly accessible on this scale.

In conclusion, we observe that in both cases the SLD shows linear behavior, that is the total intensity is given by the incoherent sum of the intensities of each point source. This showcases the advantages of the reduced coherence of the SLD, which we anticipate to decrease the amount of speckles and therefore to increase the predictability of the system.

III. PHYSICS-INSPIRED LEARNING MODEL

In this section, we describe how we use Machine Learning (ML) tools and experimental data to obtain a digital

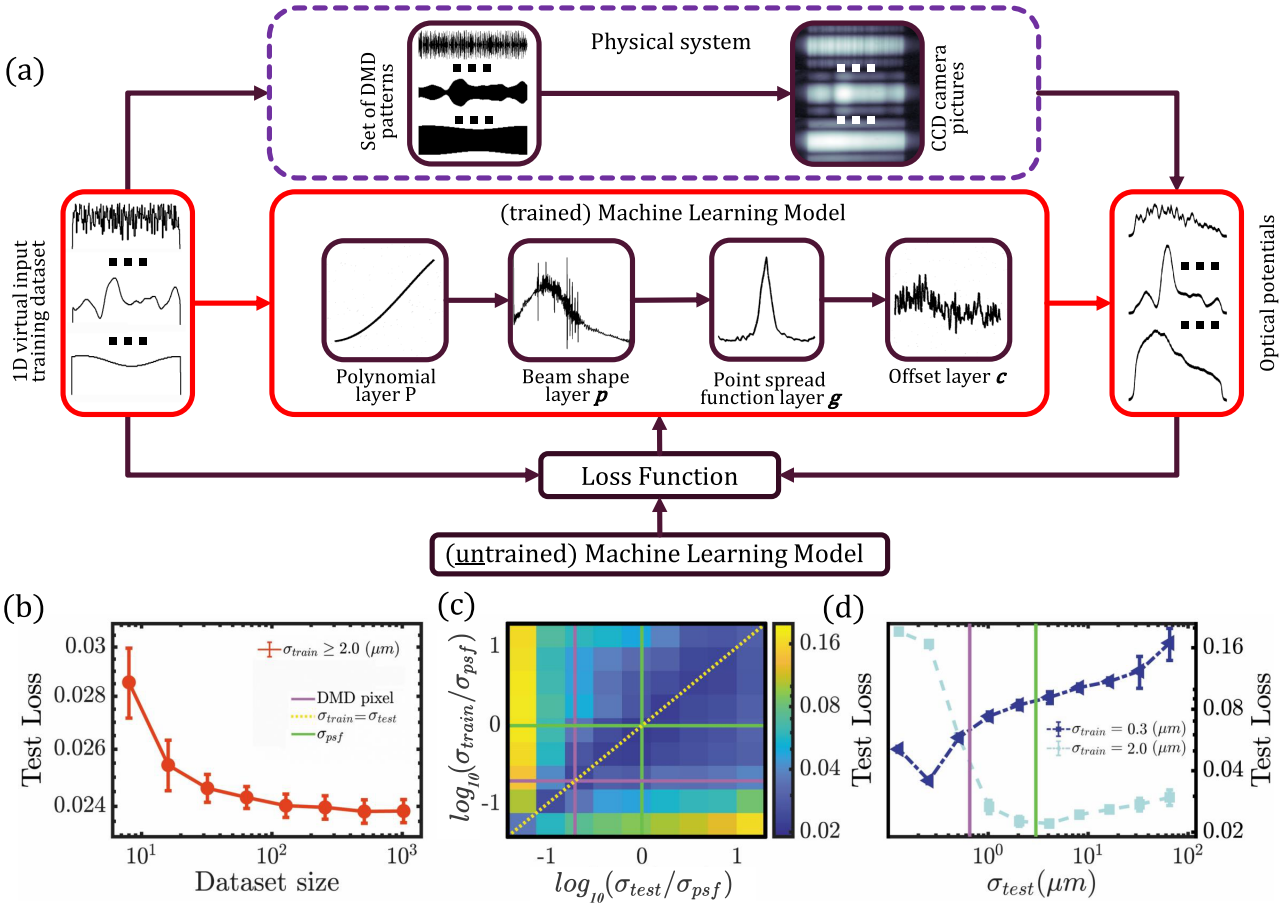

FIG. 3. **(a)** Training of the physics-inspired model. The training dataset is composed of 1D virtual inputs $\boldsymbol{\nu}$ which are transformed into 2D DMD patterns (according to Eq. (2)) and projected in the image plane. We read out a pixel row from the CCD camera to infer the 1D optical potential generated by the system. After data acquisition, the 4-layer model is trained by minimizing the loss function. The polynomial layer represents the nonlinear relation between virtual input and local grayscale, while the other three layers take into account physical properties, i.e. the beam shape, the effective point spread function, and the background. The trained model (red solid segment) is a digital twin of the physical system which converts 1D virtual input to predicted optical potential. All the signals and layers shown in the figure are actual experimental data. **(b)** Test loss dependence on the size of the training dataset. The loss converges to 2.4% for a dataset size of $\sim 10^2$, indicating that the model does not exhibit overfitting even for small training datasets. **(c)** Test loss dependence on the spatial frequency content of training and test datasets (see main text for details). **(d)** Two cuts from figure (c) for $\sigma_{train} = 0.3\ \mu m$ and $\sigma_{train} = 2.0\ \mu m$.

twin of the system. It should represent precisely the experimental system, while at the same time having a small number of parameters to reduce the required number of experimental measurements. In the following, we outline our approach based on a physics-inspired model.

To formulate our problem, we treat the DMD as a 2D array to which we associate a binary configuration matrix u_{ij}. For each pixel, a value of 1 corresponds to the mirror position that reflects light to the optical system, while 0 corresponds to the mirror position that reflects light away from the system. The potential $\boldsymbol{V}$ is a vector obtained by selecting one row of camera pixels in the imaged output. We refer to the coordinate along this row (and therefore, parallel to the slit) as z. We then cast model training in the language of regression [24]: given a set of data points composed by K couples $(u_{ij}^{(k)}, V_i^{(k)})$, and a family of functions $\mathcal{M}_{\boldsymbol{\alpha}}$, parametrized by $\boldsymbol{\alpha} = [\alpha_1, ..., \alpha_N]$, we find the values of $\boldsymbol{\alpha}$ that minimize a loss function $L(u_{ij}^{(k)}, V_i^{(k)}, \boldsymbol{\alpha})$ defined below.

Regression problems are known to be affected by the bias-variance tradeoff [25]. This implies that large models, having $\dim(\boldsymbol{\alpha}) >> K \dim(\boldsymbol{V})$, tend to approximate well the training data (low bias) but fail to accurately generalize the prediction on test data (high variance). This phenomenon is known as overfitting, and it puts a challenging limitation to our program. In order to alleviate this problem, our approach consists in developing a physics-inspired model (see Fig.3 (a)) based on the knowledge we have about optical systems. This way, we can reduce the number of its coefficients to the minimum

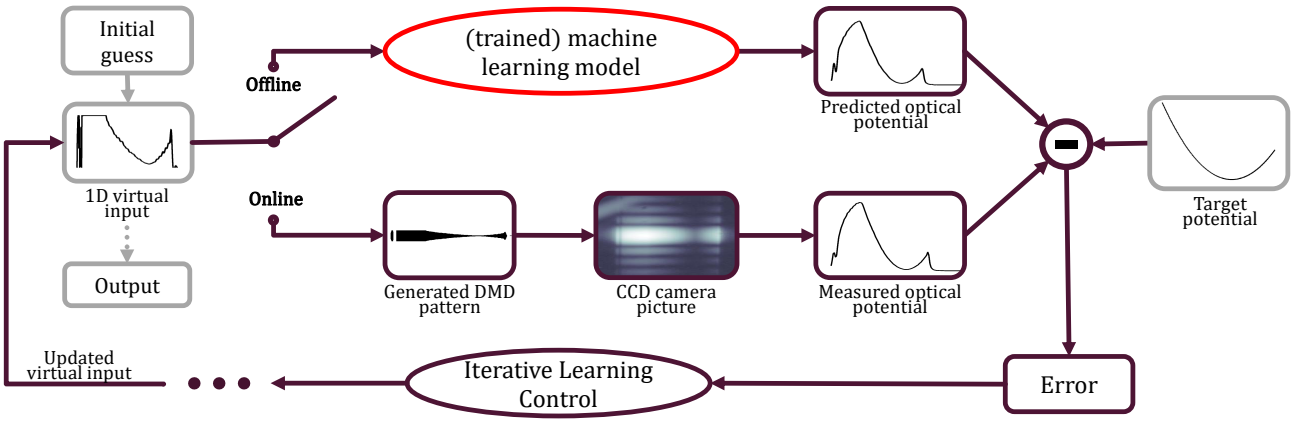

FIG. 4. Scheme for optimizing optical potentials using Iterative Learning Control. The iterative scheme is initialized by choosing an initial guess and the target potential. We use two methods that differ in the feedback source: offline and online. The offline method uses the trained ML model to predict the optical potential. The online method projects the pattern on the DMD and measures the optical potential. Then the predicted (measured) potential is subtracted from the target to get the error that ILC uses to update the 1D virtual input. The procedure is repeated iteratively until the error converges or matches conditions for loop break. For the best performance, the output of the offline optimization can be used as the initial guess for online optimization. The result of the optimization loop is an optimized DMD pattern. All the signals in the figure are actual experimental data.

necessary to represent the system precisely enough, while avoiding overfitting.

First, we select an area of interest (AOI) on the DMD where the patterns will be located, based on the beam size and position. All pixels outside the AOI are turned off. The AOI has a size of (N_{row}, N_{col}), and its columns are orthogonal with respect to the slit. Then, we employ a dimensional reduction technique which we refer to as the "virtual input" (cf. [17]). The virtual input measures the relative optical intensity induced by the pixels in the corresponding column. Since the narrow slit averages along the transverse direction, we can use the pixels on each column to create a set of gray scales at each point along the longitudinal direction z. In the limit of a very narrow slit, all the pixels in a column contribute almost equally to the final intensity, so we can define the virtual input $\boldsymbol{\nu}$ as the vector:

$$\nu_j = \frac{1}{N_{row}} \sum_i u_{ij} \,. \tag{1}$$

This quantity represents the fraction of pixels that are turned on in each column, with respect to the total number of rows inside the AOI.

In the case of non-zero slit widths, the mapping between a virtual input according to Eq. (1) and the actual relative optical intensity due to the column pattern is non-linear and depends on the beam and on the transverse shape of the point spread function (PSF). Therefore, we have to take this effect into account when we design our model.

Once u_{ij} is transformed into ν_j some information is lost since different binary matrices map onto the same virtual input. To invert this mapping we have to choose a subset of binary matrices over which the mapping is one-to-one. We define the inverse map by turning on the pixels one by one on alternating rows above and below the central row of the AOI i_c, according to the order $i_c, i_c+1, i_c-1, i_c+2, i_c-2, ...$ More compactly:

$$u_{ij} = \theta\left(\frac{N_{row}\nu_j}{2} - |i - i_c - \frac{1}{4}|\right) , \tag{2}$$

where $\theta(x)$ is the Heaviside step function.

Using the virtual input and restricting the DMD configurations to the set described by Eq.(2) reduces the dimension of the input and output space, thus simplifying the learning problem. On the other hand, the restriction of the configurations might potentially introduce an artificial limitation to the potentials that can be realized. In practice, we find the subset of binary matrices defined by this mapping to be wide enough so that it can be used for shaping arbitrary potentials.

For the task of learning the relation between the 1D input ν and the 1D potential $\boldsymbol{V}$ we propose the model $\boldsymbol{V} = \mathcal{M}_{\boldsymbol{\alpha}}(\boldsymbol{\nu})$ with (learnable) parameters $\boldsymbol{\alpha}$. This model is depicted in Fig.3 (a) and explicitly reads:

$$V_i = \left| \sum_{j=-M}^{M} g_j P(\nu_{i-j}, q_1, ..., q_{N_P}) p_{i-j} \right| + c_i \tag{3}$$

where $P(x, q_1, ..., q_{N_P}) = \sum_{n=1}^{N_P} q_n x^n$ is a polynomial of degree $N_P = 5$ with no constant term, and the vector of parameters $\boldsymbol{\alpha}$ is the concatenation of $\boldsymbol{g}$,$\boldsymbol{q}$,$\boldsymbol{p}$ and $\boldsymbol{c}$ with $\dim(\boldsymbol{\alpha}) \sim 1400$. The polynomial function P represents in an abstract way the non-linear relation between virtual input and resulting local relative intensity as discussed

above. The remaining parameters are chosen to mimic features of the physical system: the convolutional kernel $\boldsymbol{g}$ of size $N_g = 2M + 1 \sim 71$ plays the role of the longitudinal PSF, the position-dependent term $\boldsymbol{p}$, of size $N_V \sim 650$, mimics the inhomogeneous light beam and the offset term $\boldsymbol{c}$, also of size N_V, gives the background. In the convolution sum we employ zero padding, which means that the summand is zero if $i-j$ exceeds the index range $[1, N_V]$.

For the training, we choose the loss function as:

$$L(\boldsymbol{\nu}^{(k)}, \mathbf{V}^{(k)}, \boldsymbol{\alpha}) = \sum_{k=1}^{K} \frac{\sum_{i=1}^{N_V} |[\mathcal{M}_{\boldsymbol{\alpha}}(\boldsymbol{\nu}^{(k)})]_i - V_i^{(k)}|}{\sum_{i=1}^{N_V} V_i^{(k)}}, \quad (4)$$

which is the mean absolute error between prediction and measurement, normalized by the average potential itself. This way, potentials with different average values will contribute equally during the training.

To test the performance of the ML model, we create a data set of 10,000 random virtual inputs by sampling a white noise probability distribution. We filter the sample with Gaussian filters with 10 values of σ_{data} ranging from 0.2 to 100 DMD pixels (which corresponds to the range $[0.1, 65]\mu$m mapped to the image plane). This way, the dataset was subdivided into 10 subsets with varying upper spatial cutoff frequency. For each virtual input, the corresponding potential was measured and stored (along with the corresponding input).

We tested the dependence of the test loss on the training dataset size (see Fig.3 (b)). Both the training and test datasets were assembled by mixing the subsets with $\sigma_{data} \geq 2.0$ μm. The model is then trained on data chunks of increasing size K_{train}, while keeping the test dataset size fixed to $K_{test} = 300$. This sequence is repeated 4 times, choosing new data sets at each time to compute the standard deviation of the test loss. Even with $K_{train} = 8$, the test loss is already below 3% and the improvement for $K_{train} \sim 10^2$ is around half a percent. Adding even more data points does not appreciably improve the prediction quality. This result is particularly important since we aim for methods that are readily applicable to atomic density data measurement. In a typical experiment with trapped ultracold atomic gases, taking more than $K \sim \mathcal{O}(10^2)$ atom densities pictures (with averaging over a few shots each) for potential optimization is a time-costly procedure. The fact that the model we developed can be trained with less than 100 potentials means that it is in principle possible to employ a data set obtained by atom-density estimation of the potential [6].

To further analyze the learning process, we tested the dependence of the test loss on the cutoff frequency (see Fig.3 (c,d)) of both the training and test data sets. The frequency subsets are not mixed. At each time, we choose independently the cutoff frequency of the training σ_{train} and test σ_{test} datasets. The dataset sizes for training $K_{train} = 160$ and for testing $K_{test} = 39$ are fixed, and this sequence is repeated 5 times. Fig.3 (c) shows qualitatively the test loss as a function of σ_{train} and σ_{test}. The best results are obtained close to the diagonal, that is, where $\sigma_{train} \sim \sigma_{test}$, so that training and test data look most similar. The test loss becomes worse when σ_{train} or σ_{test} are around or below the DMD pixel size (magenta line on the plot), indicating a possible mismatch between the behavior of the system and the ML model on the scale of the DMD pixel. For a more quantitative interpretation, we show two different regimes in Fig.3 (d), where a representative scenario of the training on low frequencies (dark blue) is compared to another curve representing the high-frequency case. In the first case, the test loss exhibits a low plateau above the DMD pixel size and an abrupt increase below, corresponding to a breakdown of the low-frequency trained model in the high-frequency regime. The second curve instead shows how including the high frequencies during the training does not solve the problem, as now the prediction performance severely deteriorates along the whole frequency range.

IV. POTENTIAL OPTIMIZATION

In this section, we introduce an optimization algorithm based on iterative learning control (ILC) methods inspired by [17]. The algorithm was used to optimize different target potentials on the experimental setup and we compared the results with the heuristic optimization method described in [6] (see Sec.VI for details).

A. Online and Offline ILC

ILC methods employ measurements of the considered output trajectory to iteratively solve a reference tracking problem, i.e., to find an input trajectory such that the output of a system follows a desired target trajectory as closely as possible, even in the presence of model errors and uncertainties. The price to pay for this property is the requirement of running in a kind of feedback loop using experimental data. Therefore, we show how to employ ILC algorithms either using feedback from the physical experiment referred to as "online" ILC or from its digital counterpart, the physics-inspired model presented in Sec.III, referred to as "offline" ILC. See Fig.4 for a schematic of the ILC algorithm. The only difference in the algorithms is that in the first case the potential $\boldsymbol{V}$ is measured, while in the second it is predicted by the model. This way, previous (training) data is structured by the ML model while we can seamlessly improve beyond the predictive capability of the model through further online iterations.

Let us call $\boldsymbol{\nu}^n$ the virtual input at the n-th iteration, and $\boldsymbol{e}^n = \boldsymbol{V}^{tar} - \boldsymbol{V}^n$ the deviation from the target. Following standard ILC approaches, the correction of the virtual input is obtained by convolution, denoted by $*$ (see Sec.VI), with an update filter $\boldsymbol{L}_\nu$, and is then added to the old virtual input to update it, i.e.,

$$\boldsymbol{\nu}^{n+1} = \boldsymbol{\nu}^n + \boldsymbol{L}_\nu * \boldsymbol{e}^n . \quad (5)$$

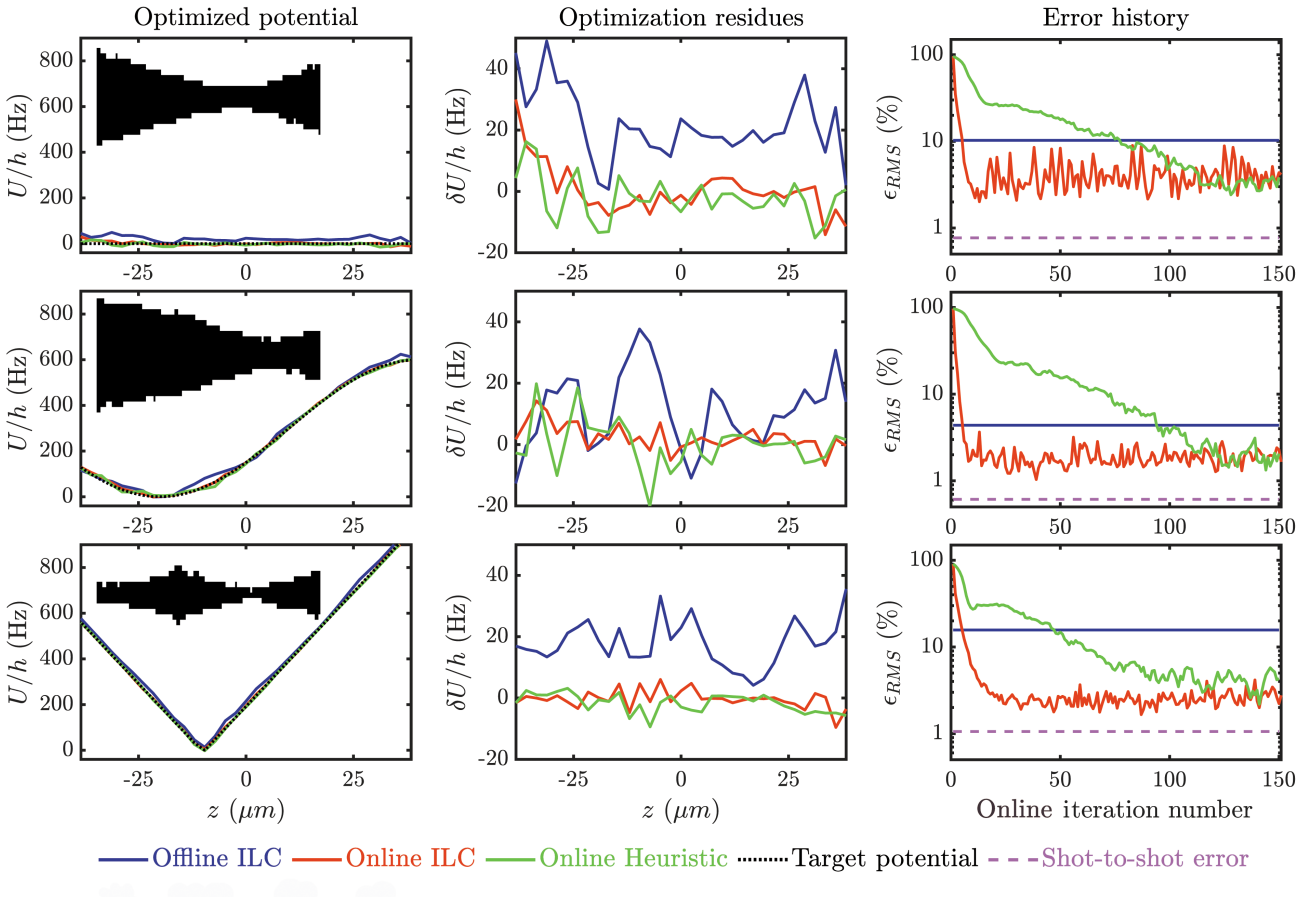

FIG. 5. Optimized effective potentials assuming a 10 Hz harmonic trap. The 1D optical potential generated by the system is inferred by reading out one selected row of the CCD. In the first column, we show the optimized effective potentials U with the offline ILC, online ILC, and the heuristic online optimization (blue, orange, and green solid lines, respectively). The black dotted line shows the effective target potential for atoms (see main text for details). The insets show DMD patterns for optimization with online ILC. All three methods give a good qualitative match with the target potentials. The second column depicts the differences δU between the target potential and the optimized potential for all 3 methods. The online ILC and the heuristic online optimization give a compatible level of optimization while the offline ILC exhibits a larger deviation from the target. In the third column, the error histories for online methods are shown in the semi-log scale, computed at each point using Eq.(8). The blue line indicates the level of offline optimization error, which is shown as a constant reference line since it requires no online iteration, showing that the offline prediction can save 50-90 heuristic iterations. The online ILC saves more than 100 iterations compared to the heuristic method to reach a similar error level. The dashed purple line shows the shot-to-shot measurement error.

The process is repeated until either convergence or the desired error level is reached. In order to choose an appropriate update filter, we approximate the system in a linear and time-invariant form

$$\boldsymbol{V} \approx \boldsymbol{g}_z * \boldsymbol{\nu}\,. \tag{6}$$

In practice, $\boldsymbol{V}$ is recorded for a trial input, and then fitted using a Gaussian guess for the longitudinal PSF $g_z(z) = A\exp(-(z-z_0)^2/(2\sigma^2))$, to obtain estimates for the parameters z_0, σ, A. We then use a pseudo-inversion-based update filter [26]

$$\boldsymbol{L}_\nu = \frac{\overline{\boldsymbol{G}}}{\gamma_\nu + \overline{\boldsymbol{G}}\boldsymbol{G}} \tag{7}$$

where $\boldsymbol{G} = \mathcal{F}[\boldsymbol{g}_z]$ is the discrete Fourier transform of the Gaussian PSF, $\overline{\boldsymbol{G}}$ its complex conjugate and all the operations are to be understood element-wise. Here, $\gamma_\nu > 0$ is a regularization parameter of the system inversion which effectively reduces the high-frequency content of the input updates and, therefore, of the explored virtual inputs $\boldsymbol{\nu}^n$. In the presence of measurement noise, optimal choices for γ_ν are ultimately given by the experimental signal-to-noise ratio [27]. The results shown in the main text are obtained using $\gamma_\nu = 0.1 \max_i |G_i|^2$.

B. Experimental results

We tested both the offline and online ILC algorithms by optimizing three target potentials, using the heuristic method as a reference. The results are shown in Fig.5. In order to emulate an experiment with a trapped BEC, we represent the longitudinal magnetic trap with a harmonic potential $V_{mag}(z) = \frac{1}{2}m\omega^2 z^2$ with $\omega = 2\pi \times 10$ Hz and m the mass of a ^{87}Rb atom. The effective potential $U(z)$ as experienced by the atoms is the sum of the magnetic trap $V_{mag}(z)$ and the optical dipole potential $V(z)$, which is estimated from measured intensity (see Sec.VI for details). We design targets V^{tar} so that the effective potential $U(z)$ will be either constant, sinusoidal, or linear. All methods start with the DMD completely turned off.

Qualitatively all the methods are successful, as the final results are barely distinguishable from the target potential. For quantitative evaluation of the optimization performance, we compute the locally normalized root mean square error (RMS):

$$\epsilon_{RMS} = \sqrt{\frac{1}{N_V}\sum_{i=1}^{N_V}\left(\frac{V_i^{tar} - V_i^n}{V_i^{tar}}\right)^2}. \tag{8}$$

We say that the normalization is local because it is computed point by point. As a measure of the minimum error that can be achieved due to shot-to-shot fluctuations, we acquire 100 pictures of the optimized potential and compute the average ϵ_{RMS}, substituting V^{tar} by the average potential over the sample.

The measured optimization error of the offline ILC is shown with a constant blue line in Fig.5 and ranges from 4% to 16% depending on the target. We find that the offline ILC scheme is able to deliver a level of accuracy already comparable to what we obtain with ~ 100 iterations of the heuristic algorithm, which would be equivalent to more than 1 hour of experimental time per potential shape for BEC experiments (and more than 5 hours with averaging over at least 5 images). The optimization error for potentials obtained through offline learning is determined by the predictive capability of the ML model, which depends on the training data set. To illustrate the robustness of the physics-inspired model and its ability to extrapolate beyond training data we here train it on a generic data set consisting only of parabolic potentials. We furthermore find that better results (with error around 3%) are achievable when the training data set resembles the optimized configurations more closely.

For a quasi-1D BEC with an atomic density of 100 μm^{-1}, atomic shot noise (which limits the precision with which we can measure the potential) is at the level of 10 % [28], and the chemical potential is of the order of 1 kHz. Since constant offsets in the optimized potential do not have an effect on the atoms, we can neglect them. The scale of the remaining imperfections is then reduced to around 10 Hz, which is 1% of the typical chemical potential. Measuring a single potential with such precision requires averaging over 100 repetitions, which is roughly an hour of experimental time.

We can compare these results with [6], where the online heuristic scheme gave $\epsilon_{RMS} \sim (4-6)\%$ for the atom density in a similar setup. These numbers suggest that the offline ILC alone is capable of generating potentials with comparable accuracy to existing heuristic schemes. Lifting the necessity of online iterations might be a particularly appealing choice for time-dependent potentials.

The online ILC and heuristic algorithms both converge to the same error level. Yet online ILC reaches convergence in around ~ 10 iterations while the heuristic algorithm needs around ~ 100 iterations, which is a great advantage for schemes incorporating experimental feedback. Moreover, the ILC algorithm does not need to manually select an optimization schedule, resulting in increased flexibility and bypassing time-consuming parameter tuning. Convergence time can be decreased even more by using the result of offline optimization as an initial guess for online ILC. In this case, we find that the online ILC reaches convergence in only a few iterations, resulting in an even larger speed-up compared to the heuristic method. In any case, the final error is far below the atomic shot noise, so it is hardly accessible in static BEC configurations.

Unlike the online and offline ILC, the heuristic algorithm does not rely on dimensional reduction and the concept of virtual input, therefore it is not restricted to the symmetric class of patterns described by Eq.(2). The fact that the same error level can be achieved despite this restriction implies that this choice does not constitute a significant bottleneck for potential optimization. While other mappings are possible, such as the optimized dithered columns suggested in [17], we found that symmetric mapping offers several advantages. In fact, it is easier to realize (as it does not require the solution of an additional optimization problem) and also more robust against variations of the beam shape along the transversal direction.

We ran the same tests using a single-frequency laser as a light source (see Sec.VI for details). The physics-inspired model performs worse in the prediction of laser-generated potentials, therefore the output of offline ILC is noticeably worse than the predictions for SLD. On the other hand, the online ILC algorithm optimizes the optical potentials created with the laser to the same error level as with the SLD. Based on these findings, we can state that our method is well suited for experimental setups employing SLDs as well as single-frequency lasers.

V. CONCLUSIONS

In this paper, we presented our experimental setup for generating and efficiently optimizing 1D optical potentials. We combined a digital micromirror device for potential shaping control together with a SLD light source. We performed measurements estimating the quantitative

difference between SLD's and laser's coherence properties and showing the advantages of using SLD due to its generally linear behavior.

We have implemented learning algorithms that enable efficient optimization of optical potentials. We have shown how to build a physics-inspired model, which acts as a digital twin of the experimental setup. The model is able to recreate the main features of the optical system based on a small set of experimental data without the need to use deep (neural) networks and large training data sets, with the advantage of saving experimental time.

The application of the Iterative Learning Control optimization method provides a more than ten-fold speed-up compared to heuristic approaches. The ILC algorithms used offline with the trained models are able to optimize optical potentials with a precision acceptable for most experiments with trapped 1D ultracold gases. Using online ILC with experimental feedback allows us to optimize optical potentials to error levels comparable to measurement error, giving a ten-fold speed-up compared to more straightforward heuristic algorithms. By combining both ILC strategies, namely using the result of an optimized digital twin configuration as an initial guess for online ILC, we get the optimized potential with just a few experimental iterations.

Regarding optimization performance for the SLD and laser, we find that the SLD outperforms the laser using offline ILC. However, the online ILC performs equally well for laser and SLD giving the same level of optimization error. The model we developed combined with Iterative Learning Control provides a very fast way to optimize optical potentials with a DMD which might be used in a large variety of experimental setups. Our work offers a prospect for fast optimization of optical dipole potentials which is very important for time-costly experiments or for very large sequences of patterns in dynamic situations.

VI. APPENDIX

A. DMD mount

Due to the specific construction of the Texas Instruments DLP650LNIR DMD's micromirror control mechanism [29], the DMD is mounted 45° rotated so all the optical elements are placed in one plane. Any pattern getting rotated 45° right before projecting on the DMD. We verified that during optimization the rotation only leads to distortions of the potential which are below the resolution and therefore optimization is not affected.

B. Intensity to optical dipole potential conversion

The number of pixels in the output V_i as obtained from the camera does not necessarily coincide with the number of pixels in the input ν_i. In order to use the model, we first interpolate V_i to the input grid size, using the `interp1` function in Matlab.

We assume the relation between light intensity and optical dipole potential to be linear $V = \alpha_V I$ with α_V as found in [30]. Since we work with red-detuned light, α_V is negative. We also suppose the relation between the CCD readout R and intensity to be linear. We then compute $\alpha_{CCD} = I r_{pow}/R$ by measuring the light intensity with a power meter. To not saturate the CCD we use a reduced amount of light intensity. To calculate the finally expected dipole potentials we employ a factor $r_{pow} = I_{full}/I_{low}$ that accounts for the source operating at low power.

C. Mathematical details

This paper heavily relies on the discretization of functions of a real variable $f(z)$ in order to obtain finite size vectors. If we define a coordinate grid $z_i = (i-1)\Delta - \bar{z}$ for $i = 1, ..., n$, then we refer to any discretized function as $f_i = f(z_i)$ and we denote with $\boldsymbol{f}$ the $\mathbb{R}^n$ vector whose elements are f_i.

Let us call $\mathcal{F}$ the discrete Fourier transform acting on a vector of size n, and $\mathcal{F}^{-1}$ its inverse:

$$\mathcal{F}[\boldsymbol{a}]_k = \sum_{j=1}^{n} a_j e^{-\frac{2\pi i}{n}(j-1)(k-1)} \tag{9}$$

$$\mathcal{F}^{-1}[\boldsymbol{b}]_k = \frac{1}{n}\sum_{j=1}^{n} b_j e^{\frac{2\pi i}{n}(j-1)(k-1)} \tag{10}$$

Then, we can define the discrete convolution of two vectors $\boldsymbol{a} * \boldsymbol{b}$ as

$$\boldsymbol{a} * \boldsymbol{b} = \mathcal{F}^{-1}[\mathcal{F}[\boldsymbol{a}]\mathcal{F}[\boldsymbol{b}]] \tag{11}$$

where the product on the right-hand side is element-wise.

D. Heuristic algorithm

In order to assess the advantages of the ILC methods, we employed an adapted version of the heuristic algorithm described in [6] as a reference. It is an iterative algorithm that updates the state of each pixel based on the local differences between measured and target potentials. The optimization happens in two phases, the first fast but rough and the second slower but more precise. During the rough phase (see Fig.5 first ∼20 iterations) in each column DMD pixels are turned on until the difference gets lower than the chosen threshold. During the precise phase pixels can be moved away or turned off.

E. Laser vs SLD

We show in Tab. I a detailed comparison of the performance of the offline and online ILC algorithms for the

	Target 1	Target 2	Target 3
SLD, offline	0.10	0.04	0.16
Laser, offline	0.15	0.13	0.18
SLD, online	0.02	0.01	0.02
Laser, online	0.02	0.02	0.03

TABLE I. Values of ϵ_{RMS} for SLD and laser sources.

SLD and laser light sources. The values of ϵ_{RMS}, cf. Eq.(8), should be compared with the shot-to-shot errors of 0.01 for the SLD and 0.02 for the laser, which are the error components that cannot be eliminated via optimization.

Acknowledgments

We would like to thank Mohammadamin Tajik and João Sabino for the discussions and technical support. This project was funded by the DFG/FWF CRC 1225 'Isoquant', Project-ID 273811115, by the Austrian Science Fund (FWF) P 36236-N (financed by the European Union - NextGenerationEU), by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – Cluster of Excellence Matter and Light for Quantum Computing (ML4Q) EXC 2004/1 – 390534769, and from the German Federal Ministry of Education and Research through the funding program quantum technologies—from basic research to market under the project FermiQP, 13N15891. M.P. has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 101032523.

[1] G. Zhao, C. Zheng, C. Kuang, R. Zhou, M. M. Kabir, K. C. Toussaint, W. Wang, L. Xu, H. Li, P. Xiu, and X. Liu, "Nonlinear Focal Modulation Microscopy," Phys. Rev. Lett. **120**, 193901 (2018).

[2] C. Gross and I. Bloch, "Quantum simulations with ultracold atoms in optical lattices," Science **357**, 995–1001 (2017).

[3] K. Henderson, C. Ryu, C. MacCormick, and M. G. Boshier, "Experimental demonstration of painting arbitrary and dynamic potentials for Bose–Einstein condensates," New Journal of Physics **11**, 043030 (2009).

[4] G. Gauthier, I. Lenton, N. M. Parry, M. Baker, M. J. Davis, H. Rubinsztein-Dunlop, and T. W. Neely, "Direct imaging of a digital-micromirror device for configurable microscopic optical potentials," Optica **3**, 1136–1143 (2016).

[5] J. L. Ville, T. Bienaimé, R. Saint-Jalm, L. Corman, M. Aidelsburger, L. Chomaz, K. Kleinlein, D. Perconte, S. Nascimbène, J. Dalibard, and J. Beugnon, "Loading and compression of a single two-dimensional Bose gas in an optical accordion," Phys. Rev. A **95**, 013632 (2017).

[6] M. Tajik, B. Rauer, T. Schweigler, F. Cataldini, J. Sabino, F. S. Møller, S.-C. Ji, I. E. Mazets, and J. Schmiedmayer, "Designing arbitrary one-dimensional potentials on an atom chip," Opt. Express **27**, 33474–33487 (2019).

[7] J. Liang, J. Rudolph N. Kohn, M. F. Becker, and D. J. Heinzen, "High-precision laser beam shaping using a binary-amplitude spatial light modulator," Appl. Opt. **49**, 1323–1330 (2010).

[8] A. L. Gaunt, T. F. Schmidutz, I. Gotlibovych, R. P. Smith, and Z. Hadzibabic, "Bose-Einstein Condensation of Atoms in a Uniform Potential," Phys. Rev. Lett. **110**, 200406 (2013).

[9] N. Navon, R. P. Smith, and Z. Hadzibabic, "Quantum gases in optical boxes," Nature Physics **17**, 1334–1341 (2021).

[10] G. Gauthier, T. A. Bell, A. B. Stilgoe, M. Baker, H. Rubinsztein-Dunlop, and T. W. Neely, "Chapter One - Dynamic high-resolution optical trapping of ultracold atoms," Advances In Atomic, Molecular, and Optical Physics, **70**, 1–101 (2021).

[11] P. Zupancic, P. M. Preiss, R. Ma, A. Lukin, M. E. Tai, M. Rispoli, R. Islam, and M. Greiner, "Ultra-precise holographic beam shaping for microscopic quantum control," Opt. Express **24**, 13881–13893 (2016).

[12] A. Smith, T. Easton, V. Guarrera, and G. Barontini, "Generation of optical potentials for ultracold atoms using a superluminescent diode," Phys. Rev. Research **3**, 033241 (2021).

[13] R. W. Floyd and L. Steinberg, "An Adaptive Algorithm for Spatial Greyscale," Proceedings of the Society for Information Display **17**, 75–77 (1976).

[14] C. Dorrer and J. D. Zuegel, "Design and analysis of binary beam shapers using error diffusion," J. Opt. Soc. Am. B **24**, 1268–1275 (2007).

[15] P. Caramazza, O. Moran, R. Murray-Smith, and D. Faccio, "Transmission of natural scene images through a multimode fibre," Nature Communications **10**, 2029 (2019).

[16] D. Bristow, M. Tharayil, and A. Alleyne, "A survey of iterative learning control," IEEE Control Systems Magazine **26**, 96–114 (2006).

[17] A. Deutschmann-Olek, M. Tajik, M. Calzavara, J. Schmiedmayer, T. Calarco, and A. Kugi, "Iterative shaping of optical potentials for one-dimensional Bose-Einstein condensates," 2022 IEEE 61st Conference on Decision and Control (CDC) , 5801–5806 (2022).

[18] C. A. Sackett and B. Deissler, "A white-light trap for Bose-Einstein condensates," Journal of Optics B: Quantum and Semiclassical Optics **6**, 15–20 (2003).

[19] A. d. Campo and M. G. Boshier, "Shortcuts to adiabaticity in a time-dependent box," Scientific Reports **2**, 648 (2012).

[20] M. Gluza, J. Sabino, N. H. Ng, G. Vitagliano, M. Pezzutto, Y. Omar, I. E. Mazets, M. Huber, J. Schmiedmayer, and J. Eisert, "Quantum Field Thermal Machines," PRX Quantum **2**, 030310 (2021).

[21] S. Manz, R. Bücker, T. Betz, C. Koller, S. Hofferberth, I. E. Mazets, A. Imambekov, E. Demler, A. Perrin,

J. Schmiedmayer, and T. Schumm, "Two-point density correlations of quasicondensates in free expansion," Phys. Rev. A **81**, 031610(R) (2010).

[22] C. K. Hitzenberger, M. Danner, W. Drexler, and A. F. Fercher, "Measurement of the spatial coherence of superluminescent diodes," Journal of Modern Optics **46**, 1763–1774 (1999).

[23] B. Redding, M. A. Choma, and H. Cao, "Speckle-free laser imaging using random laser illumination," Nature Photonics **6**, 355–359 (2012).

[24] V. N. Vapnik, "Estimation of Dependences Based on Empirical Data," , 1–505 (2006).

[25] S. Geman, E. Bienenstock, and R. Doursat, "Neural Networks and the Bias/Variance Dilemma," Neural Computation **4**, 1–58 (1992).

[26] J. Ghosh and B. Paden, "A Pseudoinverse-Based Iterative Learning Control," IEEE Transactions on Automatic Control **47**, 831–837 (2002).

[27] A. Deutschmann-Olek, G. Stadler, and A. Kugi, "Stochastic Iterative Learning Control for Lumped- and Distributed-Parameter Systems: A Wiener-Filtering Approach," IEEE Transactions on Automatic Control **66**, 3856–3862 (2021).

[28] T. Berrada, S. van Frank, R. Bücker, T. Schumm, J.-F. Schaff, and J. Schmiedmayer, "Integrated Mach-Zehnder interferometer for Bose-Einstein condensates," Nature Communications **4**, 2077 (2013).

[29] B. Lee, "Introduction to Digital Micromirror Devices (DMDs)," Texas Instruments , 1–13 (2008).

[30] R. Grimm, M. Weidemüller, and Y. B. Ovchinnikov, "Optical Dipole Traps for Neutral Atoms," Academic Press Advances In Atomic, Molecular, and Optical Physics, **42**, 95–170 (2000).

Chapter 5

Fast neutral atom transport and transfer between optical tweezers

Authors:	Cristina Cicali, **Martino Calzavara**, Eloisa Cuestas, Tommaso Calarco, Robert Zeier, and Felix Motzoi
Year:	2025
Journal:	Physical Review Applied
Volume/Issue:	24/2
DOI:	10.1103/7r3w-8m61

5.1 Summary and context

This theoretical paper [97] concerns a common quantum control task in a cold atom platform, namely the transport in space of an atom. This auxiliary operation is crucial in many ways, for example to deterministically prepare atom arrays [9, 10], to operate space-segmented designs in which storage regions are separated by computing and readout regions and to allow the application of local interaction gates between distant atoms [11]. It is also a rather slow operation compared to quantum gates, creating a potential bottleneck to the performance of cold atom platforms. We consider a setup involving a static trapping potential, capable of holding an array of atoms in its sites. By means of moving optical tweezers, we show how it is possible to collect an atom initially residing in one of these sites and then to move and deposit to another site, while minimizing residual vibrational excitations.

In order to achieve this goal, we employ both analytical STA methods and numerical optimal quantum control (see Sec. 3.3 and Sec. 3.2). More specifically, we make use of approximate STA solutions using invariant-based engineering [91]

for both the (un)loading of the atom in the moving tweezers and its transport to the final site [90], and concatenate them to create the full protocol. We show that while analytical methods give rise to both performant and stable solutions, that under some approximation can be also put in closed form for ease of experimental deployment, closed-loop control methods can further refine them while avoiding longer convergence times due to cold starts. We estimate that with our methods the transport task we consider can be significantly sped up compared to state-of-the-art experiments.

In the context of this thesis, this publication shows an example of how different quantum control methods can work together and be combined with physical information about the quantum system to implement fast and precise auxiliary (but crucial) hardware operations, while minimizing overhead in experimental work.

5.2 Author contribution

The Author is responsible for the ideation of the STA methods employed in the paper and of their implementation. He contributed to the development of the numerical software used in the paper and to the discussion of the results. He is responsible for drafting the Sections related to the STA transport pulses, especially Sec. III.B and App. A, and contributed to the writing of all sections and to the design of the Figures.

Fast neutral-atom transport and transfer between optical tweezers

Cristina Cicali,[1,2,*] Martino Calzavara,[1,2] Eloisa Cuestas,[1,3] Tommaso Calarco,[1,2,4] Robert Zeier,[1] and Felix Motzoi[1,2,†]

[1] *Forschungszentrum Jülich GmbH, Peter Grünberg Institute, Quantum Control (PGI-8), 52425 Jülich, Germany*
[2] *Institute for Theoretical Physics, University of Cologne, Zülpicher Straße 77, 50937 Cologne, Germany*
[3] *Quantum Systems Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan*
[4] *Dipartimento di Fisica e Astronomia, Università di Bologna, 40127 Bologna, Italy*

(Received 24 January 2025; revised 28 May 2025; accepted 14 July 2025; published 28 August 2025)

We study the optimization of the transport and transfer of neutral atoms between optical tweezers, both critical steps in the implementation of quantum computers and simulators. We analyze four experimentally relevant pulse shapes (piecewise linear, piecewise quadratic, minimum jerk, and a combination of linear and minimum jerk), and we also develop a protocol using shortcuts-to-adiabaticity (STA) methods to crucially incorporate the time-dependent effects of static traps. By computing a measure of the final transport error and two measures of the heating during transport, we show that our proposed STA protocol comprehensively outperforms all the experimentally inspired pulses. After further optimizing the pulse shapes, we find a lower bound on the protocol duration, compatible with the time at which the vibrational excitations exceed half of the states hosted by the moving tweezer. This lower bound is at least eight times faster than the one reported in recent experiments, which highlights the importance of including and optimizing the transfer from and to static traps, which may be the largest bottleneck to speed. Finally, our STA results demonstrate that a modulation in the depth of the moving tweezer designed to time-dependently counteract the effect of the static traps is key to reducing errors and reducing the pulse duration. To motivate the implementation of our STA pulses in future experiments, we provide a simple analytical approximation for the moving-tweezer position and depth controls.

DOI: 10.1103/7r3w-8m61

I. INTRODUCTION

Over the last decade we have witnessed a sustained growth in the capabilities for quantum information processing and simulation, based on significant progress in both the confinement and control of atomic arrays [1–7], with a critical role played by the precise manipulation and transport of neutral atoms in optical tweezers [8,9]. In state-of-the-art experiments, the atoms are usually cooled to a temperature of 10–100 μK and then stochastically loaded (with a probability of about 0.5, or half-filling) from a magneto-optical trap into a set of space- and time-controlled Gaussian traps or optical lattices. In general, the experimental setup involves a spatial light modulator to create the static trap array, while the moving optical tweezers are controlled by acousto-optic deflectors (AODs) [1,6,8]; other approaches for similar experiments rely only on tweezers generated and controlled via AODs [9–12]. These tweezers allow for a precise rearrangement of stochastically loaded atoms, enable nonlocal connectivity, and eliminate the need to prepare a new ensemble of atoms after each measurement, thus enhancing experimental efficiency [13].

The basic requirements for a quantum processor include initialization and storage of qubits in a quantum register, bringing qubits sufficiently close to realize quantum gates and the final readout [14]. Therefore, efficient atom transport is not only a key step in performing controlled translations from the preparation or cooling chamber to the science cell but also allows for on-demand interactions to realize quantum operations in the correct location (including both processing and storing sites) and with precise timing. In this context, fast and accurate manipulation of atomic motion emerges as a central requirement for quantum technologies to preserve coherence while attaining high fidelities between the final state and a predefined

*Contact author: c.cicali@fz-juelich.de
†Contact author: f.motzoi@fz-juelich.de

target state [15]. Moreover, atom transport appears as a potential long-term bottleneck in quantum computing with cold atoms, given that other operations are expected to be significantly faster [16].

Several experiments have been undertaken to study atom shuttling [17–20], where the ideal goal is to obtain a final state as close as possible to the initial state while avoiding losses and vibrational excitations. Although the use of long transport times (adiabatic processes) may seem the natural path to high final fidelities, they also translate into the accumulation of decoherence and experimental noise [18]. Conversely, diabatic (i.e., fast) processes lead by default to higher excitations and thus to overheating and losses. It is therefore necessary to find a trade-off between high fidelity and transport time [6,7]. In this work, we focus on the optimization of the transport and transfer of atoms between optical tweezers to speed up these operations while avoiding unwanted excitations. In particular, the protocols we design include the capturing and releasing of the atom from and to static traps (which we refer to as *transfer*) as well as the *transport* itself. At the end of the process, the atom is located in the desired position. This results in increased flexibility in experimental operations (once the protocol is completed, the moving tweezers are free to be used for other tasks) and also allows for a shorter transfer part, which has been reported to be the most time-consuming stage [17].

To achieve our goals, we apply optimal quantum control techniques to drive the system toward the desired state by minimizing the infidelity [18,21–25], a measure of the deviation between the target state and the state after the evolution of a particle moving from one optical trap to the next. Even though the engineered evolution path should by construction lead to a final state close to the desired one, there is no guarantee regarding the no-heating or no-loss condition, meaning that the system might be excited into the upper levels of the trap during the protocol and atoms may potentially be lost [26]. To solve this, we first analyze and characterize the performance of several tweezer trajectories used in state-of-the-art experiments (we consider four families: piecewise linear, minimum jerk, piecewise quadratic, and a hybrid between the linear and minimum-jerk ones [1,7,8,10–12,17]), with a particular focus on a measure of the error after transport and two measures of the vibrational excitations during transport as a function of the total time of the protocol.

In addition to considering experimentally motivated tweezer trajectories, we further develop a shortcuts-to-adiabaticity (STA) approach [15,27–31] to generate analytical pulses, which can also serve as a seed for subsequent optimization. More specifically, the STA method leverages a Hamiltonian invariant to derive pulses that drive the system in a nonadiabatic (or fast) way to the same final state as their adiabatic (or slow) counterparts [15,30]. The STA method provides great flexibility for computing the tweezer trajectory because it depends on ansatz functions that can be chosen freely, provided that they satisfy the necessary boundary conditions. When addressing the atom-transport problem, a wide variety of STA solutions have been derived for a harmonic trap or a power-law potential with time-independent frequency [15,25,31]. In our approach, we include the static-tweezer potential in the harmonic approximation to obtain solutions tailored to our specific setting. We show that our STA protocol outperforms the best experimentally inspired transport pulses, and our analytical approximation formulas can be straightforwardly implemented in experiments.

Following the characterization of the initial pulses, we carry out optimal pulse shaping using the user-friendly QuOCS toolkit [32]. Taking advantage of an expansion of the pulse over a randomized function basis, the included d-CRAB algorithm [33–36] allows for a drastic reduction of the number of free parameters to be optimized, so that a direct search method (such as Nelder-Mead) can be employed. At the same time, the need for computing the gradient of the control objective (or figure of merit) is alleviated. With this approach, low infidelities with respect to the target state were achieved, improving by more than one order of magnitude compared to the analytical pulse shapes. All the optimized pulses present a more stable behavior of the fidelity with respect to the target state after transport. Notably, while the optimized pulses reach an error threshold of 10^{-4} in a shorter total time compared to the nonoptimized experimentally motivated pulses (with reductions in time of 10%–30%), our proposed STA pulse requires the shortest time to obtain an error below 10^{-4}, and that time does not change after optimization. We interpret this as a signature of the high quality and suitability of our STA solution, as it approaches the numerically observed quantum speed limit (QSL) [18,37].

We aim to identify parameter regions of suitable pulses that can be used in the design of realistic experimental protocols. Thus, we choose the two pulses with the best performance and provide a heat map of the error after transport as a function of the total time and the depth of the moving tweezer. We determine a region in the parameter space for the minimum-jerk pulse leading to errors below 10^{-4} and certain *magical time windows* for the STA pulse where the error is suppressed by at least one order of magnitude [25]. In the investigated parameter interval, we observe that the performance of our STA pulse is almost independent of the moving-tweezer depth.

Finally, we identify a constraint on the choice of the total transport duration linked to the fundamental QSL of the system [18,37]. For an experimentally achievable moving-trap depth of about $3.57 \times 2\pi$ MHz and a transport distance of 7 μm, our results point toward a QSL of about 8 τ_{st} (here, for ^{39}K atoms, $\tau_{st} \approx 0.03$ ms is the characteristic time of the static tweezers, and this is related to that of the moving trap τ_{mt} via $\tau_{st} \approx 2\,\tau_{mt}$ for the

considered maximum tweezer depth), a value that coincides with the time at which the population of vibrational states is equal to half of the states hosted by the trap. Although our obtained time threshold for the transport process is roughly five times larger than the one reported in Ref. [20], our results show a lower time threshold for the atom-transfer stage; in our case the time invested in the transfer stage is nine times smaller than the experimental time reported in Ref. [19]. Since the time required for capturing or releasing the atom was estimated to be 12 times larger than the time needed for the pure transport process [17], and the protocol of Ref. [20] does not include this time cost, our results highlight the importance of including the transfer between tweezers in the model.

The remainder of this paper is structured as follows. In Sec. II, we present our model for the atom-transport problem. In Sec. III, we describe the considered experimental pulses together with the method to generate analytical pulses by means of an STA procedure. In Sec. IV, we characterize the transport process for all the considered pulses and report the optimization results obtained using realistic experimental parameters. A summary and conclusions are given in Sec. V. Appendixes A–D provide details about the STA calculations, the numerical time evolution method, and the transport dynamics.

II. ATOM TRANSPORT IN AN EXTERNAL POTENTIAL

The transport of an atom in the presence of a static external potential $V_{\rm st}(x)$, from the initial position $x_{\rm i}=0$ at time $t_{\rm i}=0$ to the final position $x_{\rm f}=d$ at time $t_{\rm f}=T$ can be generically described by the Hamiltonian

$$H(x,t)=-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}+V_{\rm mt}(x,t)+V_{\rm st}(x), \tag{1}$$

where m is the mass of the atom. In current experiments, the external potential $V_{\rm st}(x)$ is typically generated either by a series of stationary optical tweezers, leading to a sum of Gaussian traps, or by counterpropagating laser beams, resulting in a sinusoidal pattern [1,8,9,14,38–43]. Here, we consider ^{39}K atoms with $m=6.47\times10^{-26}$ kg, and we focus on static Gaussian traps, meaning that the external potential $V_{\rm st}(x)$ consists of a set of $n_{\rm st}$ time-independent Gaussian wells defined by the depth amplitude $A_{\rm st}/\hbar=0.53\times2\pi$ MHz, width $\sigma_{\rm st}=0.35$ μm, and minima positions $x_{\rm st}^n=(n-1)d$, with $d=7$ μm being the distance between two adjacent minima of the periodic potential and $n=1,2,\ldots,n_{\rm st}$ [44]. To model the moving tweezer, we use a time-dependent Gaussian potential $V_{\rm mt}(x,t)$ and assume that the width of the tweezer is fixed to $\sigma_{\rm mt}=0.47$ μm during the evolution [17,19,44,45]. Then, the control parameters for our problem are the amplitude of the moving tweezer $A_{\rm mt}(t)$ (trap depth) and the position of the center of the Gaussian well $x_{\rm mt}(t)$. The static and moving potentials are respectively

$$V_{\rm st}(x)=-A_{\rm st}\sum_{n=1}^{n_{\rm st}}\exp\left(-\frac{(x-x_{\rm st}^n)^2}{2\sigma_{\rm st}^2}\right) \quad \text{and} \tag{2a}$$

$$V_{\rm mt}(x,t)=-A_{\rm mt}(t)\exp\left(-\frac{(x-x_{\rm mt}(t))^2}{2\sigma_{\rm mt}^2}\right). \tag{2b}$$

For the particle to be correctly transferred from the initial static tweezer to the moving one, transported through the total distance d, and then transferred to the final static tweezer, the following boundary conditions need to be satisfied:

$$\begin{aligned} &A_{\rm mt}(t_{\rm i})=0 \quad \text{and} \quad A_{\rm mt}(t_{\rm f})=0,\\ &x_{\rm mt}(t_{\rm i})=x_{\rm i}=0 \quad \text{and} \quad x_{\rm mt}(t_{\rm f})=x_{\rm f}=d. \end{aligned} \tag{3}$$

While the maximum amplitude for both the moving tweezer and the static traps are given by the power of the laser beams, the dynamics of the transport protocol depend on the trade-off between the static and moving amplitudes. Comparable amplitudes translate into a strong interaction between the moving and static tweezers, with the consequent oscillations of the atom in the traps that, in turn, lead to losses during transport. To prevent this from happening, we impose an extra condition on the amplitude during transport: $A_{\rm mt}>A_{\rm st}$, usually chosen as $A_{\rm mt}\approx10A_{\rm st}$ [6,17,19].

Figure 1 presents a scheme of the atomic-transport protocol. The initial state $|\psi_{\rm i}\rangle$ and final target state $|\psi_{\rm tg}\rangle$ are the ground states, computed via exact diagonalization, of the static Hamiltonian with a single Gaussian potential centered at $x_{\rm i}=0$ and $x_{\rm f}=d$, respectively. The goal is to determine the controls $x_{\rm mt}(t)$ and $A_{\rm mt}(t)$ such that the state after the complete time evolution, denoted by $|\psi(t_{\rm f})\rangle$, closely approximates the localized ground state of the final targeted trap $|\psi_{\rm tg}\rangle$. To quantify the error after the transport protocol, we compute the infidelity $\mathcal{I}(t_{\rm f})=1-\mathcal{F}(t_{\rm f})$, where $\mathcal{F}(t)$ is the fidelity, which is defined as the overlap $\mathcal{F}(t)=|\langle\psi(t)|\psi_{\rm tg}\rangle|^2$. To shape the control pulses that execute the transport task with low infidelity, we first characterize the performance of several different pulse shapes based on current experimental realizations together with pulses that we derive within the STA formalism. In a second stage, we use these pulse profiles as initial guesses for optimal pulse shaping using the QuOCS toolkit, which allows us to reach even higher and more stable fidelities.

III. TRANSPORT PULSES

This section presents a theoretical description of the considered pulses. In Sec. III A, we describe the initial guesses motivated by experimental considerations. Section III B contains the analytical derivation of the STA solution.

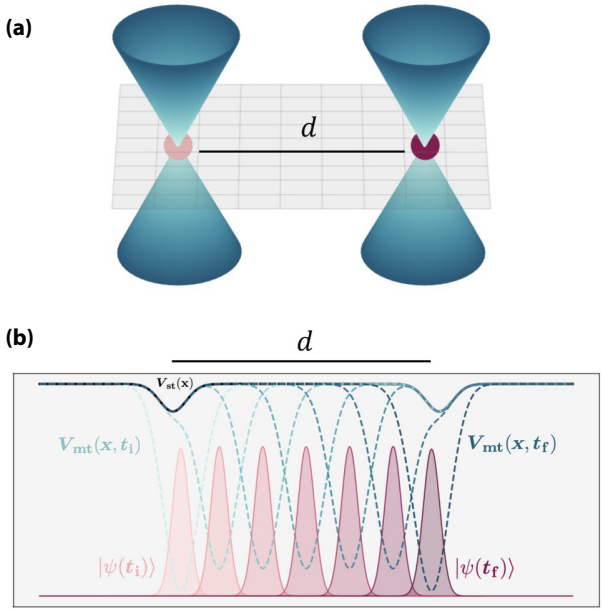

FIG. 1. Illustration of the transport protocol. (a) Two adjacent static tweezers (blue) separated by a distance d. (b) The atom initially trapped at $x_i = 0$ is transported to the position $x_f = d$ in a total time $T = t_f - t_i$. The amplitude of the external potential (static tweezers) denoted by $V_{st}(x)$ is shown as a solid black line, while that of the moving tweezer $V_{mt}(x, t)$ is depicted by dashed blue lines with a color gradient indicating the time evolution (from t_i to t_f lighter to darker—left to right). The probability of finding the particle in the position x is given by $|\psi(t)|^2$ and it is represented by the filled curves with a purple color gradient associated with the time arrow.

A. Pulses based on experiments

In current experiments, the choice of the ramps that control the position of the moving tweezer aims to suppress heating and losses, or equivalently, to minimize excitations during and after transport [1,7,8,10–12,17]. In our model, the control pulses time-dependently define the spatial trajectory followed by the bottom of the Gaussian trap $x_{mt}(t)$ and its depth, determined by the amplitude $A_{mt}(t)$. In practice, the minimum of the tweezer potential coincides with the focal spot [46], while the amplitude is proportional to the power of the beam [47]. The four transport trajectories based on current experiments that we consider are piecewise linear, piecewise quadratic, minimum jerk (derivative of the acceleration), and a hybrid trajectory that combines the linear and minimum-jerk ones. These are defined below.

The linear pulse, used for instance in Ref. [17], has the advantage of simplicity and the fact that the constant velocity can be implemented straightforwardly in experiments (using a constant sweep rate for the AOD frequency), but it is known to cause excitation due to the abrupt changes in the transport velocity at the discontinuity points. In the case of the piecewise-quadratic trajectory used in the experiments of Ref. [7], the atoms experience alternating accelerations of the same magnitude in the first and second half of the trajectory. This kind of quadratic pulse can be readily implemented with frequency-control systems [48]; however, the discontinuity in the acceleration might induce heating. The need for smooth position functions then arises to avoid unwanted excitations or energy excess. Although it was introduced in a completely different context (voluntary movements in primates [49]) 40 years ago, the minimum-jerk trajectory satisfies the latter requirement and was recently used to study diatomic molecule formation with the associated requirement of having not only the two atoms close enough but also in their relative motional ground state [10,11,19]. This trajectory minimizes the square of the jerk (derivative of the acceleration) over the full path, and it is also obtained when a polynomial ansatz is used to solve the position as a function of time for a translation with zero initial and final velocity and acceleration. In Ref. [10] the authors argue that the selection of the trajectory should aim to minimize heating due to jerk at the end points, and to avoid parametric heating due to trap depth oscillations [50]. The latter condition is further explained in Ref. [19]; when the frequency of the AOD is driven with a constant sweep rate (associated with a linear translation) resonant intensity modulations arising from imperfections in the AOD are avoided (see also Ref. [12]). A good compromise between preventing heating from changes in the acceleration at the beginning and end of the movement and at the same time avoiding resonant intensity modulations is found by using a hybrid pulse that implements a minimum-jerk trajectory for the start and end of the ramp and a constant-velocity (linear) translation for intermediate times.

As polynomial or piecewise-polynomial functions of time, the linear, quadratic, minimum-jerk, and hybrid trajectories have a simple analytical form. Since the target state in the final tweezer should closely match the initial state in the initial tweezer, and the considered static tweezers only differ in their position, one can chose symmetric "palindromic" pulses to reduce the degrees of freedom. In particular, the four families of experimentally motivated pulses that we consider are time-reversal symmetric. The linear trajectory x^l is given by

$$x^l(d, \tau, t) = d\frac{t}{\tau}, \tag{4}$$

where d is the distance covered and τ is the total time required for transport.

The piecewise-quadratic trajectory x^q, with a position described by two parabolas that intersect in the midpoint and have piecewise-constant acceleration, can be

written as

$$x^{\mathrm{q}}(d,\tau,t) = d\begin{cases} 2\left(\dfrac{t}{\tau}\right)^2 & \text{for } 0 \le t \le \dfrac{\tau}{2}, \\ -2\left(\dfrac{t}{\tau}\right)^2 + 4\dfrac{t}{\tau} - 1 & \text{for } \dfrac{\tau}{2} \le t \le \tau. \end{cases} \tag{5}$$

In turn, the minimum-jerk trajectory x^{mj} reads

$$x^{\mathrm{mj}}(d,\tau,t) = d\left(10\left(\frac{t}{\tau}\right)^3 - 15\left(\frac{t}{\tau}\right)^4 + 6\left(\frac{t}{\tau}\right)^5\right), \tag{6}$$

where it is straightforward to check that the initial ($t = 0$) and final ($t = \tau$) velocity and acceleration are equal to zero.

Finally, the hybrid trajectory x^{hyb} is defined upon the fraction ξ of the total transport time that follows a linear motion. This fraction parameter is also called hybridicity [19]. Since the total time under linear motion $\xi\tau$ ranges between 0 and τ, we have $0 \le \xi \le 1$. Moreover, for $\xi = 0$, the trajectory reduces to the minimum-jerk trajectory, while for $\xi = 1$, the trajectory coincides with the piecewise-linear one, so it is reasonable to expect that when changing the parameter ξ, the dynamics of the system should interpolate between the dynamics under the linear and minimum-jerk pulses. The hybrid trajectory can be written as

$$x^{\mathrm{hyb}}(d,\tau,t) = \begin{cases} x^{\mathrm{mj}}\left(d\dfrac{8(1-\xi)}{8+7\xi}, \tau(1-\xi), t\right) & \text{for } 0 \le t \le \dfrac{1-\xi}{2}\tau, \\ d\left(\dfrac{15}{8+7\xi}\dfrac{t}{\tau} - \dfrac{7(1-\xi)}{2(8+7\xi)}\right) & \text{for } \dfrac{1-\xi}{2}\tau \le t \le \dfrac{1+\xi}{2}\tau, \\ x^{\mathrm{mj}}\left(d\dfrac{8(1-\xi)}{8+7\xi}, \tau(1-\xi), t-\tau\xi\right) + d\dfrac{15\xi}{8+7\xi} & \text{for } \frac{1+\xi}{2}\tau \le t \le \tau. \end{cases} \tag{7}$$

In Ref. [10], the experimental values of ξ are 0 and 0.95, i.e., a full minimum-jerk trajectory and a 95% linear one. The authors of Ref. [19] use $\xi = 0.1$ [see their Fig. 3(b)], value with which they report being able to avoid resonant intensity modulations for a translation of 4.5 μm and $\tau \approx 1.3$ ms.

As mentioned before, we account not only for the transport process but also the transfer from the initial static trap to the moving tweezer and from the moving tweezer toward the target static one. Therefore, the protocol consists of three main stages: capturing the particle initially in the first static trap, the transport, and the final release of the particle in the target static tweezer. In the capturing stage, the depth of the moving tweezer is monotonically raised from zero to its maximum amplitude $A_{\mathrm{mt}}^{\max}$ while the position is kept constant. In the second transport stage, the position of the focal point follows one of the trajectories of Eqs. (4)–(7), while the power of the laser is kept constant, consistent with the experiments described in Ref. [11]. The third releasing stage is the inverse process of capturing: while maintaining the target position, the depth of the moving tweezer is decreased to zero. To assess the dynamics and quality of the final state, we add a fourth stage, which we call waiting time. During this time, the moving tweezer is off, and we evaluate several statistical measurements over the final state, mainly to examine its time stability and the possible presence of oscillations after transport (see for example Ref. [20]). Moreover, this waiting time can be used in the experiments to re-cool the atom before, for instance, the next concatenated transport step. The amplitude that we consider is then given by

$$A_{\mathrm{mt}}(t) = A_{\mathrm{mt}}^{\max}\begin{cases} \dfrac{3}{1-\eta}\dfrac{t}{T} & \text{for } 0 \le t \le \dfrac{1-\eta}{3}T, \\ 1 & \text{for } \dfrac{1-\eta}{3}T \le t \le \dfrac{1+2\eta}{3}T, \\ \dfrac{2+\eta}{1-\eta} - \dfrac{3}{1-\eta}\dfrac{t}{T} & \text{for } \dfrac{1+2\eta}{3}T \le t \le \dfrac{2+\eta}{3}T, \\ 0 & \text{for } \dfrac{2+\eta}{3}T \le t \le T, \end{cases} \tag{8}$$

where $t_i = 0$ and $t_f = T$. The total time T is divided into fractions defined by the parameter η, the transport time is assigned to be ηT, and the remaining time is divided into three intervals of equal length $(1-\eta)T/3$ for the capturing, releasing, and waiting stages. By doing so, we invest equal times for the capturing, releasing, and waiting stage. The obtained total position pulse is

$$x_{\mathrm{mt}}(t) = x_i + \begin{cases} 0 & \text{for } 0 \le t \le \dfrac{1-\eta}{3}T, \\ x(d, \eta T, t - \dfrac{1-\eta}{3}T) & \text{for } \dfrac{1-\eta}{3}T \le t \le \dfrac{1+2\eta}{3}T, \\ d & \text{for } \dfrac{1+2\eta}{3}T \le t \le \dfrac{2+\eta}{3}T, \\ d & \text{for } \dfrac{2+\eta}{3}T \le t \le T, \end{cases} \tag{9}$$

where $x(d, \tau, t)$ denotes any of the trajectories defined in Eqs. (4)–(7).

The respective pulses are shown in Figs. 2(a) and 2(b). Figure 2(a) shows the amplitude obtained via Eq. (8) for a maximum amplitude $A_{\mathrm{mt}}^{\max}/\hbar = A_{\exp}/\hbar = 3.57 \times 2\pi$ MHz, $d = 7$ μm, $\eta = 2/5$, and $T = 2.5$ ms [45]. The start and end times of the transport stage are indicated as gray dashed vertical lines. Both controls $A_{\mathrm{mt}}(t)$ and $x_{\mathrm{mt}}(t)$ satisfy the boundary conditions specified by Eq. (3). While the amplitude for the experimentally motivated pulses is always a piecewise-linear function, all the trajectories have an s-shape position dependence. In Fig. 2(b), the piecewise-linear trajectory (solid magenta) and minimum-jerk trajectory (solid blue) are presented as the limiting cases of the hybrid trajectory, which is depicted for several ξ values (gradient from magenta to purple). The minimum-jerk trajectory has the smallest velocity at the beginning (for details about the differences between the pulses at the beginning of the transport interval, see the insets of Fig. 2 depicting enlarged views of the pulses for times between 0.5 and 1 ms and positions between 0 and 1 μm) and at the end of the transport and reaches the highest one at the middle time point. For the hybrid pulse, the change ratio at the beginning and at the end of the transport decreases for decreasing hybridicity ξ. The quadratic pulse (dashed yellow line) has a higher velocity in the first and final part of the transport when compared to the minimum-jerk one, and it also has a slightly higher velocity in the middle time or inflection point.

B. Shortcut-to-adiabaticity solution

Since the piecewise-linear, hybrid, minimum-jerk, and quadratic control pulses are experimentally inspired ansätze for transport, they do not include the specifics of the system, which are encoded in the Hamiltonian of Eq. (1). In particular, there is no consideration of the shape of the moving tweezers $V_{\mathrm{mt}}(x, t)$, let alone the presence of the static tweezers $V_{\mathrm{st}}(x)$ in the background [see Eq. (2)]. To derive pulses that solve the fast-transport problem taking into account these effects while offering flexibility for satisfying desirable and realistic boundary conditions, we draw upon the STA framework. The main idea of STA methods is to develop protocols that lead to the same final state as their adiabatic (slow) counterpart in considerably shorter times [51], enabling us to avoid noise and decoherence effects.

The minimum-jerk trajectory was proposed to perform a translation of the minimum point of a perfect harmonic oscillator with minimal motional excitation at the beginning and end of the transport by constraining the velocity and acceleration to be zero at the initial and final times (see for instance Refs. [15,25,52]). Here, we use the minimum-jerk functional form in intermediate steps of the derivation of a new set of controls $A_{\mathrm{mt}}(t)$ and $x_{\mathrm{mt}}(t)$, which take into account the specific form of the moving and static tweezers. While several STA-based approaches for atom transport have been proposed [15,25,31,53–57], to the best of our knowledge, the background of static tweezers has not previously been incorporated into an STA solution. To address a more realistic scenario, we derive an STA control pulse that includes the transfer between optical tweezers in the transport, therefore avoiding errors during this critical stage.

The first step in our approach is to approximate the Hamiltonian of Eq. (1) with a time-dependent harmonic oscillator and then use known results on fast eigenstate transfer for the harmonic case [15,58]. In this way, the time-dependent oscillator accounts for both the static and moving tweezers. We start by expanding the complete potential in a Taylor series up to the second order around the center of the moving tweezer $x_{\mathrm{mt}}(t)$. By completing the resulting polynomial in x to a square form and ignoring a time-dependent energy offset, we obtain a

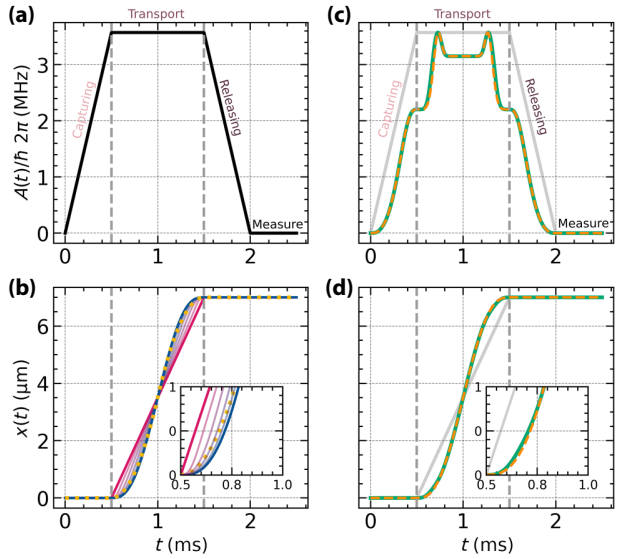

FIG. 2. Control pulses: position $x_{\rm mt}(t)$ and amplitude $A_{\rm mt}(t)$. The depth of the moving tweezer or amplitude control is shown in panels (a) and (c), while the position of the minimum of the moving trap is shown in panels (b) and (d) for a total time of a total time of $T = 2.5$ ms, a maximum amplitude $A_{\rm mt}^{\rm max}/\hbar = 3.57 \times 2\pi$ MHz, a transport time ηT with $\eta = 2/5$, and a total distance between static tweezers $d = 7$ μm. The start and end of the transport stage are indicated by gray dashed vertical lines. (a) Piecewise amplitude ramp consisting of four stages: transferring the atom from the initial static tweezer to the moving one, transport, transferring the atom from the moving tweezer to the target static one, and a final waiting or measuring time, see Eq. (8); (b) Position of the moving tweezer as a function of time for a piecewise-linear (magenta), minimum-jerk (blue), and quadratic (yellow dashed line) pulses [see Eqs. (4)–(6)]. The linear and minimum-jerk trajectories are particular cases of the more general hybrid pulse given in Eq. (7), which is also depicted for hybridizations $\xi = 0.2$, 0.4, 0.6, and 0.8 (curves with color gradients from blue to magenta; see insets for details of the position control at the beginning of the transport stage). The velocity of the moving tweezer at the beginning and end of the transport interval decreases with ξ and is minimum for the minimum-jerk pulse. (c) Amplitude shape calculated within the STA approach, showing the full numerical solution (solid turquoise curve) and the approximation obtained using Eq. (21) (dashed orange line). The piecewise-linear amplitude associated with the piecewise-linear, hybrid, minimum-jerk, and quadratic ramps is shown as a light gray curve to highlight that the maximum amplitude is the same for all pulses. The two peaks at maximum amplitude depicted by the STA pulse are developed to counteract the restoring force of the static tweezers. (d) Full numerical and approximated position of the moving tweezer obtained with the STA formalism, same color code as in panel (c).

moving harmonic oscillator with a time-dependent effective frequency

$$H_0(x,t) = -\frac{\hbar}{2m}\frac{\partial^2}{\partial x^2} + \frac{m\omega(t)^2}{2}[x - x_0(t)]^2, \tag{10}$$

where $\omega(t)$ and $x_0(t)$ are related to the controls and the derivatives of the static potential through

$$m\omega^2(t) = \left.\frac{d^2V_{\rm st}}{dx^2}\right|_{x_{\rm mt}(t)} + \frac{A_{\rm mt}(t)}{\sigma_{\rm mt}^2}, \tag{11}$$

and

$$x_0(t) = x_{\rm mt}(t) - \frac{\left.\dfrac{dV_{\rm st}}{dx}\right|_{x_{\rm mt}(t)}}{m\omega^2(t)}. \tag{12}$$

For the time-dependent Hamiltonian of Eq. (10), it is possible to construct a dynamical invariant $I(x,t)$ such that for any wave function $\psi(t)$ evolving with $H_0(x,t)$, we have $(\mathrm{d}/\mathrm{d}t)\langle\psi(t)|I(t)|\psi(t)\rangle = 0$. This means that with the appropriate time dependence of $\omega(t)$ and $x_0(t)$, if the system is initially in an eigenstate of $H_0(t_{\rm i})$, the time evolution will map it onto the corresponding eigenstate of $H_0(t_{\rm f})$ (see Appendix A for details). For our particular form of $H_0(x,t)$, the dynamical invariant involves two auxiliary functions $\alpha(t)$ and $\rho(t)$ that satisfy

$$x_0(t) = \frac{\ddot{\alpha}(t)}{\omega^2(t)} + \alpha(t) \quad \text{and} \tag{13}$$

$$\omega^2(t) = \frac{\omega_0^2}{\rho^4(t)} - \frac{\ddot{\rho}(t)}{\rho(t)}, \tag{14}$$

where ω_0 is a free constant. From the first of these equations, we see that $\alpha(t)$ can be identified with a classical trajectory of a driven generalized classical harmonic oscillator [59], and $\rho(t)$ can be seen as a spatial rescaling factor. While Eqs. (13) and (14) ensure that $H_0(x,t)$ has fast transport modes given by the eigenstates of $I(t)$ up to a phase, we also need them to coincide with the eigenstates of $H_0(x,t)$ at the initial and final times. To fulfill the latter requirement, the two auxiliary functions must also satisfy the following boundary conditions:

$$\dot{\alpha}(t_{\rm i}) = \ddot{\alpha}(t_{\rm i}) = \dot{\alpha}(t_{\rm f}) = \ddot{\alpha}(t_{\rm f}) = 0 \quad \text{and} \tag{15}$$

$$\dot{\rho}(t_{\rm i}) = \ddot{\rho}(t_{\rm i}) = \dot{\rho}(t_{\rm f}) = \ddot{\rho}(t_{\rm f}) = 0. \tag{16}$$

The strategy to obtain the desired pulse is essentially reverse engineering. First, we choose $\alpha(t)$ and $\rho(t)$ satisfying the boundary conditions, which we then use to compute $x_0(t)$ and $\omega(t)$ using Eqs. (13) and (14). After that, we numerically solve Eq. (12) to find the position of the moving tweezer $x_{\rm mt}(t)$, which is finally plugged into Eq. (11) to obtain the tweezer amplitude $A_{\rm mt}(t)$.

A simple way to design functions satisfying boundary conditions on the derivatives such as those of Eqs. (15) and (16) is to look for an appropriate polynomial function [15,52,53,60]. This polynomial interpolation has the advantage that most of the calculations can be carried out

analytically. As already mentioned, the lowest-order polynomial in a variable s having vanishing first and second derivative at the initial and final points ($s = 0, 1$) is given by the minimum-jerk polynomial $p_{\rm mj}(s) = 10s^3 - 15s^4 + 6s^5$ (notice that $p_{\rm mj}(0) = 0$ and $p_{\rm mj}(1) = 1$); therefore, we choose both $\alpha(t)$ and $\rho(t)$ to have the functional form of $p_{\rm mj}$ with suitable multiplicative and additive constants. Because our transport protocol is divided into stages, we apply the reverse-engineering process to each time interval and require continuous outputs. In other words, for each stage, we propose $\alpha(s) = \alpha_{\rm i} + p_{\rm mj}(s)(\alpha_{\rm f} - \alpha_{\rm i})$ and $\rho(s) = \rho_{\rm i} + p_{\rm mj}(s)(\rho_{\rm f} - \rho_{\rm i})$, where $s = t/(t_{\rm f} - t_{\rm i})$ is the dimensionless time and the i and f subscripts denote initial and final values over the considered interval. Using this form for the auxiliary functions, Eqs. (15) and (16) are satisfied for any $\alpha_{\rm i,f}$ and $\rho_{\rm i,f}$, which are to be fixed by the initial and the target position and by requiring the physical controls to be continuous functions of time.

We now turn to the calculation of the auxiliary functions. As mentioned before, for the proposed form for $\alpha(t)$, it is easy to see that $\ddot{\alpha}(t)$ vanishes at the initial and the final time of each stage, i.e., for $t = 0$, $(1-\eta)T/3$, $(1+2\eta)T/3$, $(2+\eta)T/3$, and T; therefore, for those values of time, we have $\alpha(t) = x_0(t)$. Since the position of the moving optical tweezer $x_0(t)$ only changes during the transport interval, the calculation of $\alpha(t)$ is straightforward and leads to

$$\alpha(t) = d\begin{cases} 0 & \text{for } 0 \le t \le \dfrac{1-\eta}{3}T, \\ 10\left(\dfrac{t-\frac{(1-\eta)}{3}T}{\eta T}\right)^3 - 15\left(\dfrac{t-\frac{(1-\eta)}{3}T}{\eta T}\right)^4 + 6\left(\dfrac{t-\frac{(1-\eta)}{3}T}{\eta T}\right)^5 & \text{for } \dfrac{1-\eta}{3}T \le t \le \dfrac{1+2\eta}{3}T, \\ 1 & \text{for } \dfrac{1+2\eta}{3}T \le t \le \dfrac{2+\eta}{3}T, \\ 1 & \text{for } \dfrac{2+\eta}{3}T \le t \le T. \end{cases} \tag{17}$$

To compute the auxiliary function $\rho(t)$, it is useful to define the frequency associated with the static traps

$$\omega_{\rm st} = \sqrt{\frac{1}{m}\left.\frac{d^2V_{\rm st}}{dx^2}\right|_{x=0,d}} = \sqrt{\frac{A_{\rm st}}{m\sigma_{\rm st}^2}}, \tag{18}$$

and the maximum frequency for the moving tweezer over the capturing and releasing intervals, which is the same as the initial frequency in the transport interval and is given in terms of the maximum depth of the moving tweezer over the capturing (c) or releasing (r) stages $A_{\rm mt}^{\rm max,cr}$:

$$\omega_{\rm mt}^{\rm max,cr} = \sqrt{\frac{A_{\rm mt}^{\rm max,cr}}{m\sigma_{\rm mt}^2}}. \tag{19}$$

Notice that in the case of the experimentally inspired pulses described in the previous section, $A_{\rm mt}^{\rm max,cr}$ matches the global maximum amplitude of the pulse $A_{\rm mt}^{\rm max}$. At the beginning of the capturing stage, at the end of the releasing interval, and at the end of the protocol (i.e., for $t = 0$, $(2+\eta)T/3$, and T), the moving tweezer is completely off, and we therefore have $\omega(t) = \omega_{\rm st}$ [see Eq. (11)].

In contrast, at the end of the capturing interval and at the beginning of the releasing interval, i.e., for $t = (1-\eta)T/3$ and $(1+2\eta)T/3$, the amplitude of the moving tweezer is at its maximum value over the capturing or releasing stage, meaning that $\omega(t) = \sqrt{\omega_{\rm st}^2 + (\omega_{\rm mt}^{\rm max,cr})^2}$. Using this in Eq. (14), we get

$$\rho(t) = \frac{\sqrt{\omega_0}}{\sqrt{\omega_{\rm st}}}\begin{cases} 1 + \left(\frac{1}{\sqrt[4]{\tilde{\omega}^2+1}}-1\right)\left\{10\left(\frac{t}{\frac{(1-\eta)}{3}T}\right)^3 - 15\left(\frac{t}{\frac{(1-\eta)}{3}T}\right)^4 + 6\left(\frac{t}{\frac{(1-\eta)}{3}T}\right)^5\right\} & \text{for } 0 \le t \le \frac{1-\eta}{3}T, \\ \frac{1}{\sqrt[4]{\tilde{\omega}^2+1}} & \text{for } \frac{1-\eta}{3}T \le t \le \frac{1+2\eta}{3}T, \\ \frac{1}{\sqrt[4]{\tilde{\omega}^2+1}} - \left(\frac{1}{\sqrt[4]{\tilde{\omega}^2+1}}-1\right)\left\{10\left(\frac{t-\frac{(1+2\eta)}{3}T}{\frac{(1-\eta)}{3}T}\right)^3 - 15\left(\frac{t-\frac{(1+2\eta)}{3}T}{\frac{(1-\eta)}{3}T}\right)^4 + 6\left(\frac{t-\frac{(1+2\eta)}{3}T}{\frac{(1-\eta)}{3}T}\right)^5\right\} & \text{for } \frac{1+2\eta}{3}T \le t \le \frac{2+\eta}{3}T, \\ 1 & \text{for } \frac{2+\eta}{3}T \le t \le T, \end{cases} \tag{20}$$

with $\tilde{\omega}^2 = (\omega_{\rm mt}^{\rm max,cr}/\omega_{\rm st})^2 = A_{\rm mt}^{\rm max,cr}\sigma_{\rm st}^2/(A_{\rm st}\sigma_{\rm mt}^2)$. Now that we have the expression for $\rho(t)$, the calculation of $\omega(t)$ is straightforward via Eq. (14). The resulting $\omega(t)$ can be used together with $\alpha(t)$ in Eq. (13) to obtain $x_0(t)$. It is important to note that since $\rho(t) \propto \sqrt{\omega_0}$, and also because $\alpha(t)$ does not depend on this quantity, $\omega(t)$ (and as a consequence $x_0(t)$) are independent of ω_0. After that, the obtained $x_0(t)$ is plugged into Eq. (12), and the equation is numerically solved to get $x_{\rm mt}(t)$, which we then use in Eq. (11) to derive the amplitude control $A_{\rm mt}(t)$.

The numerical solutions for the amplitude and the position of the bottom of the moving trap are shown in Figs. 2(c) and 2(d) as solid turquoise curves. To compare these engineered solutions with the pulses described in the previous section, we superimpose the piecewise-linear pulse and the piecewise-linear amplitude as light-gray solid curves. Note that the inset of Fig. 2(d) provides an enlarged view of the pulses for times between 0.5 and 1 ms and positions between 0 and 1 μm. While $x_0(t)$ strongly resembles the minimum-jerk trajectory, the amplitude presents two peaks that counteract the effect of the static tweezers during the atom-transport task and are directly related to the second-derivative term in Eq. (11) (see also Appendix A). For a general problem, the quantity and spacing of these peaks will be given by both the number of static traps and the spatial separation between them. In most experimental implementations, the power of the laser that generates the moving tweezer is kept constant during the movement (see for instance Ref. [11] or Fig. 9 of Ref. [19]), as we already consider for the piecewise-linear, hybrid, minimum-jerk, and quadratic ramps; however, in the supplemental material of Ref. [8], the authors mention that when transferring from the static to the moving tweezers, they use a quadratic intensity profile, highlighting the possibility of implementing our STA amplitude pulse.

Before moving to the next section, we provide a quite simple and directly applicable analytical approximation for the STA controls. The approximated solutions can be obtained using a quadratic approximation for the external potential, $V_{\rm st}(x) \approx -A_{\rm st}(1-2x/d)^2$, which is only valid between the two minima, i.e., for $0 \le x \le d$. Within this approximation, we get

$$x_{\rm mt}(t) \approx \frac{x_0(t) + \frac{4A_{\rm st}}{dm\omega^2(t)}}{1 + \frac{8A_{\rm st}}{d^2 m\omega^2(t)}}, \tag{21}$$

an expression that when combined with Eq. (11) leads to an approximated solution for $A_{\rm mt}(t)$. It is important to note that, for this approximated $x_{\rm mt}(t)$, the boundary conditions are also approximately fulfilled. In the transport interval, the approximated solution can be written as

$$x_{\rm mt}(t) \approx d\,\frac{10\left(\frac{t-\frac{1-\eta}{3}T}{\eta T}\right)^3 - 15\left(\frac{t-\frac{1-\eta}{3}T}{\eta T}\right)^4 + 6\left(\frac{t-\frac{1-\eta}{3}T}{\eta T}\right)^5 + \frac{60\left(\left(\frac{t-\frac{1-\eta}{3}T}{\eta T}\right) - 3\left(\frac{t-\frac{1-\eta}{3}T}{\eta T}\right)^2 + 2\left(\frac{t-\frac{1-\eta}{3}T}{\eta T}\right)^3\right)}{(\eta T)^2\left(\frac{A_{\rm st}}{m\sigma_{\rm st}^2} + \frac{A_{\rm mt}^{\rm max,cr}}{m\sigma_{\rm mt}^2}\right)} + \frac{4A_{\rm st}}{d^2\left(\frac{A_{\rm st}}{\sigma_{\rm st}^2} + \frac{A_{\rm mt}^{\rm max,cr}}{\sigma_{\rm mt}^2}\right)}}{1 + \frac{8A_{\rm st}}{d^2\left(\frac{A_{\rm st}}{\sigma_{\rm st}^2} + \frac{A_{\rm mt}^{\rm max,cr}}{\sigma_{\rm mt}^2}\right)}}. \tag{22}$$

A direct evaluation of the coefficients of the second and third terms of the numerator, as well as the second term in the denominator, leads to values below 10^{-3}, meaning that the STA trajectory closely resembles the minimum-jerk one. Importantly, with the same Taylor expansion, we also derived an approximation for the maximum amplitude of the STA pulse in terms of the maximum amplitude over the capturing or releasing stage (or, equivalently, the initial amplitude in the transport interval), $A_{\rm mt,STA}^{\rm max} \approx A_{\rm mt}^{\rm max,cr} + A_{\rm st}(1 + 2/e^{3/2})\sigma_{\rm mt}^2/\sigma_{\rm st}^2 \ge A_{\rm mt}^{\rm max,cr}$. The latter expression allows us to compare pulses with the same global maximum amplitude, which is the relevant experimental parameter. The approximations obtained for the STA control pulses using the same experimentally realistic values for the parameters as in the previous sections are depicted in Figs. 2(c) and 2(d) as dashed orange lines. We checked that the approximation for the maximum amplitude of the STA ramp presents a very good agreement with the full numerical results, and also that the relative error between the full numerical solution and the approximation for the amplitude is less than 4% during the transport interval.

In a recent work, Hwang *et al.* experimentally demonstrated the advantage of an STA-based trajectory over the constant-velocity and constant-jerk trajectory [61]. Also using STA techniques and a harmonic truncation of the tweezer potential, Jaewook Ahn's team developed a trajectory very similar to that given in Eqs. (21) and (22),

and they demonstrated that the STA survival probability of atoms after transportation outperforms the non-STA ones. We would like to highlight that even though our approaches are similar, there are two key differences: first, their calculations consider only the moving tweezer and not the static ones, and second, they do not develop the amplitude pulse to make use of the depth of the tweezer as a second control variable. To compare both STA approaches, we simulate the transport protocol using the trajectory described in Ref. [61] for a total distance of 7 μm and $A_{\mathrm{mt}}^{\max}/\hbar = 3.57 \times 2\pi$ MHz. For a protocol time of 2.5 ms, we obtain an infidelity of 1.5×10^{-5}, to be compared with the minimum-jerk infidelity of 1.4×10^{-5} and the one for our full STA solution of 5.3×10^{-6}; moreover, the infidelity obtained with our STA approach is one order of magnitude below the infidelity reported in Ref. [61] for total pulse times in the range $[0.01, 3]$ ms. As we show in the next section, incorporating a modulation into the amplitude of the moving tweezer that counteracts the restoring force of the static traps allows the STA pulse to outperform the experimentally motivated ones, even without optimization. We trust that the simple approximations for our STA solutions provided in this section will motivate their experimental implementation in the near future [62].

IV. ATOM TRANSPORT CHARACTERIZATION AND OPTIMIZATION

In this section, we present a performance analysis of the considered pulses and their subsequent optimization.

A. Performance of the closed-form pulses

To evaluate the performance of the transport of a single atom under the piecewise-linear, hybrid, minimum-jerk, quadratic, and STA pulses, we are mainly concerned with two features: first, that the transport is faithful (high-fidelity condition), and second, that the vibrational excitations of higher states are reasonably bounded to prevent atom losses [63,64]. To check the first condition, we analyze the transport infidelity. As explained in Sec. II, the infidelity quantifies the error after the transport, and it is closer to zero when the state of the atom is closer to the target state. Since our target state is the ground state of the target tweezer, a lower infidelity at the end of the transport is equivalent to limited heating after the complete transport protocol.

Since the potential of an optical tweezer has finite depth, it is also necessary to assess whether the no-heating condition is satisfied during transport. As discussed in Ref. [64], the main effect of heating is to expel the atom from the trap not as a result of an increase in the mean energy but as a consequence of the spreading of the width of the distribution over the energy states (the physical picture is that when the upper tail of the distribution reaches untrapped levels, the atom is lost). Based on this argument, to quantify the increase in the vibrational quantum number, we calculate the mean value $\langle N \rangle$ and the width ΔN of the distribution of the atomic state over the instantaneous eigenstates of the moving tweezer. Following Ref. [65], we also calculate an effective temperature given by the expectation value of the kinetic operator $T_{\mathrm{eff}} = 2\langle K \rangle / k_B$, with k_B denoting the Boltzmann constant and $K = p^2/(2m)$, where p is the momentum operator. By means of the virial theorem for a harmonic trap, we have $\langle H \rangle = 2\langle K \rangle$ [66]. Thus, T_{eff} is a measure of the mean energy of the system, which is usually reported to analyze the stability of the system during transport [25,56]. In our case, the maximum depth of the moving tweezer during transport hosts about 55 oscillator levels, and the deviation of the lower states with respect to the harmonic ones is small. In the following, we use Système International (SI) units and characteristic units for the time; from the frequency of the static potential [see Eq. (18)], we can define a characteristic time as $\tau_{\mathrm{st}} = 2\pi/\omega_{\mathrm{st}} \approx 0.03$ ms. We use this value, since it is fixed for any maximum amplitude of the moving tweezer; however, the relationship between this value and the characteristic time during transport is $\tau_{\mathrm{mt}} = 2\pi/\omega_{\mathrm{mt}}^{\max} \approx \tau_{\mathrm{st}}/2$.

The infidelity as a function of the total time T is reported in Fig. 3 for the piecewise-linear, hybrid ($\xi = 0.8, 0.4$), minimum-jerk, quadratic, and STA (full numerical solution and approximation) pulses from left to right. We consider total times ranging between 0.01 and 3 ms, in agreement with the experimental values (see for instance Ref. [10] or Ref. [19]). We calculate the maximum (solid curve), average (dashed line), and last value (dotted line) of the infidelity over the final interval of the pulse. Since it constitutes an upper bound, we propose the maximum value as the quantity to be used when comparing with experimental results. The shaded area corresponds to one standard deviation as a measure of the error in the fidelity.

As expected, the piecewise-linear pulse produces the highest error after transport (with an infidelity higher than 10^{-2} for all T) and sustained oscillations (with a period of about $12\,\tau_{\mathrm{st}}$). The hybrid-pulse behavior interpolates between the minimum-jerk and the linear behavior and presents more oscillations for higher ξ (i.e., going toward the linear ramp, as expected). The quadratic-pulse infidelity has a similar behavior to that of the hybrid pulse with $\xi \lesssim 0.5$ but with a slightly higher oscillatory behavior. Among all the pulses, the minimum-jerk and the STA ones present the smoothest and most stable behavior. Moreover, the STA pulse also enables reaching the lowest infidelity in the shortest time; the infidelity rapidly decreases for T between 10 and $20\,\tau_{\mathrm{st}}$ and then reaches a stable value around 10^{-4}. In general, we observe that all the pulses fail in the transport task for $T \lesssim 10\,\tau_{\mathrm{st}}$. For total times greater than $10\,\tau_{\mathrm{st}}$, the infidelity decreases when the total time increases, and the hybrid, minimum-jerk, quadratic,

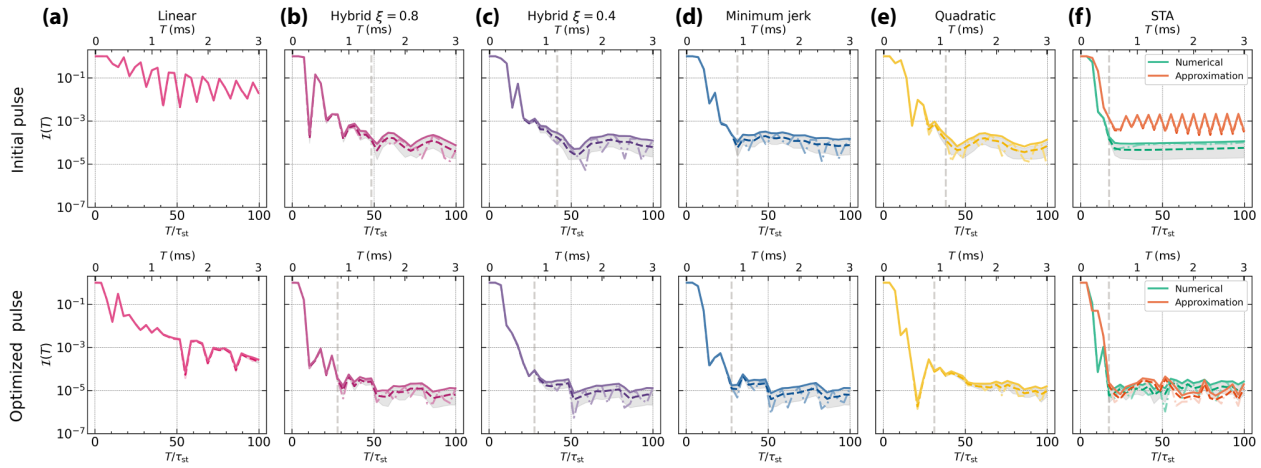

FIG. 3. Error after transport. Infidelity $\mathcal{I}$ versus total time T for the initial (top) and optimized (bottom) pulses with: (a) piecewise linear; hybrid with (b) $\xi = 0.8$ and (c) $\xi = 0.4$; (d) minimum jerk; (e) quadratic; and (f) STA. All the pulses have a maximum amplitude of $A_{\rm mt}^{\rm max}/\hbar = 3.57 \times 2\pi$ MHz, the transport time is given by $\eta\, T$ with $\eta = 2/5$, and the total distance between static tweezers is $d = 7$ μm. The maximum, average, and last value of the infidelity over the last stage of the protocol [see Eqs. (8) and (9)] are shown as solid, dashed, and dotted lines, respectively. The shaded area indicates the standard deviation calculated over the same time interval and quantifies the error in the infidelity. The vertical gray dashed line highlights the value of the total time for which the infidelity reaches the threshold of 10^{-4}. For the hybrid pulse with $\xi = 0.8$ and 0.4, minimum-jerk, quadratic, and STA ramps, the corresponding times are approximately 48 $\tau_{\rm st}$, 41 $\tau_{\rm st}$, 31 $\tau_{\rm st}$, 38 $\tau_{\rm st}$, and 18 $\tau_{\rm st}$, with $\tau_{\rm st}$ being the characteristic time of the static traps (the characteristic time during transport is $\tau_{\rm mt} \approx \tau_{\rm st}/2$). For the two hybrid and minimum-jerk optimized pulses, these times decrease to 28 $\tau_{\rm st}$, the time for the quadratic ramp decreases to 31 $\tau_{\rm st}$, the STA threshold time does not change, and the piecewise-linear optimized pulse reaches the infidelity threshold for a total time of approximately 55 $\tau_{\rm st}$. For the considered total pulse times and distance, the range for the mean velocity during transport goes from 1.75 m/s at a total protocol time of 0.01 ms to 5×10^{-3} m/s for a total time of 3 ms.

and STA pulses present two more regimes: a diabatic regime with some oscillations between $T \approx 10\,\tau_{\rm st}$ and a value in the range 20–50 $\tau_{\rm st}$, and the adiabatic regime for sufficiently large T. The total time at which the second regime ends depends on the pulse and corresponds in increasing order to the ramps for STA, minimum jerk, hybrid with $\xi = 0.4$, quadratic, and hybrid with $\xi = 0.8$. After that, the fidelity saturates with some smooth oscillations that improve when the upper time threshold of the diabatic regime decreases. Interestingly, the STA approximation captures the two first regimes well, including the value of the diabatic threshold, but it presents strong oscillations with a period of about 8 $\tau_{\rm st}$ in the last adiabatic regime. For total times $T \gtrsim 50\,\tau_{\rm st}$, the hybrid, minimum-jerk, quadratic, and STA pulses reach the adiabatic regime, and therefore, any of those pulses can faithfully perform the transport task.

Since the STA ramp presents a clear advantage for short T, we conclude that our proposed STA protocol has the best performance, achieving the infidelity threshold of 10^{-4} (vertical dashed line) in nearly half of the time taken by the minimum-jerk trajectory (which is the second-best one). Moreover, using the fidelity as a measure of the probability of a faithful transport process between two tweezers and assuming independent transport processes (see Ref. [61]) between tweezers, we can estimate the fidelity after a transport protocol involving $n_{\rm st}$ traps. For instance, using the infidelity obtained for our STA solution for a typical experimental transport velocity of 10 nm/μs [17], we obtain an infidelity of 10^{-2} after 30 consecutive independent transport movements. In Appendix C, we present examples of the evolution of the infidelity for total pulse durations $T = 0.5$, 1.5, 2.5, and 3 ms, while in Appendix D, we report the infidelities for various total distances and total pulse times, as well as their dependence on the velocity.

To test the performance of the STA solutions in a different experimental setting, we compute the pulses and simulate the evolution using the parameter values described in Ref. [17]. Our proposed STA solutions (both fully numerical and approximate) allow us to complete the transport and transfer between two neighboring tweezers separated by a distance of 5 μm in 160 μs with an infidelity of about 10^{-4}. Comparing this time to a total time of 1.25 ms reported in Ref. [17], our pulse performs the task 7.8 times faster, even when the transport stage is carried out more slowly (for us, it takes 80 μs, while in the experiments of Ref. [17], it takes 50 μs). More strikingly, our STA protocols reduce the transfer time devoted to capturing or releasing from and to static traps by a factor of 15 (in our case 40 μs, versus 600 μs for the experiments reported in Ref. [17]). We would like to stress that these results could

be further improved by adapting the ratio between the time devoted to the transport and transfer stage to parameters of the system at hand.

Figure 4 depicts the maximum expected value $\max(\langle N\rangle)$ and the associated uncertainty or width of the distribution $\max(\Delta N)$ for the occupied states of the moving tweezer, together with the maximum effective temperature $\max(T_{\text{eff}})$ as a function of the total pulse time T [67]. As expected, both quantities increase for shorter T because diabatic processes induce a higher mixture of states. The increase in the vibrational modes shows a similar behavior for all the considered pulses; for total times above 0.5 ms,

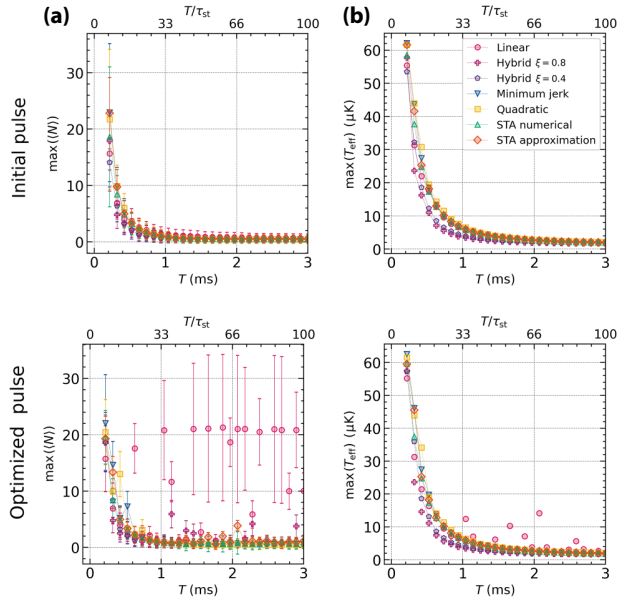

FIG. 4. Upper-bound measures of heating during transport. (a) Maximum of the mean occupied level (points) and width (error bars) of the distribution over the moving-tweezer states (denoted respectively by $\langle N\rangle$ and ΔN). (b) Maximum effective temperature T_{eff} over the complete duration of the pulse as a function of the total time T for the initial pulses (top row) and the optimized ones (bottom row). The measures obtained for the piecewise-linear, hybrid with $\xi = 0.8$ and $\xi = 0.4$, minimum-jerk, quadratic, and STA (fully numerical solution and approximation) ramps are depicted as magenta circles, burgundy crosses, purple pentagons, blue down-pointing triangles, yellow squares, turquoise up-pointing triangles, and orange diamonds, respectively, with a line of the same color as a guide for the eye. For a total pulse time of $T \approx 0.22$ ms $\approx 8\,\tau_{\text{st}}$, the mean occupied level plus its uncertainty is equal to half of the states hosted by the moving tweezer. All the initial pulses have a maximum amplitude of $A_{\text{mt}}^{\max}/\hbar = 3.57 \times 2\pi$ MHz, a transport time of ηT with $\eta = 2/5$, and a total distance $d = 7$ μm. As before, τ_{st} denotes the characteristic time of the static traps, and this is related to the characteristic time during transport via $\tau_{\text{mt}} \approx \tau_{\text{st}}/2$. Given the ranges for the total pulse time and distance, the mean velocity during transport lies between 1.75 m/s for a total protocol time of 0.01 ms to 5×10^{-3} m/s for a total time of 3 ms.

the wave function has a dominant weight over the ground state of the moving trap and smaller weights over the first and second excited states. For total pulse times shorter than 0.5 ms, the expectation value of the occupied moving tweezer states $\langle N\rangle$ and its uncertainty ΔN increase rapidly for decreasing total times. For $T \approx 0.22$ ms $\approx 8\,\tau_{\text{st}}$, the mean occupied level plus the associated uncertainty reaches half of the trap states, i.e., 27 levels for our trap hosting a total of 55 states. An estimation of the vibrational number increase for a moving harmonic oscillator can be obtained by evaluating the Fourier transform of the acceleration at the frequency of the oscillator [1,56,68]. By computing the Fourier transform of the acceleration of our approximated STA trajectory [Eq. (22)], we find an agreement with less than 10% of relative difference between the time at which our protocol reaches an increase of 27 oscillator levels and the time that would be needed by the STA solution proposed by Jaewook Ahn's group; see Eq. (7) of Ref. [61].

The hybrid ramps demonstrate a slight advantage when compared to the rest of the pulses, as they present a smaller effective temperature for total times $T \lesssim 1.5$ ms. Furthermore, the difference in the effective temperature among the hybrid ramps and the other pulses increases for shorter total times, reaching a difference of about 22 μK for $T \approx 0.3$ ms between the quadratic and the hybrid ramp with $\xi = 0.8$. Since the change of the energy after the evolution of a moving harmonic oscillator for a particle initially in one of the oscillator states is given by the Fourier transform of the acceleration evaluated at the frequency of the oscillator [1,56,68], a purely linear pulse does not induce heating during transport. The incorporation of the linear part in the hybrid trajectory presented in Ref. [10] uses the latter advantage and at the same time reduces the error after transport (as discussed on the basis of Fig. 3, the piecewise-linear pulse infidelity is more than 2 orders of magnitude higher when compared to the other pulses) induced by the velocity discontinuities at the endpoints of the pulse by replacing them with the minimum-jerk subintervals. In Appendix C, we report $\langle N\rangle$, ΔN, and T_{eff} as functions of time for total protocol times $T = 0.5$, 1.5, 2.5, and 3 ms. Here, it is possible to observe that for $T = 0.5$ ms, the hybrid pulse with $\xi = 0.8$ has the smaller ΔN and T_{eff}, followed by the hybrid pulse with $\xi = 0.4$.

B. Atom-transport optimization

To further reduce the transport error, we implement an optimization protocol using the d-CRAB algorithm introduced by Rach *et al.* in 2015 [33–35,69]. This algorithm offers several advantages over other optimization methods, mainly because of its efficiency managing high-dimensional control spaces; by expanding the control pulses in a randomized basis, it reduces the dimensionality of the search space while circumventing local traps that

could arise due to this restriction. Since within d-CRAB, only a small number of parameters are optimized at the same time, the need for gradient information is mitigated [70,71]. This makes it possible to perform adaptive optimal pulse shaping using experimental data, which implicitly allows experimental uncertainties—such as varying parameters and instrumentation transfer functions [72]—to be taken into account in the optimization cycle. We observe that, in practice, the flexibility offered by d-CRAB comes at the price of increased dependence on the initial guess for the controls; this is likely the result of a more local search in the control space. In light of this, the in-depth discussion of the relevant experimentally feasible pulses and the STA protocol that we presented in Sec. III becomes especially important.

We use the QuOCS library [32], which incorporates, among others, the d-CRAB algorithm in a user-friendly interface [35]. The QuOCS library allows for a straightforward setting of the parameters related to the pulse, such as the total time and basis parameters. It also allows the implementation of scaling functions that set appropriate limits on the position and amplitude control. We use an expansion of the control pulse in a basis of sigmoid functions and select only some of the coefficients as optimization control. We then conduct a systematic pulse optimization involving four different optimization schemes. Following the methodology outlined in Ref. [18], we optimize the trajectory for different total times and adjust the initial controls to identify the shortest achievable time during the numerical analysis.

The infidelity after transport as a function of the total time T for all the considered pulses is shown in Fig. 3 (bottom row). The optimization improves the fidelity by 2 orders of magnitude for the linear pulse and by one order of magnitude for the remaining pulses. For all the pulses, the shaded area corresponding to one standard deviation in the error over the last waiting interval of the pulse is reduced by the optimization, pointing toward a much more stable state after transport (see also Appendix C). The optimized piecewise-linear pulse reaches the infidelity value of 10^{-4} for a total time pulse of 55 τ_{st}. This threshold time remains at 18 τ_{st} for the STA pulse, not changing appreciably after the optimization. The threshold time for the remaining pulses decreases between 10% and 30%; it changes from 48 τ_{st} to 28 τ_{st} for the hybrid pulse with $\xi = 0.8$, from 41 τ_{st} to 28 τ_{st} for the hybrid ramp with $\xi = 0.4$, from 32 τ_{st} to 28 τ_{st} for the minimum-jerk ramp, and from 38 τ_{st} to 31 τ_{st} for the quadratic pulse (all these times are indicated with dashed gray vertical lines). We conclude that the optimization yields a significant improvement for all the pulses; however, we would like to highlight that the full numerical solution for the STA protocol shows highly desirable features, even without optimization. In particular, our nonoptimized STA solution implies an improvement of 42% in the total time required to reach the fixed infidelity threshold of 10^{-4} with respect to the nonoptimized minimum-jerk pulse. In the case of the optimized ramps, the STA solution improves that time by 36% with respect to the optimized minimum-jerk pulse.

Since the objective of the optimization is to minimize the final infidelity, there is no guarantee that high-energy instantaneous eigenstates are not populated during the obtained transport dynamics, allowing the particle to escape the trap. For this reason, it is very important to check that the optimized pulses does not induce heating during the transport. As can be seen in Fig. 4 (bottom row), the optimization leads to higher vibrational excitations for the piecewise-linear and hybrid pulse with higher hybridization ($\xi = 0.8$). This feature is revealed when inspecting the effective temperature, but it is magnified by the behavior of the mean occupied level and the width of the distribution over the oscillator states, in line with the discussion presented in Ref. [64]. Our results suggest avoiding the use of piecewise-linear pulses, which are known to induce heating due to their intrinsic velocity discontinuities; this is consistent with the good fidelity obtained in Ref. [19] using a pulse with constant velocity in the central 10% of the transport interval ($\xi = 0.1$). The evolution of the infidelity, $\langle N \rangle$, ΔN, and T_{eff} is shown in Appendix C for total times $T = 0.5$, 1.5, 2.5, and 3 ms. Appendix C also contains a comparison between some examples of initial and optimized pulses. We would like to mention that a very small change in the pulse can translate into a considerable improvement of the fidelity, as was already exposed from the difference in the performance of the full numerical solution for the STA pulse when compared to the behavior of the approximated solution. Since all the optimized pulses except for the piecewise-linear one show limited heating, our results suggests that the use of d-CRAB for pulse shaping would not induce collateral heating that might cause major atom loss during transport.

As a final goal, we turn to the determination of good regions in the parameter space where the transport protocol can be reliably implemented in current experiments. We therefore focus on the two parameters that can be changed easily in the experiments, namely the total time of the pulse and the maximum amplitude (depth) of the moving tweezer. In fact, while the amplitude can be experimentally adjusted by manipulating the laser power, other parameters (such as the width of the tweezer) might be challenging to modify since they are ultimately determined by the optical elements within the apparatus. To identify good intervals for the amplitude and total time to run the experiments, in Fig. 5, we present heat maps for the optimized infidelity (minimizing the maximum of $\mathcal{I}$ over the last waiting interval) for the two pulses that enable reaching the infidelity threshold of 10^{-4} in the shortest total time for a fixed amplitude of $A_{\mathrm{mt}}^{\mathrm{max}}/\hbar = A_{\mathrm{exp}}/\hbar = 3.57 \times 2\pi$ MHz,

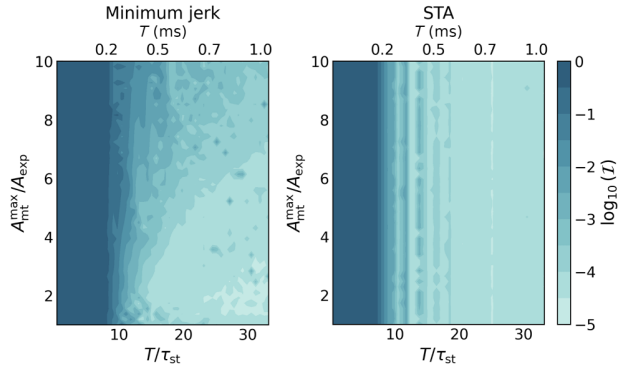

FIG. 5. Determining optimal regions in the experimentally accessible parameter space for high-fidelity transport pulses. The infidelity is shown for the minimum-jerk ramp (left) and our proposed STA pulse (right) after optimization as a function of the maximum amplitude $A_{\mathrm{mt}}^{\mathrm{max}}$. The amplitude is in units of $A_{\mathrm{exp}}/\hbar = 3.57 \times 2\pi$ MHz, and the total time of the pulse T is given in units of the characteristic time of the static traps τ_{st} (bottom scale) and in milliseconds (upper scale). Lighter colors show regions of smaller infidelity, which quantifies the error after transport. As before, the characteristic time during transport is given by $\tau_{\mathrm{mt}} \approx \tau_{\mathrm{st}}/2$ and the mean velocity during transport varies between 1.75 m/s for a total protocol time of 0.01 ms to 5×10^{-3} m/s for a total time of 3 ms.

i.e., the optimized minimum-jerk and STA pulses. We consider total times up to 1 ms $\sim 33\,\tau_{\mathrm{st}}$ (for larger total times, the infidelity reaches the stable regime) and maximum amplitudes between A_{exp} and $10\,A_{\mathrm{exp}}$, both being realistic ranges of the parameters for experimental realizations.

Our simulations show that the infidelity of the minimum-jerk pulse has a strong dependence on both the total time and the maximum amplitude. For larger amplitude values, longer times are needed to achieve a desired fixed value of the infidelity. We interpret this on the basis of the relation between the capturing and releasing time and the total time of the transport, which are here fixed to $T/5$ and $2\,T/5$ respectively. A higher amplitude of the moving tweezer implies a larger difference between the moving and static potential, meaning that the system requires a longer releasing time to adjust toward the target state. Fixing an error threshold of 10^{-4}, the best region in the parameter space appears for times longer than $20\,\tau_{\mathrm{st}}$ and amplitudes between 1 and $4\,A_{\mathrm{exp}}$. On the other hand, by incorporating a modulation in the amplitude of the moving tweezer that counteracts the restoring force of the static traps, the STA pulse depicts a quite robust infidelity against amplitude variations. Using our STA pulse, infidelities below 10^{-4} can be obtained for times longer than $20\,\tau_{\mathrm{st}}$ independently of $A_{\mathrm{mt}}^{\mathrm{max}}$. The STA ramps also seem to present *magic time windows*, exposed as lighter vertical stripes, indicating that the infidelity is suppressed by at least one order of magnitude around $T \approx 10.5\,\tau_{\mathrm{st}}$, $13.5\,\tau_{\mathrm{st}}$, $15.5\,\tau_{\mathrm{st}}$, $17.5\,\tau_{\mathrm{st}}$, and $25\,\tau_{\mathrm{st}}$. Such magic times have already been found in previous works addressing efficient atom transport. See for instance Refs. [25,31], where the authors combine reverse-engineering methods with quantum optimal control and design trajectories that show a strong suppression of the population across excited states for a given set of magic times [73].

Figure 5 shows a drastic increase in the infidelity for total pulse durations below $8\,\tau_{\mathrm{st}}$. This feature can already be inferred from Fig. 3 for all the considered pulses and for $A_{\mathrm{mt}}^{\mathrm{max}} = A_{\mathrm{exp}}$. Our results strongly point toward the presence of a QSL, imposing a bound for the minimum time required for the transport task [22,37]. Taking into account that the transport takes $2/5$ of the total time, T, the bound for the transport time is about $3\,\tau_{\mathrm{st}}$. When comparing with the results for a quadratic pulse presented in Ref. [20] for the same setting (with a distance in oscillator units and a trap depth hosting the same amount of states) we observe that our lower bound for the transport time is five times larger; however, it is important to highlight that our protocol incorporates more information-processing details than the one considered in Ref. [20], where only the transport stage is taken into account with no consideration of the time devoted to capture and release the atom. For a moving trap hosting the same number of levels, we obtain a lower time threshold for the transfer between tweezers, which is nine times faster than the time reported in Ref. [19] (eight oscillator units versus 75 oscillator units). Considering that in Ref. [17], the authors state that the time needed for the capturing or releasing is 12 times that required for the transport task, our results suggests that an improvement of the transfer between tweezers can translate into an overall speedup, even at the cost of devoting a larger time to the transport stage. Our findings call attention to the importance of including the transfer between tweezers in the model; by optimizing over the entire process, which also includes the effect of the static traps, we see the importance of tailored solutions, such as those based on our specialized STA, for reaching error rates in the fault-tolerant regime.

V. SUMMARY AND CONCLUSIONS

In this work, we focused on the transport and transfer between tweezers of neutral atoms, a relevant problem (and possibly a bottleneck) for the creation and improvement of quantum processors and simulators [1,8]. We considered four different types of experimentally motivated pulses: piecewise-linear, piecewise-quadratic, minimum-jerk, and a family of hybrid linear and minimum-jerk ramps. These pulses were used to transport not only atomic ensembles but also few and single atoms in experiments with optical tweezers [1,7,8,10–12,17] and optical conveyor belts [48]. To generate a pulse that is specifically tailored for our problem, we developed a transport protocol using STA techniques. The main difference between our STA approach and previous ones is that we take into

account the potential of the static tweezers to optimize the transfer between static and moving traps. To facilitate the adoption of our pulse in future experiments, we provided an easy-to-implement approximation for the full numerical solution. A close inspection of the approximated STA solution shows that the associated trajectory is a modification of the widely known minimum-jerk trajectory, with the key difference that its amplitude modulation (related to the depth of the moving trap) counteracts the effect of the static tweezers. We characterized the performance of all the considered pulses and found that our proposed STA protocol outperforms the experimentally inspired ramps, even without optimization.

All the considered pulses were used as initial guesses for QuOCS optimization, leading to a reduction in the error (quantified by the infidelity with respect to the target state) by at least one order of magnitude. By setting a threshold for the infidelity of 10^{-4}, we obtained a reduction of the total time of the pulse between 10% and 30% after optimization for the hybrid, minimum-jerk, and quadratic pulses.

Based on the discussions on heating and losses presented in Refs. [1,65], and to check that our optimized pulses do not induce extra heating during transport, we calculated the mean occupied level and the width of the distribution over the states of the moving tweezer, together with an effective temperature related to the motional energy. For total pulse times below $8\,\tau_{st} \approx 16\,\tau_{mt}$ (with τ_{st} and τ_{mt} being the characteristic times of the static and moving traps, respectively), the increase in the vibrational excitations exceeds half of the states hosted by the moving trap. We focused on the two pulses with the best performance, namely the minimum-jerk and STA pulses, and we quantified the error after transport as a function of the total time of the pulse and the maximum depth of the moving tweezer (which are both experimentally accessible parameters). We identified good regions in the parameter space to reliably run the transport protocol. For total times larger than $10\,\tau_{st}$, both pulses provide good performance (with an infidelity below 10^{-2}); however, the minimum-jerk ramp entails limitations on the maximum amplitude that can be selected, while our proposed STA pulse provides much more freedom. We conclude that our STA solution constitutes an improvement compared to the usual minimum-jerk trajectory.

Our findings also suggest the presence of a lower bound for the total time of the pulses; we obtain a numerical quantum speed limit for the complete protocol time of about $8\,\tau_{st}$. Since the transport time is a fraction of the total duration of the protocol, this value corresponds to a bound of $3\,\tau_{st}$ for the transport stage, which is around five times the value reported in Ref. [20] for a protocol that does not consider the transfer time required to capture and release the atom. Taking into account that the capturing or releasing time was estimated to be 12 times larger than the time needed for the transport process alone in the experiments of Ref. [17], and that our obtained transfer time is nine times faster than the one reported in Ref. [19], our results draw attention to the relevance of including the transfer between tweezers in the model. In other words, to fully capture the information protocol, it is necessary to consider not only the transport process but also the transfer between optical tweezers, which makes up a significant portion (if not the majority) of the error and the time budget.

Our analytical and numerical results show that small deformations in the pulses (for instance, less than 4% for the approximated STA amplitude when compared to the full numerical solution) can translate into a decrease in the error after transport by 2 orders of magnitude. We also contribute to filling a knowledge gap identified in Ref. [48] in the case of atomic transport by means of optical conveyor belts, namely, providing a systematic characterization of the effectiveness of different trajectories.

Our proposed STA pulse also demonstrates that a modulation of the depth of the tweezer (in contrast to the widely used piecewise-linear ramps) counteracts the effect of the static tweezers, producing smaller and more stable errors for shorter pulse durations. This adds extra controllability, similarly to amplitude and phase control in optical lattices [74]. To the best of our knowledge, our proposed amplitude modulation could be implemented in current experiments since, as mentioned in Ref. [8], amplitudes with a quadratic dependence on time are possible. The implementation of the modulation in the tweezer depth as a second control that corrects the effects of the static tweezers is particularly promising in light of the recent experimental implementation of an STA trajectory which has many common elements to ours [61], with the key difference of correcting the static-tweezer effects during the transfer process. Our results contribute to the generalization and improvement of previously known ramps for transporting neutral atoms in state-of-the-art quantum processors and quantum simulators based on tweezer arrays. Given the flexibility of our approach to accommodate different external potentials, we are currently working on its application to optical lattice platforms.

ACKNOWLEDGMENTS

We thank Thomas Reisser for technical support related to QuOCS. We thank Jan Reuter and Matteo Rizzi for discussions about numerical stability. We also thank Christian Groß, Philip Osterholz, Yu Hyun Lee, Peter Bojovic, and Titus Franz for valuable discussions about the experimental possibilities. We acknowledge funding from the German Federal Ministry of Education and Research through the funding program quantum technologies—from basic research to market under FermiQP Project No. 13N15891, under Horizon

Europe programme HORIZON-CL4-2022-QUANTUM-02-SGA via Project No. 101113690 (PASQuanS2.1) and under HORIZON-CL4-2021-DIGITALEMERGING-02-10 via Grant Agreement 101080085 QCFD, by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—Cluster of Excellence Matter and Light for Quantum Computing (ML4Q) EXC 2004/1—390534, and from the Jülich Supercomputing Center through the JUWELS and JURECA cluster. E.C. was supported by JSPS KAKENHI Grant No. JP23K13035.

DATA AVAILABILITY

The data that support the findings of this article are openly available [75].

APPENDIX A: STA THEORY

For the convenience of the reader, we regroup and summarize here the well-known results from the theory of shortcuts to adiabaticity (STA), which are relevant in the derivation of the STA-based transport pulse in Sec. III B.

1. Dynamical invariants and STA

Shortcuts to adiabaticity are a collection of methods whose aim is to obtain a fast-forward version of the adiabatic time evolution of a system. We exploit a known dynamical invariant of the system to reverse engineer an STA pulse that realizes the desired dynamics. Let us start by considering a time-dependent Hamiltonian $H(t)$ and the solutions $|\psi(t)\rangle$ to the Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = H(t)\,|\psi(t)\rangle\,. \tag{A1}$$

A dynamical invariant of $H(t)$ is a time-dependent operator $I(t)$ that satisfies the relation

$$\frac{dI}{dt} \equiv \frac{\partial I}{\partial t} + \frac{1}{i\hbar}[I,H] = 0, \tag{A2}$$

meaning that $(d/dt)\langle\psi(t)|I(t)|\psi(t)\rangle$. Provided that $I^\dagger = I$ is Hermitian, it is possible to write any solution of Eq. (A1) as a time-independent linear combination of eigenvectors $\left|\phi_n^{\mathrm{LR}}(t)\right\rangle$ of $I(t)$ [30], which we denote as dynamical modes

$$|\psi(t)\rangle = \sum_n c_n \left|\phi_n^{\mathrm{LR}}(t)\right\rangle, \quad \text{with}$$
$$I(t)\left|\phi_n^{\mathrm{LR}}(t)\right\rangle = \lambda_n \left|\phi_n^{\mathrm{LR}}(t)\right\rangle, \tag{A3}$$

and where we explicitly state that the eigenvalues λ_n of $I(t)$ do not depend on time. Additionally, in the case that $I(t)$ does not involve a differentiation in time, the dynamical modes can be explicitly constructed by means of a gauge transformation of the (generic) eigenvectors $|\phi_n(t)\rangle$ of $I(t)$ [30], featuring the so-called Lewis-Riesenfeld phases $\gamma_n(t)$:

$$\left|\phi_n^{\mathrm{LR}}(t)\right\rangle = e^{i\gamma_n(t)}\,|\phi_n(t)\rangle\,, \tag{A4}$$
$$\hbar\frac{d\gamma_n}{dt} = \langle\phi_n(t)|\,i\hbar\frac{\partial}{\partial t} - H(t)\,|\phi_n(t)\rangle\,. \tag{A5}$$

In addition to being a useful tool to solve the equation of motion [Eq. (A1)], these modes can be used to find the Hamiltonian operator that produces a certain prescribed time evolution by means of reverse engineering. The goal is to realize a time evolution that maps eigenstates of $H(t_\mathrm{i})$ to eigenstates of $H(t_\mathrm{f})$ in a given amount of time $T = t_\mathrm{f} - t_\mathrm{i}$ (the subscripts i and f denote, as usual, initial and final). One way to obtain this is to impose that every dynamical mode coincides (up to a phase factor) with an eigenstate of the Hamiltonian at $t = t_\mathrm{i}, t_\mathrm{f}$. This requires the dynamical invariant to commute with the Hamiltonian at the endpoints of the time evolution, leading to the following boundary conditions:

$$[I(t_\mathrm{i}),H(t_\mathrm{i})] = [I(t_\mathrm{f}),H(t_\mathrm{f})] = 0. \tag{A6}$$

These are just necessary conditions, but they allow us to effectively speed up the adiabatic transfer between ground states at $t = t_\mathrm{i}$ and $t = t_\mathrm{f}$ in concrete cases, such as for atom transport [15].

2. STA for atom transport

A general family of Hamiltonians that is useful for atom transport and manipulation with a known dynamical invariant is given by Ref. [58]:

$$H(t) = \frac{p^2}{2m} - F(t)x + \frac{m}{2}\omega^2(t)x^2 + \frac{V\left(\dfrac{x-\alpha(t)}{\rho(t)}\right)}{\rho^2(t)}. \tag{A7}$$

It is possible to verify that for any constant ω_0, the operator

$$I(t) = \frac{[\rho(p - m\dot\alpha) - m\dot\rho(x-\alpha)]^2}{2m} + \frac{m\omega_0^2}{2}\left(\frac{x-\alpha}{\rho}\right)^2 + V\left(\frac{x-\alpha}{\rho}\right) \tag{A8}$$

is an invariant for $H(t)$, meaning that it satisfies Eq. (A2) provided that the following auxiliary conditions are also

verified:

$$\ddot{\alpha} + \omega^2(t)\alpha = \frac{F(t)}{m} \quad \text{and} \tag{A9}$$

$$\ddot{\rho} + \omega^2(t)\rho = \frac{\omega_0^2}{\rho^3}\,. \tag{A10}$$

As proven in Ref. [58], there is a unitary transformation $|\phi\rangle = U(t)\,|\psi\rangle$ such that the transformed dynamical invariant $J = UIU^\dagger$, expressed in the coordinate $\zeta = \rho^{-1}(x-\alpha)$, is time independent. We can then obtain the eigenstates of I by solving the stationary eigenvalue problem for J and then transforming back with $U^\dagger$:

$$J\,|\phi_n\rangle = \lambda_n\,|\phi_n\rangle\,, \quad \text{with } |\psi_n(t)\rangle = U^\dagger(t)\,|\phi_n\rangle\,, \tag{A11}$$

which, in the coordinate space, gives rise to [15,58]

$$\left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial\zeta^2} + \frac{1}{2}m\omega_0^2\zeta^2 + V(\zeta)\right]\phi_n(\zeta) = \lambda_n\phi_n(\zeta), \tag{A12}$$

$$\psi_n(x,t) = \rho^{-\frac{1}{2}}e^{\frac{im}{\hbar}[\dot{\rho}x^2/2\rho + (\dot{\alpha}\rho - \alpha\dot{\rho})x/\rho]}\phi_n\left(\frac{x-\alpha}{\rho}\right). \tag{A13}$$

Finally, the Lewis-Riesenfeld phases are obtained as [30,58]

$$\gamma_n(t) = -\frac{1}{\hbar}\int_0^t dt'\left(\frac{\lambda_n}{\rho^2} + \frac{m(\dot{\alpha}\rho - \alpha\dot{\rho})^2}{2\rho^2}\right), \tag{A14}$$

and the boundary conditions from Eq. (A6), which are necessary for eigenstate transfer, become

$$\dot{\alpha}(t_\mathrm{i}) = \ddot{\alpha}(t_\mathrm{i}) = \dot{\alpha}(t_\mathrm{f}) = \ddot{\alpha}(t_\mathrm{f}) = 0, \tag{A15}$$

$$\dot{\rho}(t_\mathrm{i}) = \ddot{\rho}(t_\mathrm{i}) = \dot{\rho}(t_\mathrm{f}) = \ddot{\rho}(t_\mathrm{f}) = 0, \tag{A16}$$

justifying Eqs. (15) and (16) in the main text. By fixing $F(t) = m\omega^2(t)x_0(t)$ and $V(x) = 0$ in Eq. (A7), we obtain the time-dependent harmonic-oscillator Hamiltonian [15] from Eq. (10), while Eqs. (A9)–(A10) become Eqs. (13)–(14), respectively, of the main text.

APPENDIX B: TIME-EVOLUTION METHOD AND DISCRETIZATION-ERROR ANALYSIS

Numerical simulations inherently contain errors, particularly if an infinite space is approximated by a finite one, as in the discretization of quantum-optical systems. For this reason, we examine the stability of the time evolution and investigate the error due to the space and time discretization. This is a crucial step in validating our numerical implementation. To evaluate the numerical errors present in our implementation, we study the harmonic oscillator under static and time-dependent conditions (using the same frequency ω_st and mass m as in the main text).

1. Space-discretization error analysis

The well-known solutions of the harmonic-oscillator problem allow us to study the error introduced by the space discretization, constrained only by the machine precision. We first compare the wave-function amplitude computed via exact diagonalization and the corresponding exact solution involving the Hermite polynomials for different space-discretization steps Δx. The analysis offers an understanding of the error range in the considered interval for space discretization. Calculating the root-mean-square error, $\mathcal{E}$, between the two different computation methods and fitting the resulting data, we relate $\mathcal{E}$ to the discretization step Δx as $\mathcal{E} = a\Delta x^b/(1+c\Delta x)$, with $a = 0.46$, $b = 2.31$, and $c = 7.03$, allowing for a selection of Δx once the error threshold is fixed.

Taking into account the required computational time and discretization precision, we found that $\Delta x \sim 0.02\ \mu\mathrm{m} \approx 0.2l_\mathrm{st} \approx 0.4l_\mathrm{mt}$ (where l_st and l_mt are respectively the characteristic lengths of the static and moving trap during transport, i.e., with maximum tweezer depth) is sufficient to capture the wave-function details while maintaining a discretization-error threshold of approximately 10^{-5} and a computational time in the range of milliseconds.

2. Time-discretization error analysis

The time evolution is performed using the split-step Fourier method [76–80], as explained in detail in Ref. [55]. This method is particularly effective thanks to the Baker-Campbell-Hausdorff relation [81]

$$e^{iH\Delta t} \simeq e^{-iV\Delta t/2}e^{-iT\Delta t}e^{-iV\Delta t/2} + \mathcal{O}(\Delta t^3), \tag{B1}$$

and the fast Fourier transform (FFT) algorithm [82]. In Eq. (B1), T and V denote the kinetic and potential terms of the Hamiltonian. We have chosen to reduce the error by introducing a symmetric splitting, also known as Strang splitting, where the potential is applied for a half time step before and after the kinetic operator, obtaining an error of $\mathcal{O}(\Delta t^3)$ [76,80]. The Fourier transform is used in the intermediate steps to diagonalize the operators T and V in their respective bases, reducing each multiplication in Eq. (B1) to a scaling of $\mathcal{O}(N)$. This combination allows us to develop a straightforward algorithm to compute the evolution of the state $|\psi(t)\rangle$ from time t to time $t+\Delta t$ that can be summarized in the following steps:

$$\begin{aligned} e^{-iH\Delta t}\,|\psi(t,x)\rangle &\approx e^{-iV\Delta t/2}e^{-iT\Delta t}e^{-iV\Delta t/2}\,|\psi(t,x)\rangle \\ &\propto U_V(x)\mathcal{F}^{-1}\mathcal{F}U_T(x)\mathcal{F}^{-1}\mathcal{F}U_V(x)\,|\psi(t,x)\rangle \\ &\propto U_V(x)\mathcal{F}^{-1}U_T(p)\,|\psi(t,p)\rangle \\ &\propto U_V(x)\,\left|\psi''(t,x)\right\rangle \\ &\propto |\psi(t+\Delta t,x)\rangle\,. \end{aligned} \tag{B2}$$

We have used $U_V = e^{-iV\Delta t/2}$, $U_T = e^{-iT\Delta t}$, and $\mathcal{F}$ denotes the Fourier transform. Moreover, the method assumes that the wave function does not change significantly within each small time step Δt.

To study our system described by a Hamiltonian

$$H(t) = \frac{p^2}{2m} + V(x) + V(x,t), \tag{B3}$$

we need to recursively apply the previous algorithm to the discretized potential $V(x,t_n) = V_n$, where $n = 0, 1, \ldots, N_t$, with N_t being the total number of time steps, which in our computation for $T = 3$ ms was chosen to be 5000 (see below), leading to $\Delta t = 0.6\ \mu\text{s} \approx 0.02\ \tau_{\text{st}} \approx 0.06\ \tau_{\text{mt}}$.

The FFT, with a scaling of order $N\log(N)$, is the primary source of the high computational cost, and a large number of time steps can further slow down the computation. At first glance, the single-splitting method given by

$$e^{iH\Delta t} \simeq e^{iV\Delta t}e^{iT\Delta t} + \mathcal{O}(\Delta t^2) \tag{B4}$$

could reduce the computational time of the time evolution; however, it incurs the cost of introducing high-order oscillations for the intermediate steps. To quantify the error and the amplitude of the induced oscillations, we studied the time evolution of a static harmonic oscillator by analyzing the infidelity between the time-evolved state and the analytical one for each time step (we use a fixed grid of 5000 time steps). We expect a vanishing infidelity due to the static nature of the harmonic oscillator; however, oscillations are observed at all time steps except the initial and final ones. The infidelity for the single-splitting method oscillates around 1.7×10^{-4}, while the infidelity obtained with the Strang splitting oscillates around 2.2×10^{-7} with the same oscillation period of two times the characteristic time of the oscillator. Even though the Strang splitting method is 30% slower, it leads to an improvement of more than 2 orders of magnitude in the infidelity, which is crucial for us to obtain faithful results for infidelities under 10^{-4}.

The time discretization Δt was chosen by analyzing the splitting methods for different total numbers of time steps. After a careful quantification of the effects of the oscillations in the evolution method, we compute the infidelity between the numerical and analytical harmonic-oscillator solution as a function of Δt and chose a grid of 5000 points for the larger considered total time of 3 ms to reach infidelities below 10^{-7}. With this procedure, smaller times will keep the error bound. When varying either the total evolution time in the range $[0.01, 3]$ ms or the frequency of the oscillator in the interval $[\omega_{\text{st}}, 10\,\omega_{\text{st}}]$, the relation between the two methods (Strang splitting and single splitting) is consistent, even though the error increases for increasing total times and frequencies. Finally, we used the transport trajectory obtained for the harmonic oscillator within our STA approach and calculated the root-mean-square error between the numerically computed and analytical solutions. In the latter case, we observe that the error for the Strang splitting is one order of magnitude less than that obtained using the single-splitting approach. For the final simulations, the grid in space and time has been chosen to find a compromise between obtaining higher fidelities while spending a reasonable computational time. Given that the experimental measurable errors are around 10^{-2} to 10^{-3}, we choose these grids to reach an infidelity of the order of 10^{-7} for a moving harmonic oscillator.

APPENDIX C: TRANSPORT DYNAMICS

In this section, we present the time evolution of the quantities studied in Sec. IV for some particular values of the total duration T of the protocol. We have chosen $T = 0.5$, 1.5, 2.5, and 3 ms, which corresponds to $16.6\ \tau_{\text{st}}$, $49.8\ \tau_{\text{st}}$, $83\ \tau_{\text{st}}$, and $99.6\ \tau_{\text{st}}$, or $55.7\ \tau_{\text{mt}}$, $167.1\ \tau_{\text{mt}}$, $278.4\ \tau_{\text{mt}}$, and $334.1\ \tau_{\text{mt}}$. At this point, it is important to remember that τ_{st} and τ_{mt} are the characteristic times of the static and moving trap during transport (i.e., with maximum tweezer depth), which are related by $\tau_{\text{st}} \approx 2\,\tau_{\text{mt}}$.

In Fig. 6 we report the error after transport measured as the maximum of the infidelity with respect to the target state over the last waiting interval of the pulse [see Eqs. (8) and (9)] for the five families of pulses considered in the main text. The upper row shows the results without optimization, while the error during transport obtained for the optimized pulses is depicted in the lower row. The darker the color of the curve, the larger the total time T. The evolution of the infidelity depicts more or less the same behavior for all the pulses; the fidelity is very low until the atom reaches the target tweezer, and after that, the fidelity rapidly increases. During the waiting stage of the pulse, the value of the infidelity presents some oscillations that can be seen as the thick final part of each curve. To interpret these results, it is important to have in mind that the data are presented in logarithmic scale; therefore, even though the STA final evolution seems to be very noisy, the final variations are spread on a 10^{-6} scale. Furthermore, for the hybrid pulse with $\xi = 0.4$ and the minimum-jerk ramp, the final infidelity is higher for $T = 3$ ms than for $T = 2.5$ ms, which is consistent with the oscillations of the infidelity when plotted against T (see Fig. 3 in the main text). All the optimized pulses improve the fidelity by at least one order of magnitude.

Figure 7 shows the time evolution of the mean occupied level of the moving trap states $\langle N\rangle$ and its uncertainty ΔN given by the width of the distribution over the states of the moving tweezer during transport (upper row), as well as the motional effective temperature T_{eff} (lower row). Since the nonoptimized and optimized curves are very similar except for the linear pulse, the values for the initial (nonoptimized) ramps are shown as gray curves,

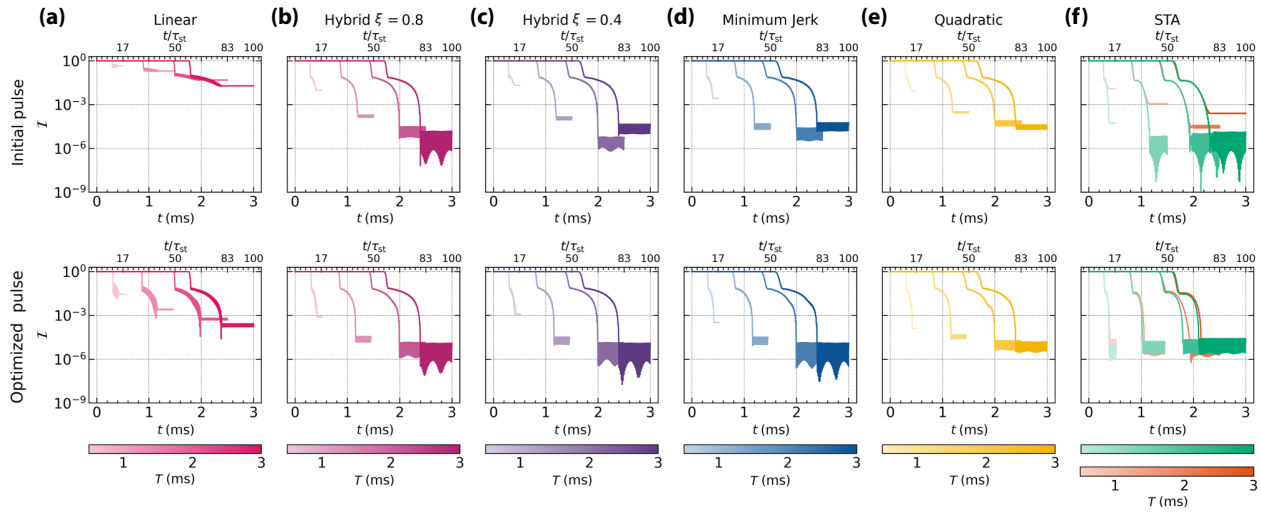

FIG. 6. Error during transport. Infidelity $\mathcal{I}$ vs time for total pulse times $T = 0.5$, 1.5, 2.5, and 3 ms (lighter to darker colors) for all the considered initial pulses (top row) and the corresponding optimizations (bottom row). The time scales are given in SI units and characteristic units, where the characteristic time τ_{st} of the static traps is related to the one of the moving trap via $\tau_{st} \approx 2\,\tau_{mt}$. All the pulses have a maximum amplitude of $A_{mt}^{max}/\hbar = 3.57 \times 2\pi$ MHz, the transport time is given by $\eta\,T$ with $\eta = 2/5$, and the total distance between static tweezers is $d = 7$ μm.

and the values obtained for the optimized pulses are in color. As before, darker colors indicate longer total times. Shorter pulse times are related to higher $\langle N\rangle$, ΔN, and T_{eff} values. Furthermore, the optimization induces heating mainly for the piecewise-linear ramp, the hybrid one with hybridicity $\xi = 0.8$ (closer to the linear pulse), and the STA approximation. We notice that the distribution over the moving-tweezer states is more sensitive as a heating measure for transient times than the effective temperature. The two peaks depicted by $\langle N\rangle$ and ΔN correspond to the mixing of states induced when the wave function abandons one of the tweezers and readjusts into the following one.

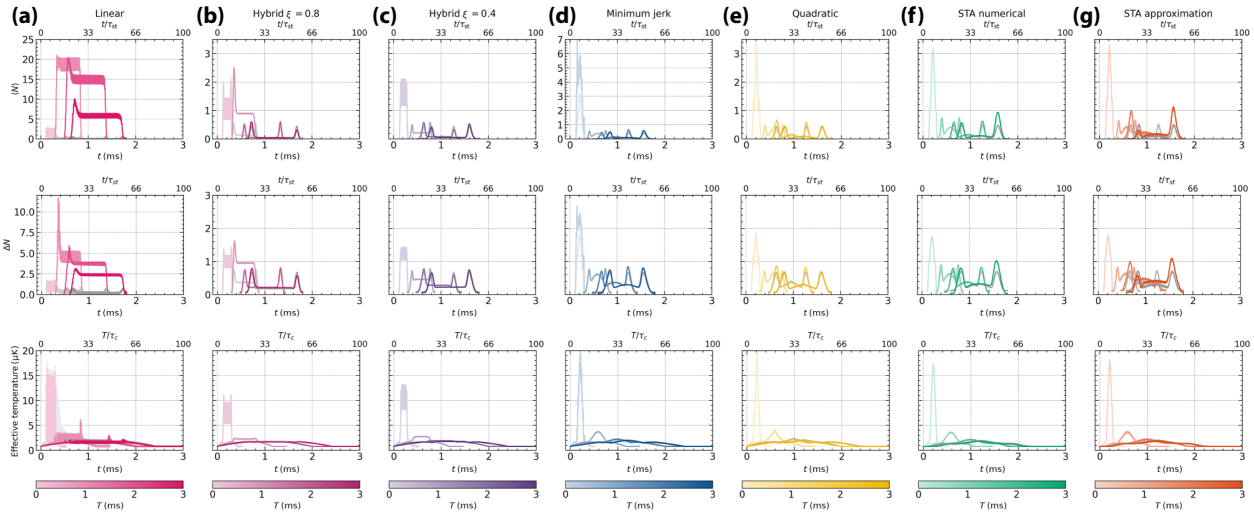

FIG. 7. Measures of heating during transport. Expected value $\langle N\rangle$ for the occupied level of the moving-trap states (top row) with the associated uncertainty ΔN (middle row), and effective temperature T_{eff} (bottom row) vs time for total pulse times $T = 0.5$, 1.5, 2.5, and 3 ms (lighter to darker colors) and for all the considered pulses. The colored curves correspond to the optimized pulses; the results obtained with the initial (nonoptimized) pulses are very similar (except for the piecewise-linear pulse) and are shown in gray behind each optimized curve. The time scales are given in SI units and characteristic units, with τ_{st} being the characteristic time of the static traps related to the one of the moving trap via $\tau_{st} \approx 2\,\tau_{mt}$. All the pulses have a maximum amplitude of $A_{mt}^{max}/\hbar = 3.57 \times 2\pi$ MHz, the transport time is given by $\eta\,T$ with $\eta = 2/5$, and the total distance between static tweezers is $d = 7$ μm.

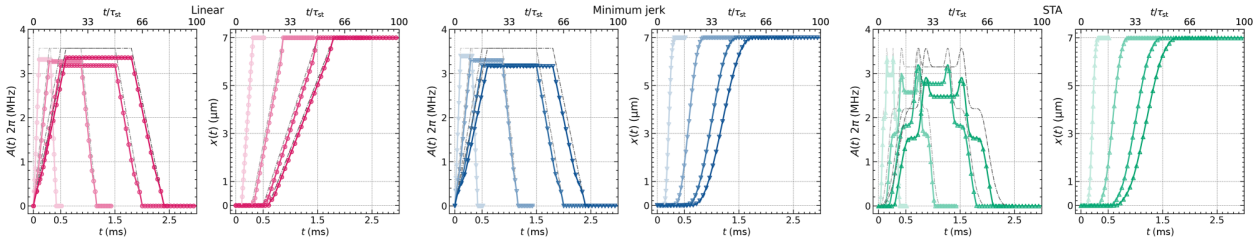

FIG. 8. Examples of optimized pulses. The amplitude $A_{mt}(t)$ and the position $x_{mt}(t)$ of the moving tweezer are shown for the linear (left), minimum-jerk (center), and STA (right) ramps. The optimized pulses are plotted in colored solid curves while the initial (nonoptimized) ones are depicted as gray dashed lines. We observe that the optimization results in small changes in the minimum-jerk and STA trajectories compared to the rest of the pulses. As before, the time scales are given in SI units and characteristic units, with τ_{st} being the characteristic time of the static traps related to the one of the moving tweezer via $\tau_{st} \approx 2\,\tau_{mt}$. All the pulses have a maximum amplitude of $A_{mt}^{max}/\hbar = 3.57 \times 2\pi$ MHz, the transport time is given by $\eta\, T$ with $\eta = 2/5$, and the total distance between static tweezers is $d = 7$ μm.

As expected due to the jumps in the velocity, in general, the piecewise-linear ramp depicts the highest $\langle N \rangle$ and ΔN values.

Finally, Fig. 8 presents examples of comparisons between the optimized pulses (colored solid lines) and the initial guesses or nonoptimized pulses (gray dashed lines) for the same total times, $T = 0.5$, 1.5, 2.5, and 3 ms, as considered before. We would like to highlight that the optimized position pulse for the linear ramp is not symmetric. This can also be seen in the asymmetric form depicted by $\langle N \rangle$ and ΔN in Fig. 7. While all the considered initial pulses were taken as symmetric, the automatic optimization algorithm breaks this symmetry, particularly for the piecewise-linear and approximated STA pulses. Furthermore, we observe that the optimization changes the piecewise-linear pulse mainly at the beginning and end of the transport stage; also, for the STA pulse, the optimization tends toward smaller amplitudes.

APPENDIX D: TRANSPORT AND TRANSFER ACROSS DIFFERENT DISTANCES AND VELOCITIES

The transport distance can range from a few micrometers to several hundreds of micrometers, depending on whether the transport occurs within the computational zone or between the storage and computational zones. We analyzed here the behavior of the minimum-jerk trajectory and the full numerical solution of our proposed STA pulse over distances between 1 and 14 μm. By way of example, in Fig. 9(a), we report the error after transport quantified by the infidelity $\mathcal{I}$ with the target state for a total distance of

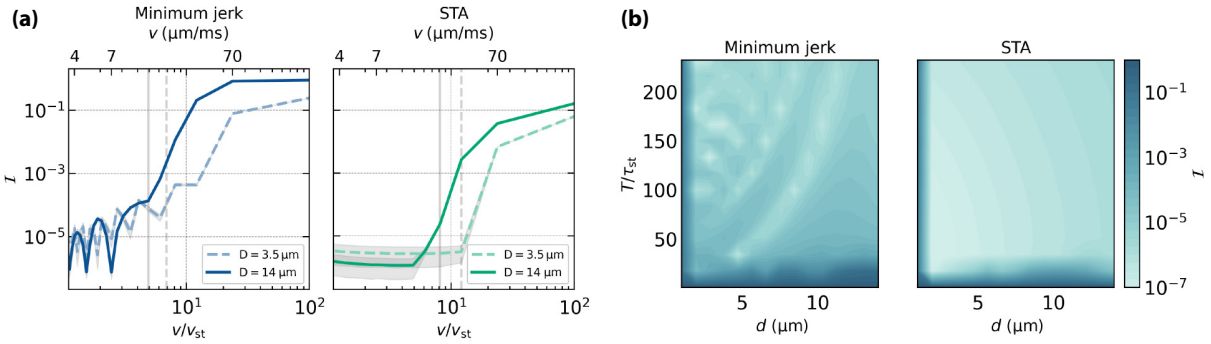

FIG. 9. Error during transport for larger distances. (a) Infidelity $\mathcal{I}$ vs velocity v in units of the characteristic velocity of the static oscillator v_{st} for distances of 3.5 and 14 μm (lighter-colored dashed line and darker-colored full line, respectively). The plots show the average infidelity over the last interval of the pulse (see Fig. 2) with the associated standard deviation (shaded area). The STA shows more stable behavior for both distances, followed by a speed up to reach the threshold infidelity of 10^{-4}; in particular, for a distance of 14 μm, that time reduces from 40 $\tau_{st} \approx 1.21$ ms ($v = 0.014$ m/s) for the minimum jerk to 24 $\tau_{st} \approx 0.75$ ms ($v = 0.024$ m/s) for the STA, while for a distance of 3.5 μm, it reduces from 32 $\tau_{st} \approx 0.97$ ms ($v = 0.018$ m/s) to 16 $\tau_{st} \approx 0.5$ ms ($v = 0.036$ m/s). (b) Infidelity $\mathcal{I}$ as function of the distance d and total time T. The time is varied between 0.01 and 7 ms, and the distance is ranges between 1 and 14 μm. For both pulses, the infidelity is higher at short times; however, while the minimum-jerk pulse shows a stronger dependence on T and d, the STA pulse exhibits much more stable behavior with lower infidelities.

3.5 μm and 14 μm (i.e., half and double the distance considered in the main text; transport over larger distances, although feasible, requires higher grid discretization and consequently longer computational times; for that reason, we stick to the typical range of distances found when moving atoms inside a determined zone and not among zones [17]), and a for total pulse time T ranging between 0.01 and 7 ms, corresponding to a mean velocity during transport varying between 1.75 and 0.0025 m/s. The figure shows the average infidelity over the last interval of the pulses and the associated standard deviation (shaded area) as a function of the mean velocity calculated as $d/(\eta T)$. Similar to what was observed for the original distance of 7 μm, the STA pulses depict a more stable behavior when compared to the minimum-jerk pulses. The infidelity increases as a function of the velocity. In both cases, the time needed for reaching the infidelity threshold of 10^{-4} increases with the distance. As before, the STA pulse allows for a reduction of approximately 40% in the time needed to obtain an infidelity of 10^{-4}.

In Fig. 9(b), we show the error $\mathcal{I}$ as function of both the distance d and the total time of the protocol T. The time ranges between 0.01 and 7 ms. while the distance is varied between 1 and 14 μm. For short distances and short total times, both protocols lead to errors above 10^{-1} due to the excitations induced by higher transport speeds. The bottom dark areas in the plots show the dependence of the quantum speed limit on the distance of the transport. Longer distances require a longer time; however, while the minimum jerk reaches infidelities of 10^{-6} or 10^{-7} in specific pairs of T and d, consistent with the oscillations present in Fig. 9(a), the STA shows a more stable and robust behavior on d and T. This suggest that the STA pulse not only reduces the error and the protocol time but it is also much more versatile than the considered experimentally inspired pulses.

[1] D. Bluvstein, H. Levine, G. Semeghini, T. T. Wang, S. Ebadi, M. Kalinowski, A. Keesling, N. Maskara, H. Pichler, M. Greiner, *et al.*, A quantum processor based on coherent transport of entangled atom arrays, Nature **604**, 451 (2022).

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

[3] A. J. Daley, I. Bloch, C. Kokail, S. Flannigan, N. Pearson, M. Troyer, and P. Zoller, Practical quantum advantage in quantum simulation, Nature **607**, 667 (2022).

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

[5] F. Gyger, M. Ammenwerth, R. Tao, H. Timme, S. Snigirev, I. Bloch, and J. Zeiher, Continuous operation of large-scale atom arrays in optical lattices, Phys. Rev. Res. **6**, 033104 (2024).

[6] D. Barredo, S. de 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).

[7] 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).

[8] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, *et al.*, Logical quantum processor based on reconfigurable atom arrays, Nature **626**, 58 (2024).

[9] A. M. Kaufman and K.-K. Ni, Quantum science with optical tweezer arrays of ultracold atoms and molecules, Nat. Phys. **17**, 1324 (2021).

[10] L. R. Liu, J. D. Hood, Y. Yu, J. T. Zhang, K. Wang, Y.-W. Lin, T. Rosenband, and K.-K. Ni, Molecular assembly of ground-state cooled single atoms, Phys. Rev. X **9**, 021039 (2019).

[11] L. R. B. Picard, G. E. Patenotte, A. J. Park, S. F. Gebretsadkan, and K.-K. Ni, Site-selective preparation and multistate readout of molecules in optical tweezers, PRX Quantum **5**, 020344 (2024).

[12] J. T. Zhang, L. R. B. Picard, W. B. Cairncross, K. Wang, Y. Yu, F. Fang, and K.-K. Ni, An optical tweezer array of ground-state polar molecules, Quantum Sci. Technol. **7**, 035006 (2022).

[13] J. Tindall, M. Fishman, E. M. Stoudenmire, and D. Sels, Efficient tensor network simulation of IBM's eagle kicked Ising experiment, PRX Quantum **5**, 010308 (2024).

[14] J. Beugnon, C. Tuchendler, H. Marion, A. Gaëtan, Y. Miroshnychenko, Y. R. P. Sortais, A. M. Lance, M. P. A. Jones, G. Messin, A. Browaeys, and P. Grangier, Two-dimensional transport and transfer of a single atomic qubit in optical tweezers, Nat. Phys. **3**, 696 (2007).

[15] E. Torrontegui, S. Ibáñez, X. Chen, A. Ruschhaupt, D. Guéry-Odelin, and J. G. Muga, Fast atomic transport without vibrational heating, Phys. Rev. A **83**, 013415 (2011).

[16] Y. Chew, T. Tomita, T. P. Mahesh, S. Sugawa, S. de Léséleuc, and K. Ohmori, Ultrafast energy exchange between two single Rydberg atoms on a nanosecond timescale, Nat. Photon. **16**, 724 (2022).

[17] K.-N. Schymik, V. Lienhard, D. Barredo, P. Scholl, H. Williams, A. Browaeys, and T. Lahaye, Enhanced atom-by-atom assembly of arbitrary tweezer arrays, Phys. Rev. A **102**, 063107 (2020).

[18] M. R. Lam, N. Peter, T. Groh, W. Alt, C. Robens, D. Meschede, A. Negretti, S. Montangero, T. Calarco, and A. Alberti, Demonstration of quantum brachistochrones between distant states of an atom, Phys. Rev. X **11**, 011035 (2021).

[19] S. Spence, R. V. Brooks, D. K. Ruttley, A. Guttridge, and S. L. Cornish, Preparation of ^{87}Rb and ^{133}Cs in the motional ground state of a single optical tweezer, New J. Phys. **24**, 103022 (2022).

[20] A. Pagano, D. Jaschke, W. Weiss, and S. Montangero, Optimal control transport of neutral atoms in optical tweezers at finite temperature, Phys. Rev. Res. **6**, 033282 (2024).

[21] D. M. Reich, M. Ndong, and C. P. Koch, Monotonically convergent optimization in quantum control using Krotov's method, J. Chem. Phys. **136**, 104103 (2012).
[22] T. Caneva, M. Murphy, T. Calarco, R. Fazio, S. Montangero, V. Giovannetti, and G. E. Santoro, Optimal control at the quantum speed limit, Phys. Rev. Lett. **103**, 240501 (2009).
[23] L. Van Damme, Z. Zhang, A. Devra, S. J. Glaser, and A. Alberti, Motion-insensitive time-optimal control of optical qubits, Quantum Sci. Technol. **10**, 035025 (2025).
[24] M. Dalgaard, F. Motzoi, J. H. M. Jensen, and J. Sherson, Hessian-based optimization of constrained quantum control, Phys. Rev. A **102**, 042612 (2020).
[25] Q. Zhang, X. Chen, and D. Guéry-Odelin, Fast and optimal transport of atoms with nonharmonic traps, Phys. Rev. A **92**, 043410 (2015).
[26] Even though there is no dissipative mechanism considered in our model, we adhere to the identification of vibrational excitations and heating. See, for instance, Refs. [63,64,83].
[27] D. Guéry-Odelin, A. Ruschhaupt, A. Kiely, E. Torrontegui, S. Martínez-Garaot, and J. G. Muga, Shortcuts to adiabaticity: Concepts, methods, and applications, Rev. Mod. Phys. **91**, 045001 (2019).
[28] L. Theis, F. Motzoi, S. Machnes, and F. Wilhelm, Counteracting systems of diabaticities using drag controls: The status after 10 years, Europhys. Lett. **123**, 60001 (2018).
[29] S. Deffner, C. Jarzynski, and A. del Campo, Classical and quantum shortcuts to adiabaticity for scale-invariant driving, Phys. Rev. X **4**, 021013 (2014).
[30] H. R. Lewis and W. B. Riesenfeld, An exact quantum theory of the time-dependent harmonic oscillator and of a charged particle in a time-dependent electromagnetic field, J. Math. Phys. **10**, 1458 (1969).
[31] X. Chen, E. Torrontegui, D. Stefanatos, J.-S. Li, and J. G. Muga, Optimal trajectories for efficient atomic transport without final excitation, Phys. Rev. A **84**, 043415 (2011).
[32] M. Rossignolo, T. Reisser, A. Marshall, P. Rembold, A. Pagano, P. J. Vetter, R. S. Said, M. M. Müller, F. Motzoi, T. Calarco, *et al.*, QuOCS: The quantum optimal control suite, Comput. Phys. Commun. **291**, 108782 (2023).
[33] M. M. Müller, R. S. Said, F. Jelezko, T. Calarco, and S. Montangero, One decade of quantum optimal control in the chopped random basis, Rep. Prog. Phys. **85**, 076001 (2022).
[34] T. Caneva, T. Calarco, and S. Montangero, Chopped random-basis quantum optimization, Phys. Rev. A **84**, 022326 (2011).
[35] N. Rach, M. M. Müller, T. Calarco, and S. Montangero, Dressing the chopped-random-basis optimization: A bandwidth-limited access to the trap-free landscape, Phys. Rev. A **92**, 062343 (2015).
[36] A. Omran, H. Levine, A. Keesling, G. Semeghini, T. T. Wang, S. Ebadi, H. Bernien, A. S. Zibrov, H. Pichler, S. Choi, *et al.*, Generation and manipulation of Schrödinger cat states in Rydberg atom arrays, Science **365**, 570 (2019).
[37] S. Deffner and S. Campbell, Quantum speed limits: From Heisenberg's uncertainty principle to optimal quantum control, J. Phys. A **50**, 453001 (2017).
[38] I. Bloch, Ultracold quantum gases in optical lattices, Nat. Phys. **1**, 23 (2005).
[39] B. Yang, H. Sun, C.-J. Huang, H.-Y. Wang, Y. Deng, H.-N. Dai, Z.-S. Yuan, and J.-W. Pan, Cooling and entangling ultracold atoms in optical lattices, Science **369**, 550 (2020).
[40] C. Gross and I. Bloch, Quantum simulations with ultracold atoms in optical lattices, Science **357**, 995 (2017).
[41] M. Saffman, T. G. Walker, and K. Mølmer, Quantum information with Rydberg atoms, Rev. Mod. Phys. **82**, 2313 (2010).
[42] T. Chalopin, P. Bojović, D. Bourgund, S. Wang, T. Franz, I. Bloch, and T. Hilker, Optical superlattice for engineering Hubbard couplings in quantum simulation, Phys. Rev. Lett. **134**, 053402 (2025).
[43] L.-M. Steinert, P. Osterholz, R. Eberhard, L. Festa, N. Lorenz, Z. Chen, A. Trautmann, and C. Gross, Spatially tunable spin interactions in neutral atom arrays, Phys. Rev. Lett. **130**, 243001 (2023).
[44] P. Osterholz, Freely configurable holographic trap arrays for the trapping of single atoms, Master thesis, Ludwig Maximilians Universität München, 2020.
[45] Parameters suggested by Christian Groß group.
[46] G. Pesce, P. H. Jones, O. M. Maragò, and G. Volpe, Optical tweezers: Theory and practice, Eur. Phys. J. Plus **135**, 949 (2020).
[47] D. S. Simon, *Gaussian Beams and Lasers* (Morgan and Claypool Publishers, San Rafael, CA, USA, 2016), Chap. 3, pp. 3–1 to 3–14. 2053–2571.
[48] A. Matthies, Optical Conveyor-Belt Transport of Cs and Rb Atoms, Durham theses, Durham University, 2023.
[49] N. Hogan, An organizing principle for a class of voluntary movements, J. Neurosci. **4**, 2745 (1984).
[50] R. Jáuregui, Nonperturbative and perturbative treatments of parametric heating in atom traps, Phys. Rev. A **64**, 053408 (2001).
[51] D. Guéry-Odelin, A. Ruschhaupt, A. Kiely, E. Torrontegui, S. Martínez-Garaot, and J. G. Muga, Shortcuts to adiabaticity: Concepts, methods, and applications, Rev. Mod. Phys. **91**, 045001 (2019).
[52] X. Chen, A. Ruschhaupt, S. Schmidt, A. del Campo, D. Guéry-Odelin, and J. G. Muga, Fast optimal frictionless atom cooling in harmonic traps: Shortcut to adiabaticity, Phys. Rev. Lett. **104**, 063002 (2010).
[53] G. Ness, C. Shkedrov, Y. Florshaim, and Y. Sagi, Realistic shortcuts to adiabaticity in optical transfer, New J. Phys. **20**, 095002 (2018).
[54] Q. Zhang, J. G. Muga, D. Guéry-Odelin, and X. Chen, Optimal shortcuts for atomic transport in anharmonic traps, J. Phys. B **49**, 125503 (2016).
[55] S. H. Hauck and V. M. Stojanović, Coherent atom transport via enhanced shortcuts to adiabaticity: Double-well optical lattice, Phys. Rev. Appl. **18**, 014016 (2022).
[56] D. Guéry-Odelin and J. G. Muga, Transport in a harmonic trap: Shortcuts to adiabaticity and robust protocols, Phys. Rev. A **90**, 063425 (2014).
[57] S. H. Hauck, G. Alber, and V. M. Stojanović, Single-atom transport in optical conveyor belts: Enhanced shortcuts-to-adiabaticity approach, Phys. Rev. A **104**, 053110 (2021).
[58] A. K. Dhara and S. V. Lawande, Feynman propagator for time-dependent Lagrangians possessing an invariant quadratic in momentum, J. Phys. A **17**, 2423 (1984).

[59] S. K. Singh and S. Mandal, The solutions of the generalized classical and quantum harmonic oscillators with time dependent mass, frequency, two-photon parameter and external force: The squeezing effects, Opt. Commun. **283**, 4685 (2010).

[60] A. D. Campo and M. G. Boshier, Shortcuts to adiabaticity in a time-dependent box, Sci. Rep. **2**, 648 (2012).

[61] S. Hwang, H. Hwang, K. Kim, A. Byun, K. Kim, S. Jeong, M. P. Soegianto, and J. Ahn, Fast and reliable atom transport by optical tweezers, Optica Quantum **3**, 64 (2025).

[62] By fulfilling the same initial and final boundary conditions as the minimum-jerk trajectory, our STA solution also prevents heating arising from the acceleration at the beginning and end of the movement; however, since any translation with a nonconstant velocity profile implies a nonconstant sweeping rate in the AOD, which was associated with resonant intensity modulations (see Refs. [10,12]), we anticipate that our proposed STA position pulse could also be susceptible to this kind of heating. This point requires further experimental tests, and we foresee two possibilities, one is to solve it by implementing a closed-loop control optimization, and the other is to create a new hybrid pulse following the strategy of Kang-Kuen Ni's group as presented in Ref. [10] (i.e., avoid the resonant intensity modulations by including a constant sweep AOD rate for intermediate times) by mixing the STA and linear pulses.

[63] T. A. Savard, K. M. O'Hara, and J. E. Thomas, Laser-noise-induced heating in far-off resonance optical traps, Phys. Rev. A **56**, R1095 (1997).

[64] C. W. Gardiner, J. Ye, H. C. Nagerl, and H. J. Kimble, Evaluation of heating effects on atoms trapped in an optical trap, Phys. Rev. A **61**, 045801 (2000).

[65] G. T. Hickman and M. Saffman, Speed, retention loss, and motional heating of atoms in an optical conveyor belt, Phys. Rev. A **101**, 063411 (2020).

[66] C. Cohen-Tannoudji, B. Diu, and F. Laloë, *Quantum Mechanics* (Wiley, New York, 1977), Vol. 1.

[67] The expected value and maximum width of the distribution over the moving-tweezer states are computed on the transport interval as follows. For each time t, we compute the states $|\phi_n(t)\rangle$ of the moving trap via diagonalization of the moving-tweezer Hamiltonian. Then, we project the evolved wave function $|\psi(t)\rangle$ over each of these states to calculate the weights or populations over each state $p_n = |\langle\phi_n(t)|\psi(t)\rangle|^2$. Once we have the population distribution $\{p_n\}$, we calculate the variance and mean value as usual. We perform the same calculation using a harmonic approximation of the moving trap, and we check the consistency of the results. The maximum effective temperature $\max(T_{\text{eff}})$ is calculated over the complete protocol duration.

[68] P. Carruthers and M. M. Nieto, Coherent states and the forced quantum oscillator, Am. J. Phys. **33**, 537 (1965).

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

[70] V. Bergholm, W. Wieczorek, T. Schulte-Herbrüggen, and M. Keyl, Optimal control of hybrid optomechanical systems for generating non-classical states of mechanical motion, Quantum Sci. Technol. **4**, 034001 (2019).

[71] 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).

[72] J. Singh, R. Zeier, T. Calarco, and F. Motzoi, Compensating for nonlinear distortions in controlled quantum systems, Phys. Rev. Appl. **19**, 064067 (2023).

[73] The perturbative approach proposed in Ref. [25] for anharmonic traps can be applied to a fourth-order Taylor expansion of our tweezer potential to obtain an analytical estimation of the magic times; however, we would like to highlight that these theoretical results require further experimental tests.

[74] B. Khani, S. T. Merkel, F. Motzoi, J. M. Gambetta, and F. K. Wilhelm, High-fidelity quantum gates in the presence of dispersion, Phys. Rev. A **85**, 022306 (2012).

[75] C. Cicali, M. Calzavara, E. Cuestas, T. Calarco, R. Zeier, and F. Motzoi, Dataset for "Fast neutral atom transport and transfer between optical tweezers", available at: https://doi.org/10.5281/zenodo.16572794 (2025).

[76] S. Montangero, E. Montangero, and Evenson, *Introduction to Tensor Network Methods* (Springer, Cham, Switzerland, 2018).

[77] C. Lubich, *From Quantum to Classical Molecular Dynamics: Reduced Models and Numerical Analysis* (European Mathematical Society, Zürich, 2008), Vol. 12.

[78] W. Bao, S. Jin, and P. A. Markowich, On time-splitting spectral approximations for the Schrödinger equation in the semiclassical regime, J. Comput. Phys. **175**, 487 (2002).

[79] R. L. Speth, W. H. Green, S. MacNamara, and G. Strang, Balanced splitting and rebalanced splitting, SIAM J. Numer. Anal. **51**, 3084 (2013).

[80] S. MacNamara and G. Strang, in *Splitting Methods in Communication, Imaging, Science, and Engineering*, edited by R. Glowinski, W. Yin, and Stanley J. Osher (Springer, Cham, Switzerland, 2016), pp. 95–114.

[81] C. Cohen-Tannoudji, B. Diu, and F. Laloë, *Quantum Mechanics, Volume 3: Fermions, Bosons, Photons, Correlations, and Entanglement* (John Wiley & Sons, Weinheim, Germany, 2019).

[82] M. Frigo and S. G. Johnson, The design and implementation of FFTW3, Proc. IEEE **93**, 216 (2005).

[83] M. E. Gehm, K. M. O'Hara, T. A. Savard, and J. E. Thomas, Dynamics of noise-induced heating in atom traps, Phys. Rev. A **58**, 3914 (1998).

Chapter 6

Classical feature map surrogates and metrics for quantum control landscapes

---

Authors:	**Martino Calzavara**, Tommaso Calarco and Felix Motzoi
Year:	2025
Journal:	ArXiv preprint
Volume/Issue:	[quant-ph] 2509.25930
DOI:	10.48550/arXiv.2509.25930

6.1 Summary and context

In the following preprint [98], we investigate theoretically the properties of quantum control landscapes, i.e. the family of functionals whose minimization is central to quantum optimal control (see Sec. 3.2). We consider aspects relating to both the representation of such landscapes [99, 100], by designing regression models (see Sec. 3.1) and studying their performance [101], and to the problem of finding their minima, by inspecting a collection of metrics, such as ruggedness, variance and trap density, which characterize the hardness of this problem [102, 101].

The regression models we consider are linear combinations of basis functions — the so-called feature maps, using ML terminology. This simple structure allows us to relate basic properties of quantum time evolution, like transition frequencies and total evolution time, to the ones of the landscape, establishing bounds on the derivatives and bandwidth of this family of functions. Moreover, training is much easier and reliable for this class of models compared to custom models

with multiple layers, such as the one we consider in Chapter 5, substituting the loss minimization problem with a simpler matrix inversion. We test our methods numerically on a state transfer problem involving a one-dimensional Transverse Field Ising Model, and find that bandwidth-limited kernels are a particularly effective feature map choice.

Concerning optimization and metrics, we show how the bound on the derivatives ensures the Lipschitz continuity of the landscape, which connects the volume of a region in control space with the maximum improvement in the figure of merit that we can obtain by exploring it. We discuss how this effect, together with bandwidth limitedness, typically suppresses the local trap density and variance, giving rise to a rather flat landscape as time discretization of the control pulse is increased, and how these insights can inform optimizer design.

From a system-theoretical point of view, we are here addressing problems which are formally similar to the ones in Chapter 4, while switching our focus from the transfer function to the quantum cost landscape. These are, for instance, how to use sampled data to create a digital twin of the system, and how we can use physical knowledge to inform the design of optimizers and regression models.

6.2 Author contribution

The Author is responsible for all the mathematical proofs and theoretical results in this paper. He developed the software related to linear regression and Lie-Fourier coefficient evaluation using the Discrete Fourier Transform (Sec. III.D), and generated all data used throughout the manuscript. He is the main contributor to all Sections except Sec. I, VIII, IX (where he also contributed) and created all Figures.

Classical feature map surrogates and metrics for quantum control landscapes

Martino Calzavara,[1,2,*] Tommaso Calarco,[1,2,3] and Felix Motzoi[1,2,†]

[1] *Peter Grünberg Institute - Quantum Control (PGI-8), Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Straße, 52428 Jülich, Germany*
[2] *Institute for Theoretical Physics, University of Cologne, Zülpicher Straße 77, 50937 Cologne, Germany*
[3] *Dipartimento di Fisica e Astronomia, Università di Bologna, 40127 Bologna, Italy*

(Dated: October 1, 2025)

We derive and analyze three feature map representations of parametrized quantum dynamics, which generalize variational quantum circuits. These are (i) a Lie-Fourier partial sum, (ii) a Taylor expansion, and (iii) a finite-dimensional sinc kernel regression representation. The Lie-Fourier representation is shown to have a dense spectrum with discrete peaks, that reflects control Hamiltonian properties, but that is also compressible in typically found symmetric systems. We prove boundedness in the spectrum and the cost function derivatives, and discrete symmetries of the coefficients, with implications for learning and simulation. We further show the landscape is Lipschitz continuous, linking global variation bounds to local Taylor approximation error—key for step size selection, convergence estimates, and stopping criteria in optimization. This also provides a new form of polynomial barren plateaux originating from the Lie-Fourier structure of the quantum dynamics. These results may find application in local and general surrogate model learning, which we benchmark numerically, in characterizations of hardness and phase transitions in the problem instances, and for meta-parameter heuristics in quantum optimizers.

I. INTRODUCTION

Many functional processes in the natural and engineering sciences suffer from high complexity or limited observability. As a result, they are often treated and interrogated as *black box oracles* – systems with unknown structure or landscape. This challenge concerns a wide range of disparate fields, including condensed matter, control theory, complexity theory, machine learning, and quantum optimization. The respective approaches to modeling the landscapes reveal different but complementary methods with large overlaps and synergies.

A central question across these domains is how to characterize the prescriptive performance of the system, that is, how well it performs under a given configuration, and, by extension, how accurately a (partial) model can predict or optimize that performance. In condensed matter physics, average behaviour is analysed to uncover landscape features such as phase transitions and the prevalence of metastable states (or traps). In contrast, complexity theory focuses on worst case performance, often emphasizing the difficulty of sampling in landscapes plagued by barren plateaus that hinder the search for global optima. Machine learning aims to approximate and reproduce the functional behavior of the black-box system, with a focus on model expressivity. Conversely, control theory seeks to directly optimize system outputs, raising questions about controllability in relation to high-dimensional input spaces.

In this work, we examine this multifaceted problem through the lens of quantum optimization and optimal control theory. Many classical concepts, such as phase transitions, barren plateaus, and landscape complexity, have quantum analogues, and corresponding methodologies have been adapted to the quantum setting. For instance, phase transitions are known to impact optimization hardness, and have been studied in quantum optimal control (QOC) [1–4]. Similarly, insights from classical machine learning have shaped our understanding of variational quantum algorithms (VQAs) and quantum machine learning (QML), often highlighting negative results due to barren plateaus [5–7]. The presence of traps and other obstructions has also been investigated in the QOC literature [8–11], alongside various metrics to quantify the hardness of the optimization landscape [3, 12–15]. Efforts to learn and predict the quantum cost landscape have employed techniques ranging from Fourier analysis [16–21] to Gaussian process [3, 22, 23] and neural network [3] Ansätze.

Rather than relying on physically inspired Ansätze, our approach derives first-principles constraints that characterize landscape structure and its surrogates under general assumptions. We analyze the quantum functional mapping controls to figure of merit (e.g. fidelity) using three representations: (i) a Fourier basis expansion, (ii) a Taylor expansion with respect to the controls, and (iii) a low-dimensional bandwidth-limited kernel. Notably, we show that VQA and QML landscapes can be viewed as special cases of the general QOC landscape, where controls act sequentially and time-ordering becomes trivial.

We derive upper bounds on gradient and higher-order derivatives, bounds relating control variations to cost changes, and bounds on the variance of the cost landscape, which plays a critical role in the emergence of barren plateaus. These derivative bounds enable a quantitative analysis of the Taylor expansion error, establishing its local efficiency and offering a global Lipschitz bound that can guide optimizer design, including stopping con-

* m.calzavara@fz-juelich.de
† f.motzoi@fz-juelich.de

ditions and global sampling strategies. Conversely, our variance analysis indicates that, even under ideal conditions, the fraction of the landscape with meaningful variation shrinks rapidly with increasing problem dimensionality.

Finally, we compare the representational power of the three expansions. We find that the kernel representation is the most expressive in the limit of dense sampling, while the Fourier and Taylor representations are more effective in low-data regimes. Our results provide quantitative dependencies of landscape complexity on key parameters such as control operator spectral bandwidth, total evolution time (or circuit depth), and system size – offering practical insights for algorithm design and parameter tuning.

The manuscript is organized as follows: in Sec. II we define the notion of quantum dynamical landscape and introduce the three representations that we will use to study the problem. These are later investigated in depth within dedicated sections: Sec. III for the Fourier-, Sec. V for the Taylor- and Sec. VI for the sinc representation. In Sec. IV we prove *en passant* some properties of the landscape derivatives (using the results from Sec. III). We then highlight the relevance of all previous results in the context of landscape optimization and of optimizer design respectively in Sec. VII and Sec. VIII. Finally, we draw our conclusions in Sec. IX. The Appendices are primarily devoted to the mathematical proofs of the main results (especially App. C and D), and to related details like notation (App. A), problem standardization (App. B) and numerical examples (App. E).

II. PROBLEM SETTING AND OVERVIEW

The starting point of our discussion is a finite D-dimensional evolving quantum system, with states $|\psi\rangle \in \mathcal{H} \simeq \mathbb{C}^D$. To consider the general case for parametric unitary operations and extend the simple structure of layered gates, we model the circuit evolution through its generator given by a Hamiltonian $\hat{H}(t)$ that depends on time through a single bounded external control $u(t) \in [-u_{\max}, u_{\max}]$. To perform time ordering, we discretize the control to be piecewise-constant on an N-timestep grid (with uniform step $\delta t = T/N$) from $t = 0$ to $t = T$. Note that while δt can be made arbitrarily small to retain precision, in quantum circuits one usually considers gates generated on large single timesteps, while in control theory the sampling rate of the control sets a practical upper bound on δt [24].

We consider the standard bilinear form of the controls given by

$$\hat{H}[u(t)] = \hat{H}_d + \hat{H}_c u(t)$$

which we write in vector form (see App. A for more details about notation) as

$$\hat{H}(u_\nu) = \hat{H}_d + \hat{H}_c u_\nu,$$

where $\boldsymbol{u} \in \mathcal{C}^N := [-u_{\max}, u_{\max}]^N$, defining a region of interest as a hypercube. In order to obtain a minimally cumbersome treatment, all results in the main text are discussed within this single control setting. Nevertheless, most of them are proved for multiple controls in App. C, yielding the results for a single control as a special case.

Starting from the initial state $|\psi(0)\rangle$, the system will evolve to

$$|\psi(T)\rangle = \hat{U}(\boldsymbol{u}) |\psi(0)\rangle$$

at time T, where the time-ordering is given straightforwardly by

$$\hat{U}(\boldsymbol{u}) = \hat{U}(u_N) \cdots \hat{U}(u_1) = e^{-i\delta t \hat{H}(u_N)} \cdots e^{-i\delta t \hat{H}(u_1)}. \tag{1}$$

This unitary operator defines a generalized Parametrized Quantum Circuit (PQC), where the drift and control Hamiltonian $\hat{H}_d$ and $\hat{H}_c$ that generate the gates can in general act at the same time. This is in contrast with the typical VQA setting, where usually it is assumed that only one generator is acting within the gates defining each circuit layer.

We define a quantum dynamical landscape $J(\boldsymbol{u})$ as the expectation value of a Hermitian operator $\hat{O}$ over the output of the circuit $|\psi(T)\rangle$:

$$J(\boldsymbol{u}) = \langle\psi(T)| \hat{O} |\psi(T)\rangle = \langle\psi(0)| \hat{U}^\dagger(\boldsymbol{u}) \hat{O} \hat{U}(\boldsymbol{u}) |\psi(0)\rangle . \tag{2}$$

In the VQA setting, $\hat{O}$ is typically expressed as a sum of observables that can be readily measured in a typical experimental setup (e.g. Pauli strings) [22]. The case of optimal control for a state transfer problem corresponds instead to choosing $|\psi\rangle = |\psi(0)\rangle$ as initial state and $\hat{O} = |\chi\rangle\langle\chi|$ as the density matrix of the (pure) target state. In this case, $J(\boldsymbol{u})$ corresponds to the state fidelity that we want to maximize over the controls $\boldsymbol{u}$. The average gate fidelity F_{avg} with respect to a gate $\hat{U}_{\text{target}}$ can also be written as a sum of such landscapes. In fact, given a sample of states $\{|\psi_i\rangle \in \mathcal{H}\}_{i=1}^M$, we have

$$\begin{aligned} F_{\text{avg}} &:= \frac{1}{M} \sum_{i=1}^{M} |\langle\psi_i| \hat{U}_{target}^\dagger \hat{U}(\boldsymbol{u}) |\psi_i\rangle|^2 \\ &= \frac{1}{M} \sum_{i=1}^{M} \langle\psi_i| \hat{U}^\dagger(\boldsymbol{u}) \hat{O}_i \hat{U}(\boldsymbol{u}) |\psi_i\rangle = \frac{1}{M} \sum_{i=1}^{M} J_i(\boldsymbol{u}), \end{aligned}$$

where we defined the observables

$$\hat{O}_i = \hat{U}_{\text{target}} |\psi_i\rangle\langle\psi_i| \hat{U}_{\text{target}}^\dagger.$$

In the main text we will focus on the case of a state fidelity landscape, but we prove most of our results for a generic observable $\hat{O}$ in App. C.

The main aim of this paper is to study how to represent a quantum dynamical landscape as a linear combination of non-linear functions of the form

$$J(\boldsymbol{u}) = \sum_{i=1}^{N_{\text{weights}}} w_i \phi_i(\boldsymbol{u}), \tag{3}$$

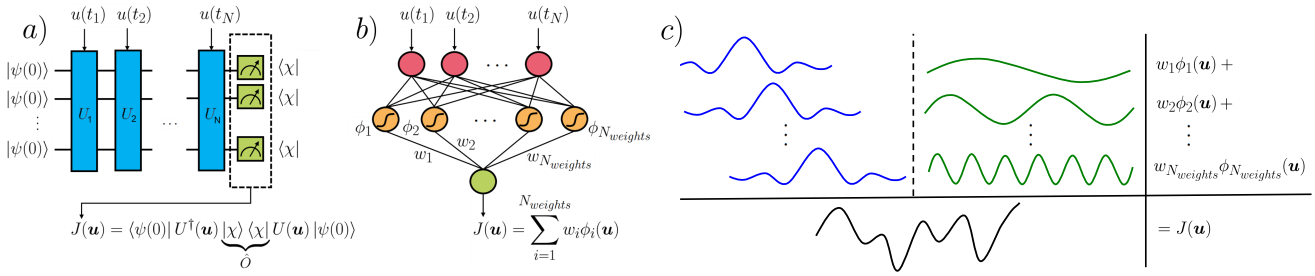

FIG. 1: The time-evolved expectation value $\langle \hat{O}(\boldsymbol{u})\rangle$ of an observable (e.g. state fidelity $\hat{O} = |\chi\rangle\langle\chi|$) for a system controlled through stepwise-constant pulses $\boldsymbol{u}$ is a Parametrized Quantum Circuit (Panel (a)). As such, it can be represented as a shallow computational network (Panel (b)), which consists in a linear combination of non-linear functions $\phi_i(\boldsymbol{u})$ called "features", with weights w_i (Panel (c)).

where i indexes the basis, specified depending on the choice of the functions $\phi_i(\boldsymbol{u})$.

In the machine learning context these are usually referred to as features, hence the name "feature map representation". See Fig. 1 for a graphical representation of this class of models. Compared to deep neural networks, for these shallow models it is easier to relate their structure both to the physical content of the system generating them and to the properties of their outputs. They can therefore be thought of as intermediate representations (or classical surrogates) of the quantum system, making them an attractive tool to analyze and even compute quantum cost landscapes [17, 18, 20, 25, 26].

A very natural and well studied choice for the feature map is the so-called Fourier representation, given by

$$\phi_{\boldsymbol{\omega}}(\boldsymbol{u}) = e^{-i\boldsymbol{\omega}\cdot\boldsymbol{u}}$$

where the weighting over $\boldsymbol{\omega}$ encodes the frequency spectrum of the landscape. We show in Sec. III that in general this spectrum is dense, unlike in the VQA and QML setting where only discrete combinations of the ω appear, corresponding to the eigenspectrum of the Hamiltonian generators. Naturally, this poses a unique challenge for classical surrogate models.

We also prove that the presence of symmetries in the dynamics constrains these frequencies, determining selection rules that we will describe in detail. We provide in Sec. III D a numerical example for the transverse field Ising model.

Another important representation is the polynomial feature map given by

$$\phi_{\boldsymbol{p}}(\boldsymbol{u}) = \prod_{\nu=1}^{N} (u_\nu - u_\nu^{(0)})^{p_\nu}$$

with $\boldsymbol{p} \in \mathbb{N}^N$, $\boldsymbol{u}^{(0)}$ a reference pulse, and where the weights $w_{\boldsymbol{p}}$ can be computed from the derivatives by Taylor expansion, as shown in Sec. V. We bound these derivatives using the Fourier representation, establishing the Lipschitz continuity of the landscape and thereby providing an estimate of the error in this expansion. This enables a low-order $\mathcal{O}(poly(N))$ truncation under certain conditions relating to constraining the total time and energy of the external fields.

The third representation is given by bandwidth-limited kernel function

$$\phi_i(\boldsymbol{u}) = \prod_{\nu=1}^{N} \frac{\sin\left[\omega_{\max}\delta t(u_\nu - u_\nu^{(i)})\right]}{\omega_{\max}\delta t(u_\nu - u_\nu^{(i)})},$$

which captures correlations between different sampled data points $\mathcal{D}_{\text{train}} = \{(\boldsymbol{u}^{(i)}, J(\boldsymbol{u}^{(i)}))\}_{i=1}^{N_{\text{train}}}$. This regression model is derived in Sec. VI by essentially integrating over the infinite Fourier spectrum, enhancing the tractability of the learning task. We further discuss efficiently training the model via RIDGE regression [17, 27].

Finally, we conclude by exploring the relevance of our findings in the context of landscape optimization. In order to quantify the difficulties inherent to this problem, several landscape measures have been proposed in the literature (ruggedness, trap density, gradient variance and barren plateaux, etc.) [3, 5, 6], each one highlighting a different way in which the optimization problem can be considered hard (or not). In Sec. VII, we relate a selection of these measures to the representations that we have studied, establishing bounds for them whenever possible. We also investigate the relevance of our findings in the tuning and design of optimization algorithms in Sec. VIII.

III. LIE-FOURIER REPRESENTATION

In order to derive a Fourier representation of the landscape J, we expand each one of the time-step unitaries in Eq. (1) by means of the Lie-Trotter product formula [28–30]

$$\begin{aligned}\hat{U}(u) &= e^{-i\delta t\hat{H}(u)} \\ &= \lim_{n\to\infty} (e^{-\frac{i\delta t}{n}\hat{H}_d} e^{-\frac{i\delta t}{n}u\hat{H}_c})^n = \lim_{n\to\infty} \hat{U}_n(u) \quad (4)\end{aligned}$$

which converges to its limit, with an error of order $\mathcal{O}(n^{-1})$ that depends on $\delta t A||\hat{H}||_\infty$. By studying this

expression for a generic n we can find a sequence of approximations J_n of the landscape J, whose convergence properties are discussed in detail in App. C. On the other hand, finite order expansions are also interesting on their own, and for example for $n = 1$ we recover an interleaved circuit which is a common Ansatz for VQA [22]. More generally, for any n and even for multiple controls (see App. C), the resulting circuits will be Periodic Structure Ansätze, as introduced in [7], which consist of stacked copies of a predefined gate sequence. In such a case, we can then apply a similar procedure to [25] and obtain an expression of Eq. (4) as a possibly infinite sum of complex exponentials. We will refer to the resulting partial Fourier sum as to the Lie-Fourier representation of the landscape. This stresses the dependence of the representation on the usage of the Lie product expansion as a preliminary step, which can be in principle substituted with higher order Suzuki-Trotter products [31] or other approximation schemes [32], giving rise to different frequencies and coefficients.

A. Commuting Hamiltonians $[\hat{H}_d, \hat{H}_c] = 0$

Let us first look at a base case of the problem. If $[\hat{H}_c, \hat{H}_d] = 0$ the Lie-Trotter expansion stops at the first order, and we have

$$\hat{U}(\boldsymbol{u}) = e^{-iN\delta t\hat{H}_d} e^{-i\delta t\hat{H}_c \sum_{\nu=1}^{N} u_\nu}.$$

The unitary operator only depends on the "effective" control $\bar{u} := N^{-1}\sum_{\nu=1}^{N} u_\nu$. We can write this unitary operator as a finite sum of Fourier components by working in the eigenbasis of $\hat{H}_c = \hat{V}^\dagger\hat{\Lambda}\hat{V}$, with $\langle i|\,\hat{\Lambda}\,|j\rangle = \delta_{ij}\lambda_i$:

$$\begin{aligned}\hat{U}(\bar{u}) &= e^{-iN\delta t\hat{H}_d}\hat{V}^\dagger e^{-iT\hat{\Lambda}\bar{u}}\hat{V} \\ &= e^{-iN\delta t\hat{H}_d}\sum_{j=1}^{D}\hat{V}^\dagger|j\rangle\, e^{-iT\lambda_j\bar{u}}\,\langle j|\,\hat{V} \\ &=: \sum_{\omega\in\mathcal{S}}\hat{B}^\omega(T)e^{-iT\omega\bar{u}},\end{aligned}$$

where we defined the spectrum of the control Hamiltonian $\mathcal{S} = \{\lambda_j\}_{j=1,\dots,D}$. We can now write also the transfer fidelity with respect to a target state $|\chi\rangle$ from Eq. (2) as a finite Fourier sum

$$\begin{aligned}J(\bar{u}) &= \sum_{\omega,\omega'\in\mathcal{S}}\underbrace{\langle\psi|\,\hat{B}^{\omega\dagger}(T)\,|\chi\rangle}_{b^*_\omega}\underbrace{\langle\chi|\,\hat{B}^{\omega'}(T)\,|\psi\rangle}_{b_{\omega'}}e^{iT(\omega-\omega')\bar{u}} \\ &=: \sum_{\omega\in\mathcal{S}^\Delta}c_\omega(T)e^{-iT\omega\bar{u}},\end{aligned}$$

where we defined the Fourier spectrum of the fidelity as the set of all possible differences between frequencies in the control spectrum

$$\mathcal{S}^\Delta = \{\omega = \lambda' - \lambda | \lambda, \lambda' \in \mathcal{S}\} \tag{5}$$

and the fidelity Fourier coefficients as

$$c_\omega(T) = \sum_{\omega',\omega''\in\mathcal{S}}\delta_{\omega''-\omega',\omega}b^*_{\omega'}(T)b_{\omega''}(T).$$

B. General case $[\hat{H}_d, \hat{H}_c] \neq 0$

In general, the expansion in Eq. (4) does not stop for any finite n, and we obtain for the cost functional a sequence of partial Fourier sums $J_n(\boldsymbol{u})$ that converges to $J(\boldsymbol{u})$. Once again, we write the control Hamiltonian in its eigenspectrum $\hat{\Lambda}$ and absorb the change of basis in the terms $\hat{W}$ which do not depend on the control u and are defined implicitly via

$$\hat{U}_n(u) = \hat{V}^\dagger\underbrace{\hat{V}e^{-\frac{i\delta t}{n}\hat{H}_d}\hat{V}^\dagger}_{\hat{W}(n^{-1}\delta t)}e^{-\frac{i\delta t}{n}u\hat{\Lambda}}\cdots e^{-\frac{i\delta t}{n}\hat{H}_d}\hat{V}^\dagger e^{-\frac{i\delta t}{n}u\hat{\Lambda}}\hat{V}$$

where we multiplied a $\hat{V}^\dagger\hat{V} = \hat{I}$ factor on the left. When the dependence on n is not crucial, we omit it to simplify the notation. By expressing the result in terms of the matrix components, we finally obtain the Lie-Fourier representation of the unitary time-step evolution

$$\begin{aligned}\langle i|\,\hat{U}_n(u)\,|k\rangle &= \sum_{j_1\dots j_n l=1}^{D} e^{-i\delta t\frac{u}{n}(\lambda_{j_1}+\cdots+\lambda_{j_n})}V^\dagger_{il}W_{lj_n}(n^{-1}\delta t)\cdots W_{j_2j_1}(n^{-1}\delta t)V_{j_1k} \\ &= \sum_{\boldsymbol{j}\in[D]^n}e^{-i\delta tu\omega(\boldsymbol{j})}A^{\boldsymbol{j}}_{ik}(n,\delta t) = \sum_{\omega\in\mathcal{S}_n}e^{-i\delta tu\omega}B^\omega_{ik}(n,\delta t),\end{aligned} \tag{6}$$

where $[D]^n \subset \mathbb{N}^n$ is the set of integer vectors with elements ranging from 1 to D, and the sum over l is absorbed into the definition of the coefficients $A^{\boldsymbol{j}}_{ik}$, since the frequency ω does not depend on l:

$$A^{\boldsymbol{j}}_{ik}(n,\delta t) = \sum_{l=1}^{D}V^\dagger_{il}W_{lj_n}(n^{-1}\delta t)\cdots W_{j_2j_1}(n^{-1}\delta t)V_{j_1k}.$$

In the last step all the coefficients $A^{\boldsymbol{j}}_{ik}$ corresponding to the same frequency $\omega = \omega(\boldsymbol{j})$ have been summed up into the matrix with entries B^{ω}_{ik},

$$B^{\omega}_{ik}(n,\delta t) = \sum_{\boldsymbol{j}\in\mathcal{J}_\omega} A^{\boldsymbol{j}}_{ik}(n,\delta t),\ \mathcal{J}_\omega = \{\boldsymbol{j}\in[D]^n \mid \omega(\boldsymbol{j}) = \omega\}$$

which is only defined for values of ω belonging to the discrete Fourier spectrum $\mathcal{S}_n$

$$\mathcal{S}_n = \{\omega(\boldsymbol{j}) = \frac{1}{n}(\lambda_{j_1} + \cdots + \lambda_{j_n}) \mid \boldsymbol{j}\in[D]^n\}.$$

Now by layering the N time-step unitaries we can build up the full unitary $\hat{U}_n(\boldsymbol{u})$:

$$\begin{aligned}\hat{U}_n(\boldsymbol{u}) &= \hat{U}_n(u_N)\cdots\hat{U}_n(u_1)\\ &= \sum_{\boldsymbol{j}_1\cdots\boldsymbol{j}_N} e^{-i\delta t\sum_\nu u_\nu\omega_{j_\nu}}\hat{A}^{\boldsymbol{j}_N}\ldots\hat{A}^{\boldsymbol{j}_1}\\ &= \sum_{\omega_1\ldots\omega_N} e^{-i\delta t\sum_\nu u_\nu\omega_\nu}\hat{B}^{\omega_N}\cdots\hat{B}^{\omega_1}\\ &= \sum_{\boldsymbol{\omega}\in\mathcal{S}_n^N} e^{-i\delta t\boldsymbol{\omega}\cdot\boldsymbol{u}}\hat{B}^{\boldsymbol{\omega}}(n,\delta t), \quad (7)\end{aligned}$$

where we defined the frequency vector $\boldsymbol{\omega}\in\mathcal{S}_n^N$ and the operator $\hat{B}^{\boldsymbol{\omega}}$ as the product

$$\hat{B}^{\boldsymbol{\omega}} = \hat{B}^{\omega_N}\cdots\hat{B}^{\omega_1}.$$

By plugging this expression for the unitary operator back into Eq. (2) we finally find the Lie-Fourier representation of the fidelity:

$$\begin{aligned}J_n(\boldsymbol{u}) &= \langle\psi|\hat{U}_n^\dagger(\boldsymbol{u})|\chi\rangle\langle\chi|\hat{U}_n(\boldsymbol{u})|\psi\rangle\\ &= \sum_{\boldsymbol{\omega}',\boldsymbol{\omega}''\in\mathcal{S}_n^N}\underbrace{\langle\psi|\hat{B}^{\boldsymbol{\omega}''\dagger}|\chi\rangle}_{b^*_{\boldsymbol{\omega}''}}\underbrace{\langle\chi|\hat{B}^{\boldsymbol{\omega}'}|\psi\rangle}_{b_{\boldsymbol{\omega}'}} e^{-i\delta t(\boldsymbol{\omega}'-\boldsymbol{\omega}'')\cdot\boldsymbol{u}}\\ &= \sum_{\boldsymbol{\omega}\in(\mathcal{S}_n^\Delta)^N} c_{\boldsymbol{\omega}}(n,\delta t)e^{-i\delta t\boldsymbol{\omega}\cdot\boldsymbol{u}} \quad (8)\end{aligned}$$

where $(\mathcal{S}_n^\Delta)^N$ is the set of frequency differences within the Fourier spectrum of the unitary. The coefficients of the expansion $c_{\boldsymbol{\omega}}$ are given by

$$c_{\boldsymbol{\omega}} = \sum_{\boldsymbol{\omega}',\boldsymbol{\omega}''\in\mathcal{S}_n^N}\delta_{\boldsymbol{\omega},\boldsymbol{\omega}'-\boldsymbol{\omega}''}b^*_{\boldsymbol{\omega}''}b_{\boldsymbol{\omega}'} \quad (9)$$

We conclude this derivation by pointing out that there is a useful trick we can use since in the main text we are working with just one control per time step. In fact, we can equivalently write the fidelity as

$$J_n(\boldsymbol{u}) = \langle\tilde{\psi}|\hat{\tilde{U}}_n^\dagger(\boldsymbol{u})|\tilde{\chi}\rangle\langle\tilde{\chi}|\hat{\tilde{U}}_n(\boldsymbol{u})|\tilde{\psi}\rangle \quad (10)$$

by transforming the dynamics to a new unitary frame through $\hat{V}$, so that the initial state becomes $|\tilde{\psi}\rangle = \hat{V}|\psi\rangle$ and the propagator $\hat{\tilde{U}} = \hat{V}\hat{U}\hat{V}^\dagger$. In this new frame, the coefficients $\tilde{b}_\omega = b_\omega$ stay the same (they are scalars), while $\tilde{A}^{\boldsymbol{j}}_{ik}$ (and consequently $\tilde{B}^{\omega}_{ik}$) take a simpler form

$$\tilde{A}^{\boldsymbol{j}}_{ik}(n,\delta t) = W_{ij_n}(n^{-1}\delta t)\cdots W_{j_2j_1}(n^{-1}\delta t)\delta_{j_1k}.$$

C. Basic properties of the Lie-Fourier representation

Let us now prove some properties of the Lie-Fourier representation. It is easy to see from Eq. (5)-(9) that, compatibly with J and J_n being real functions, both the fidelity spectrum $\mathcal{S}_n^\Delta$ and the coefficients $c_{\boldsymbol{\omega}}$ are symmetric in the following way

$$\begin{aligned}\omega\in\mathcal{S}_n^\Delta &\implies -\omega\in\mathcal{S}_n^\Delta,\\ c_{-\boldsymbol{\omega}} &= c^*_{\boldsymbol{\omega}}\end{aligned}$$

Another straightforward result is that all the approximating functions J_n are bandwidth limited

Lemma 1 (Bandwidth limitation)**.** *$\mathcal{S}_n\in[\lambda_{\min},\lambda_{\max}]$ with $\lambda_{\min}(\lambda_{\max})$ the minimum (maximum) eigenvalue of $\hat{H}_c$.*

Proof.

$$\forall n\quad \min_{\boldsymbol{j}\in[D]^n}\omega_{\boldsymbol{j}} = \min_{\boldsymbol{j}\in[D]^n}\sum_{i=1}^n\frac{\lambda_{j_i}}{n} = \sum_{i=1}^n\min_{j\in[D]}\frac{\lambda_j}{n} = \lambda_{\min},$$

and the same is true for the max, which implies $\lambda_{\min}\le\omega\le\lambda_{\max}$. □

Moreover, let us consider the subset of frequencies $\mathcal{S}'_n\subset\mathcal{S}_n$ obtained by constraining the choice of eigenvalues to only $\lambda_{\min},\lambda_{\max}$:

$$\mathcal{S}'_n = \{\omega = \frac{m}{n}\lambda_{\max} + \frac{n-m}{n}\lambda_{\min},\ m = 0,\ldots,n\}.$$

It easy to see that this set is a regular grid over $[\lambda_{\min},\lambda_{\max}]$ with a step of $n^{-1}(\lambda_{\max}-\lambda_{\min})$, which is also larger than the maximum distance between any point in $[\lambda_{\min},\lambda_{\max}]$ and $\mathcal{S}'_n$. But then as $n\to\infty$ the frequencies in $\mathcal{S}'_n$ (and therefore in $\mathcal{S}_n$) will fill up that interval densely.

Similar properties hold also for the fidelity spectrum $\mathcal{S}_n^\Delta\subset[-\omega_{\max},\omega_{\max}]$ where we defined $\omega_{\max} = \lambda_{\max} - \lambda_{\min}$. Physically, this means that the landscape displays a typical scale $\sim\omega_{\max}^{-1}$ in control space below which there are no new details within a given tolerance. This information is relevant, for instance, in the context of optimization, as can potentially be used to set step sizes, filter out noise below that length scale and avoid oversampling. We will explore this idea more in detail later in Sec. VII.

To better understand the properties of the Fourier representation in the $n\to\infty$ limit, we first state two results concerning the boundedness of the coefficients:

Lemma 2 (L_1 Boundedness of the coefficients)**.**

$$\exists r\in\mathbb{R},\ s.t.\ \ \forall n,\ \sum_{\boldsymbol{\omega}\in(\mathcal{S}_n^\Delta)^N}|c_{\boldsymbol{\omega}}(n,\delta t)|\le r$$

Proof. See App. D. □

Lemma 3 (L_2 Boundedness of the coefficients)**.**

$$\forall n, \sum_{\boldsymbol{\omega}\in(\mathcal{S}_n^{\Delta})^N} |c_{\boldsymbol{\omega}}(n,\delta t)|^2 \leq 1$$

Proof. See App. C for a proof valid for multiple controls and a generic observable $\hat{O}$. □

Due to the exponential behaviour of r with respect to the dimensionality of the Hilbert space D (which is itself exponential in the number of subsystems) and the number of timesteps N, the L_1 bound is unlikely to be helpful in numerical predictions in general, given the results of this analysis. To obtain useful results concerning quantities that are linear in the coefficients (such as the landscape derivatives in Sec. IV), we will instead directly invoke the boundedness of J. Conversely, the L_2 bound is stronger because it is constant and, as we will see in Sec. VII, it is related to the landscape variance, as it is quadratic in the coefficients.

These bounds also suggest that the coefficients are well behaved in the large n limit. Following this intuition, we can show that the coefficients with frequencies corresponding to the eigenvalues of the control Hamiltonian —that is, matching the trivial commuting case (5)— contain in general a finite contribution, so we will refer to them as "resonant". Note that these will correspond to resonances in the landscape, i.e. generalized Rabi oscillations in the system output, which in general differ from the resonances in the underlying physical system because of the presence of the drift Hamiltonian.

If we fix $\boldsymbol{j} = (p,\dots,p)$ then clearly $\omega(\boldsymbol{j}) = \lambda_p$ and the corresponding contribution to the propagator coefficient $B_{ik}^{\lambda_p}$ is therefore

$$\begin{aligned} A_{ik}^{(p,\dots,p)}(n,\delta t) &= \lim_{n\to\infty} \sum_{l=1}^{d} V_{il}^{\dagger} W_{lp} \cdots W_{pp} V_{pk} \\ &= V_{ip}^{\dagger} \lim_{n\to\infty} (W_{pp})^n V_{pk}, \end{aligned}$$

where the only term in the sum which is non-zero in the limit is $l = p$. In fact, all the other terms contain an off-diagonal element of $\hat{W}$, which is vanishingly small for large n (similarly to App. D). Moreover, we have the first-order expansion

$$W_{ij} = [\hat{V} e^{-\frac{i\delta t}{n}\hat{H}_d}\hat{V}^{\dagger}]_{ij} = \delta_{ij} - i\frac{\delta t}{n}[\hat{V}\hat{H}_d\hat{V}^{\dagger}]_{ij} + o\left(\frac{\delta t}{n}\right).$$

Thus, we can exploit the continuity of the logarithm to write

$$\begin{aligned} \lim_{n\to\infty} (W_{pp}(n^{-1}\delta t))^n &= \exp \lim_{n\to\infty} n \log\big(W_{pp}(n^{-1}\delta t)\big) \\ &= \exp \lim_{n\to\infty} n(-i\frac{\delta t}{n}[\hat{V}\hat{H}_d\hat{V}^{\dagger}]_{pp} + o\left(\frac{\delta t}{n}\right)) \\ &= e^{-i\delta t[\hat{V}\hat{H}_d\hat{V}^{\dagger}]_{pp}}, \end{aligned}$$

which gives us, as anticipated, a finite result for a resonant landscape frequency. Indeed, in the next section we will see, using numerics, that the fidelity coefficients corresponding to differences in eigenvalues are typically finite, unless discrete symmetries decide otherwise, while all the other coefficients converge to 0 as $n \to \infty$, often creating a mostly continuous spectrum in that limit. We conclude by stating a result concerning the role of dynamical symmetries, which determine which frequencies can be non-zero:

Lemma 4 (Symmetries and selection rules)**.** *Let $\hat{\Gamma}$ be a symmetry of the system, that is $[\hat{\Gamma}, \hat{H}_c] = [\hat{\Gamma}, \hat{H}_d] = 0$. Let $|\gamma_{g(i)}, \lambda_i\rangle$ be the simultaneous eigenstates of $\hat{\Gamma}$ and $\hat{H}_c$ (the respective eigenvalues being γ_g and λ_i). Let $\mathcal{I}_g$ be the set of indices belonging to the g-th symmetry sector*

$$\mathcal{I}_g = \{i \in \{1,\dots,D\} \mid g(i) = g\},$$

so that the projector $\hat{P}_g$ onto that sector is

$$\hat{P}_g = \sum_{i\in\mathcal{I}_g} |\gamma_g, \lambda_i\rangle \langle\gamma_g, \lambda_i|,$$

and let $\mathcal{G}$ be the set of sector indices with corresponding non-zero overlap on both the initial and target states

$$\mathcal{G} = \{g \in \{1,\dots,G\} \mid \hat{P}_g|\psi\rangle \neq 0 \wedge \hat{P}_g|\chi\rangle \neq 0\}.$$

Then, the coefficients $b_{\boldsymbol{\omega}}$ of the cost functional are zero unless $\boldsymbol{\omega} \in {}^{(\Gamma,\mathrm{N})}\mathcal{S}_n$, with

$${}^{(\Gamma,\mathrm{N})}\mathcal{S}_n = \bigcup_{g\in\mathcal{G}} (\mathcal{S}_n^{(g)})^N,$$

and $\mathcal{S}_n^{(g)}$ the set of frequencies built from the g-th symmetry sector

$$\mathcal{S}_n^{(g)} = \{\omega(\boldsymbol{j}) = \frac{1}{n}(\lambda_{j_1}^{(g)} + \cdots + \lambda_{j_n}^{(g)}) \mid \boldsymbol{j} \in \mathcal{I}_g^n\}.$$

The coefficients $c_{\boldsymbol{\omega}}$ are zero unless $\boldsymbol{\omega} \in {}^{(\Gamma,\mathrm{N})}\mathcal{S}_n^{\Delta}$, with

$${}^{(\Gamma,\mathrm{N})}\mathcal{S}_n^{\Delta} = \{\boldsymbol{\omega} = \boldsymbol{\omega'} - \boldsymbol{\omega''} | \boldsymbol{\omega'}, \boldsymbol{\omega''} \in {}^{(\Gamma,\mathrm{N})}\mathcal{S}_n\}.$$

Proof. See App. D. □

In essence, the lemma establishes that the eigenvalues that can appear in the Lie-Fourier representation of the cost landscape can only originate from the same symmetry sector at every time step, where the relevant sectors must contain a projection of the initial and final states.

This result relies on the decomposition of the dynamics into invariant subspaces, which causes the unitaries to have a block diagonal matrix representation. Another known consequence is that the dynamical Lie algebra of the generators factorizes, simplifying the dynamics and putting constraints on the appearance of barren

plateaux and on classical non-computability of the dynamics [7, 33]. In light of this, and given the relation between Fourier spectrum and landscape variance that we will explore in Sec. VII, we interpret the frequency selection rules as one of the manifestations of this algebraic phenomenon in the context of the Fourier representation of the landscape. This provides an alternative viewpoint on the subject and informs the further development of classical simulation algorithms [20].

D. Numerical examples

We will now pick a specific system and see how the theory detailed earlier in this section comes into play. The system under consideration is the one dimensional transverse field Ising model

$$\hat{H}[u(t)] = \sum_{i=1}^{Q} \alpha_d(\hat{\sigma}_z^{(i)}\hat{\sigma}_z^{(i+1)} + h_z\hat{\sigma}_z^{(i)}) + u(t)\hat{\sigma}_x^{(i)} \quad (11)$$

where the Hilbert space $\mathcal{H}$, with $D := \dim \mathcal{H} = 2^Q$ is the product of Q qubits and $\hat{\sigma}_a^{(i)}$ is the Pauli operator for qubit i. We use the transverse field along the x direction as control, while the constant α_d plays the role of drift Hamiltonian strength. The parameter h_z can be used to turn on the longitudinal field, and we fix it to zero unless specified otherwise. We assume periodic boundary conditions, so that as far as the qubit indices are concerned $Q+1 \equiv 1$. The Ising model is a paradigmatic example of quantum dynamics [34], and its state transfer fidelity landscape has already been object of study [2, 3]. Moreover, circuit Ansätze featuring the generators in Eq. (11) (or variations thereof) are often used within variational algorithms such as QAOA [35] and their dynamical symmetries have been studied in depth [7, 33, 36].

1. *Ising model spectrum*

As we will see now, the Ising model also gives rise to a remarkably simple Lie-Fourier representation. In fact, as we see from the eigenstates of $\hat{H}_c$, namely

$$\sum_{i=1}^{Q} \hat{\sigma}_x^{(i)} \left|-^{b_1}\cdots-^{b_Q}\right\rangle = \sum_{i=1}^{Q} (-1)^{b_i} \left|-^{b_1}\cdots-^{b_Q}\right\rangle, \quad (12)$$

where we adopt the convention $|-^0\rangle = |+\rangle, |-^1\rangle = |-\rangle$, the spectrum $\mathcal{S}$ of the control Hamiltonian

$$\mathcal{S} = \{-Q, -Q+2, \ldots, Q-2, Q\}$$

is highly degenerate (the number of distinct eigenvalues increases linearly instead of exponentially in Q, as in the non-degenerate scenario). In turn this gives rise to the following Fourier spectra for the single-timestep unitary and fidelity

$$\begin{aligned}\mathcal{S}_n &= \{-Q, -Q+\frac{2}{n}, \ldots, Q-\frac{2}{n}, Q\},\\ \mathcal{S}_n^{\Delta} &= \{-2Q, -2Q+\frac{2}{n}, \ldots, 2Q-\frac{2}{n}, 2Q\}\\ &= \{\omega = \omega_{\max}\frac{k}{k_{\max}} \mid k = -k_{\max}, \ldots, k_{\max}\}\end{aligned}$$

with $k_{\max} = Qn$, so $\omega_{\max} = 2Q$, $\#\mathcal{S}_n = k_{\max}+1$, $\#\mathcal{S}_n^{\Delta} = 2k_{\max}+1 =: n_{\Delta}$. We notice that for this model all the frequencies in $\mathcal{S}_n^{\Delta}$ are equally spaced, which allows us to compute numerically the $c_{\boldsymbol{\omega}}$ coefficients by means of the Discrete Fourier Transform (DFT) (refer to App. E for details).

By applying this numerical algorithm, we study the Lie-Fourier representation of several state transfer problems for the Ising model. More specifically, we consider the transfer between $|0_Q\rangle \mapsto |1_Q\rangle$ (also for non-zero longitudinal field $h_z \neq 0$), between eigenstates of the control Hamiltonian $|+_Q\rangle \mapsto |-_Q\rangle$ and between random states $|r\rangle \mapsto |r'\rangle$ [37]. In Fig. 2 we can see plotted the coefficients of the Lie-Fourier expansion for $N = 1, 2$ for each case.

Let us start by looking at the case $|0_Q\rangle \mapsto |1_Q\rangle$. As expected from the discussion in the previous subsection, we see that the coefficients corresponding to differences in eigenvalues correspond to finite contributions, which appear as sharp peaks, since we plotted $n^N c_{\omega}$. This plotting choice was adopted to emphasize the behaviour of the non-resonant coefficients, which we did not discuss extensively. We observe that numerically the coefficients give rise to an apparently smooth function within the intervals delimited by the resonant frequencies. This suggests that, at least in this particular case, the non-resonant part of the Lie-Fourier expansion approximates the Fourier Transform of a well behaved function.

2. *Missing resonances*

If now instead we consider the state transfer problem $|+_Q\rangle \mapsto |-_Q\rangle$, we see that some of the frequencies that were resonant in the previous case do not display a sharp peak anymore. The reason behind this lies in the discrete symmetries of the Hamiltonian in Eq. (11). Let us define the following operator(s):

$$\hat{\Sigma}_{\alpha} = \bigotimes_{i=1}^{Q} \hat{\sigma}_{\alpha}^{(i)}, \quad \alpha = x, y, z,$$

which are Hermitian $\hat{\Sigma}_{\alpha}^{\dagger} = \hat{\Sigma}_{\alpha}$ and satisfy $\hat{\Sigma}_{\alpha}^2 = \hat{I}$. One can easily check that $\hat{\Sigma}_x$ is a symmetry of both generators $[\hat{H}_d, \hat{\Sigma}_x] = [\hat{H}_c, \hat{\Sigma}_x] = 0$ with eigenvalues ± 1 and eigenvectors given by

$$\hat{\Sigma}_x \left|-^{b_1}\cdots-^{b_Q}\right\rangle = (-1)^{\sum_{i=1}^{Q} b_i} \left|-^{b_1}\cdots-^{b_Q}\right\rangle.$$

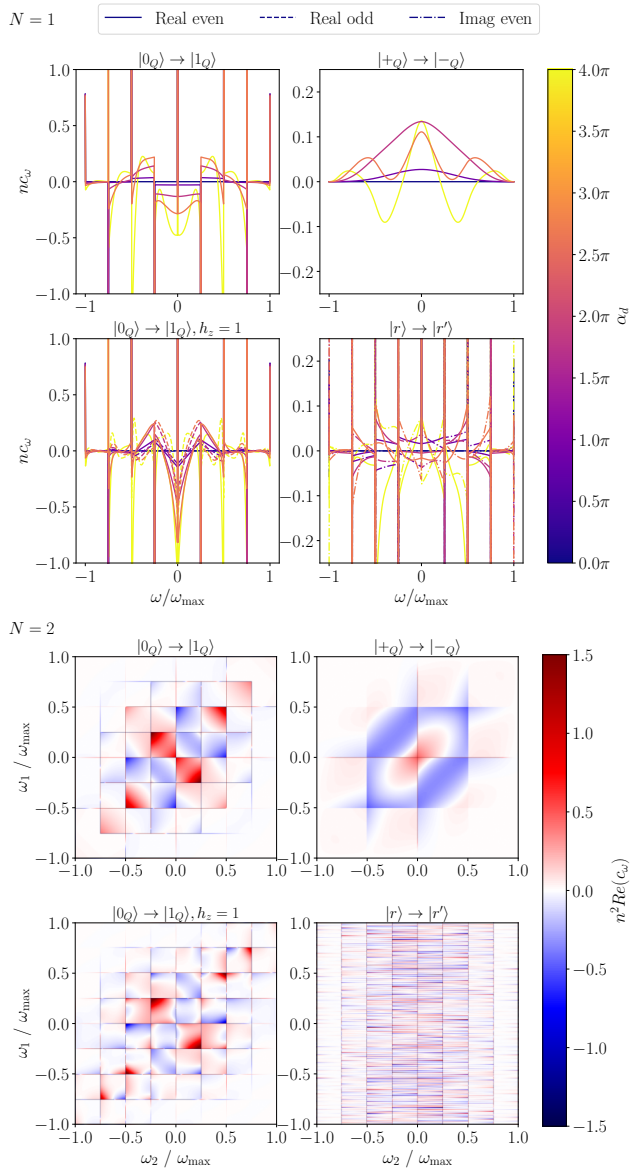

FIG. 2: Lie-Fourier representation coefficients of the Ising model dynamical landscape for selected state transfer problems. The results are computed for $n = 200$ using the DFT algorithm described in App. E. (Upper panel) The c_ω coefficients for a single timestep landscape $N = 1$ are plotted for several values of drift strength $\alpha_D \pi^{-1} = 0.0, 0.89, 1.78, 2.67, 4.0$ (color scale). The line styles correspond to the real part of the even (solid) and odd (dashed) frequency index branches and to the imaginary part of the even branch (dot-dashed). The imaginary part of the odd branch is numerically zero in all cases. (Lower panel) The real parts of the c_ω coefficients for a double timestep landscape $N = 2$ are plotted as a function of the vector frequencies $\boldsymbol{\omega}$ for $\alpha_D \pi^{-1} = 5.59$. In all cases but for random states $|r\rangle \to |r'\rangle$ the coefficients can be reorganized in two branches which exhibit continuous behaviour inside intervals defined by the resonant frequencies.

Comparing this expression with Eq. (12), we can categorize the eigenvalues in $\mathcal{S}$ according to the symmetry sector they belong to:

$$\mathcal{S}^+ = \begin{cases} \{Q, Q-4, \ldots, -Q+4, -Q\} \text{ if } Q \text{ even} \\ \{Q, Q-4, \ldots, -Q+6, -Q+2\} \text{ if } Q \text{ odd} \end{cases}$$

$$\mathcal{S}^- = \begin{cases} \{Q-2, Q-6, \ldots, -Q+6, -Q+2\} \text{ if } Q \text{ even} \\ \{Q-2, Q-6, \ldots, -Q+4, -Q\} \text{ if } Q \text{ odd} \end{cases}$$

with $\mathcal{S} = \mathcal{S}^+ \cup \mathcal{S}^-$. Moreover, the initial and target states lie inside a single symmetry sector:

$$\hat{\Sigma}_x |+_Q\rangle = |+_Q\rangle ,$$
$$\hat{\Sigma}_x |-_Q\rangle = (-1)^Q |-_Q\rangle .$$

Then, because of Lemma 4, if Q is odd all coefficients $b_{\boldsymbol{\omega}}$ and, therefore, $c_{\boldsymbol{\omega}}$, are zero. If instead Q is even, the spectrum will be limited to the eigenvalues of the $+1$ symmetry sector, that is

$$\mathcal{S}_n^+ = \{Q, Q - \frac{4}{n}, \ldots, -Q + \frac{4}{n}, -Q\}$$

which gives rise to the fidelity spectrum

$$\mathcal{S}_n^{+\Delta} = \{2Q, 2Q - \frac{4}{n}, \ldots, -2Q + \frac{4}{n}, -2Q\}.$$

This excludes from the spectrum roughly half of the frequencies, among which the resonant frequencies that do not appear in the $|+_Q\rangle \mapsto |-_Q\rangle$ case in Fig. 2. Since this symmetry is broken in all the other cases, the resonances are in general not suppressed and can appear as sharp peaks.

3. *Additional spectrum features*

It is interesting to see that, in all cases except the transfer between random states, the coefficients $c_{\boldsymbol{\omega}} \in \mathbb{C}$ are real. As far as the $|0_Q\rangle \mapsto |1_Q\rangle$ case is concerned (also for $h_z \neq 0$), this fact can be explained by the existence of additional anti-symmetries of the Ising model. These are given by the operators $\hat{\Sigma}_{y,z}$ and have the property $\hat{\Sigma}_{y,z} \hat{H}(u) \hat{\Sigma}_{y,z} = \hat{H}(-u)$. But then we have the following:

$$\begin{aligned} J(\boldsymbol{u}) &= |\langle 1_Q| \hat{U}(\boldsymbol{u}) |0_Q\rangle|^2 \\ &= |\langle 1_Q| \hat{\Sigma}_z^2 \hat{U}(u_N) \hat{\Sigma}_z^2 \cdots \hat{\Sigma}_z^2 \hat{U}(u_1) \hat{\Sigma}_z^2 |0_Q\rangle|^2 \\ &= |(-1)^Q \langle 1_Q| \hat{U}(-u_N) \cdots \hat{U}(-u_1) |0_Q\rangle|^2 = J(-\boldsymbol{u}) \end{aligned}$$

which together with $J \in \mathbb{R}$ implies $c_{\boldsymbol{\omega}} \in \mathbb{R}$. A similar reasoning using $\hat{\Sigma}_x$ shows another interesting property:

$$\begin{aligned} J(u_1, \ldots, u_N) &= |\langle 1_Q| \hat{U}(\boldsymbol{u}) |0_Q\rangle|^2 \\ &= |\langle 1_Q| \hat{\Sigma}_x^2 \hat{U}(u_N) \hat{\Sigma}_x^2 \cdots \hat{\Sigma}_x^2 \hat{U}(u_1) \hat{\Sigma}_x^2 |0_Q\rangle|^2 \\ &= |\langle 0_Q| \hat{U}(u_N) \cdots \hat{U}(u_1) |1_Q\rangle|^2 \\ &= |\langle 1_Q| \hat{U}^\dagger(u_1) \cdots \hat{U}^\dagger(u_N) |0_Q\rangle|^2 = J(u_N, \ldots, u_1), \end{aligned}$$

where the last step is justified since both $\hat{H}$ and the states $|0_Q\rangle, |1_Q\rangle$ only have real matrix elements and overlaps with the computational basis (see App. E for more details). This property shows that in this case the Lie-Fourier coefficients are approximately symmetric under reversal of the frequency order $c_{\omega_1\cdots\omega_N} = c_{\omega_N\cdots\omega_1}$. The reason why the symmetry is not exact is that in the Lie-Fourier representation the unitary $\hat{U}$ is substituted with $\hat{U}_n$, which obeys the symmetry exactly only in the $n \to \infty$ limit.

Another phenomenon that we observe is the splitting of the coefficients into branches featuring apparently continuous behaviour inside the intervals described above. In all cases without longitudinal field $h_z = 0$, the branch corresponding to frequencies in $\mathcal{S}_n^{\Delta}$ with odd indices are zero [38]. Moreover, the even/odd branches appear to be the only ones present in all cases except the random state transfer. In this last case, the structure looks much more complex, potentially featuring many branches and more discontinuities. We believe this to be related to the fact that randomly chosen states break not only the symmetries we discussed, but also the ones associated to cyclic qubit permutations [36].

The observed stepwise-continuous behaviour seems to suggest that at least in symmetric cases it would be advantageous to choose as feature maps a basis of e.g. polynomial functions in frequency space. This way large swaths of the continuous spectrum could be represented by a few basis functions, resulting in a more efficient representation. Even the discrete symmetries we highlighted could be explicitly encoded within such a feature map. While leaving this for future work, we notice that we can achieve a similar effect with the sinc-kernel representation as presented in Sec. VI.

IV. LANDSCAPE DERIVATIVES

The Lie-Fourier representation we discussed in Section III also allows us to prove some results concerning the derivatives of the landscape, which are independent of the representation itself. In fact we have that

Lemma 5 (Boundedness of the derivatives)**.** *The partial derivatives of $J(\boldsymbol{u})$ of any order P are bounded by*

$$|\partial_1^{p_1}\cdots\partial_N^{p_N} J(\boldsymbol{u})| \leq \frac{(\omega_{\max}\delta t)^P}{2}$$

where $P = \sum_\nu p_\nu$ and $\omega_{\max} = |\lambda_{\max} - \lambda_{min}|$ is the maximum transition frequency in the control Hamiltonian.

Proof. See App. C for a proof in the case of multiple controls and a generic observable $\hat{O}$. □

The proof first shows that the result holds for the Lie-Fourier approximants J_n, making use of the fact that they are bandwidth limited and bounded $0 \leq J_n \leq 1$. The result for the landscape J is obtained by invoking uniform convergence of the functions J_n and of their derivatives of any order.

As a direct consequence we notice that the L_1 norm of the gradient, Hessian and higher order derivative tensors are bounded by constants:

$$\sum_{\boldsymbol{p}\in\mathbb{N}^N} \delta_{||\boldsymbol{p}||_1,P} |\partial_1^{p_1}\cdots\partial_N^{p_N} J(\boldsymbol{u})| \leq \frac{(N\omega_{\max}\delta t)^P}{2} = \frac{L^P}{2},$$

where we defined the nondimensional parameter $L = \omega_{\max} T$, with $T = N\delta t$ the total evolution time, and $\boldsymbol{p} = (p_1, p_2, \cdots, p_N)^T$. Since the parameter L does not depend on the circuit depth N alone, this points to the bounds being relevant also in the case of continuous controls, which are approximated by the stepwise-constant pulses that we studied.

Another consequence of the bound on the derivatives is given by the following result:

Lemma 6 (Lipschitz continuity)**.** *The function $J(\boldsymbol{u})$ is Lipschitz continuous, that is*

$$|J(\boldsymbol{u}) - J(\boldsymbol{u'})| \leq K||\boldsymbol{u} - \boldsymbol{u'}||_1$$

where the Lipschitz constant satisfies $0 \leq K \leq \omega_{\max}\delta t/2$. Moreover, if $\boldsymbol{u}$ is a critical point $\boldsymbol{\nabla} J(\boldsymbol{u}) = \boldsymbol{0}$, then also the following inequality holds

$$|J(\boldsymbol{u}) - J(\boldsymbol{u'})| \leq K_c||\boldsymbol{u} - \boldsymbol{u'}||_1^2.$$

where $0 \leq K_c \leq (\omega_{\max}\delta t)^2/2$.

Proof. See App. C for a proof valid in the case of multiple controls and a generic observable $\hat{O}$. □

In practical terms, this means that the maximum amount of variation in the landscape value is fixed through the constant rate K to the distance between sampled points. This fact has important consequences for optimization that we will explore in Sec. VII-VIII. Besides computing an upper bound for K in terms of the physical properties of the systems, this results clarifies what is the relevant notion of distance between controls that should be used in this case, namely the L_1 norm $||\cdot||_1$ (also known as "taxicab" norm, see App. A for definitions). In order to study the scaling for large N, we can reabsorb the factor δt inside the norm

$$K\sum_{\nu=1}^{N} \delta t|u(t_\nu) - u'(t_\nu)| \xrightarrow{N\to\infty} \frac{\omega_{\max}}{2}\int_0^T dt|u(t) - u'(t)|.$$

We can also show that the properties we just discussed still apply if we work with parametrized controls, which is often the case in optimal control [24, 39, 40]. The following corollary to Lemma 5 shows that, provided the basis functions are appropriately normalized, the same bounds on the derivatives, and therefore on the Lipschitz constant, are satisfied:

Corollary 1 (Linear parametrizations of controls)**.** *Let us consider a linear parametrization of the controls* $\boldsymbol{u} = \boldsymbol{R}\boldsymbol{v}$*, where* $\boldsymbol{v} \in \mathbb{R}^{N_c}$ *and* $\boldsymbol{R} \in \mathbb{R}^{N\times N_c}$*, which is normalized as follows:*

$$\forall i = 1, \dots, N_c \quad ||\boldsymbol{r}_i||_1 := \sum_{\nu=1}^{N} |r_{\nu i}| = 1.$$

where $\boldsymbol{r}_i$ *are the columns of* $\boldsymbol{R}$*.*

Then, the derivatives of the parametrized landscape $\tilde{J}(\boldsymbol{v}) = J(\boldsymbol{R}\boldsymbol{v})$ *obey the bound*

$$|\tilde{\partial}_1^{p_1} \cdots \tilde{\partial}_{N_c}^{p_{N_c}} \tilde{J}(\boldsymbol{v})| \leq \frac{(\omega_{\max}\delta t)^P}{2} \prod_{i=1}^{N_c} ||\boldsymbol{r}_i||_1^{p_i} = \frac{(\omega_{\max}\delta t)^P}{2},$$

where $P = \sum_\nu p_\nu$*,* $\tilde{\partial}_{\mu'\nu'} := \frac{\partial}{\partial v_{\mu'\nu'}}$ *and* $\partial_{\mu\nu} := \frac{\partial}{\partial u_{\mu\nu}}$

Proof. See App. C for a proof valid in the case of multiple controls and a generic observable $\hat{O}$. □

Once again, we can derive an appropriate formula for scaling in N by including the factor δt in the normalization of the basis functions.

V. TAYLOR REPRESENTATION

The bound on the landscape derivatives that we found in Lemma 5 also gives important information regarding how efficiently the landscape can be locally represented by means of a Taylor expansion. By taking the reference pulse $\boldsymbol{u}_0$ as the expansion point, the Taylor representation up to order P can be written in the following form [41] [42]

$$J_P(\boldsymbol{u}) = \sum_{p=0}^{P} \sum_{\nu_1\cdots\nu_p} a_{\boldsymbol{\nu}} (\boldsymbol{u}-\boldsymbol{u}_0)_{\nu_1} \cdots (\boldsymbol{u}-\boldsymbol{u}_0)_{\nu_p},$$
$$a_{\boldsymbol{\nu}} = \frac{1}{p!} \partial_{\nu_1} \cdots \partial_{\nu_p} J(\boldsymbol{u}_0),$$

In order to see that the landscape admits a representation in this form, we find an upper bound for the approximation error by means of Lemma 5 and of the following classical result due to Lagrange and Taylor:

Lemma 7 (Taylor approximation error in the Lagrange form)**.**

$$|J(\boldsymbol{u}) - J_P(\boldsymbol{u})| \leq \frac{1}{2} \frac{(\omega_{\max}\delta t)^{P+1}}{(P+1)!} ||\boldsymbol{u}-\boldsymbol{u}_0||_1^{P+1}$$

Proof. See App. C for a proof valid for multiple controls and a generic observable $\hat{O}$. □

Inside the control region of interest $\mathcal{C}^{(N)}(\boldsymbol{u}_0)$ the bound can be relaxed to

$$\sup_{\boldsymbol{u}\in\mathcal{C}^{(N)}(\boldsymbol{u}_0)} |J(\boldsymbol{u}) - J_P(\boldsymbol{u})| \leq \frac{(u_{\max}L)^{P+1}}{2(P+1)!} =: \epsilon(P), \quad (13)$$

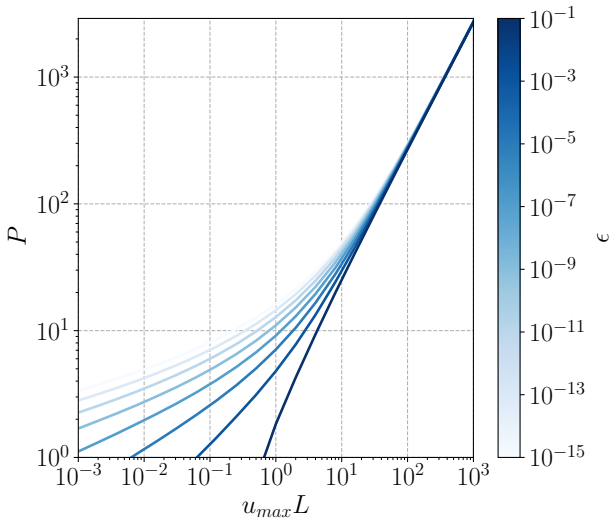

FIG. 3: Given an error threshold ϵ, we plot the solution for P of the equation $\epsilon(P) = \epsilon$. This quantity represents the minimum order of Taylor expansion P to represent J up to an error ϵ. The different lines show the results for several values for the error threshold $\epsilon = \{10^{-1}, 10^{-3}, 10^{-5}, 10^{-7}, 10^{-9}, 10^{-11}, 10^{-13}, 10^{-15}\}$. There is a crossover from $u_{\max}L < 1$, where $P \lesssim 10$ and it depends weakly on $u_{\max}L$, enabling an efficient local representation of the landscape, to the region $u_{\max}L \gg 1$ where the dependence becomes linear $P \sim e u_{\max}L$.

where we recall that $L = \omega_{\max}N\delta t$. Intuitively, this implies that for $u_{\max} < L^{-1}$, the error of the $P-$order expansion is suppressed by both the numerator (exponentially) and the denominator (factorially), defining a control region of improved convergence in P. On the other hand, for $u_{\max} \geq L^{-1}$ the numerator diverges exponentially, and the factorial suppression of the denominator must first kick in for the error to vanish. Arbitrary precision can be obtained in both cases by choosing P to be large enough.

In Fig. 3 we study what is the minimum required order of the Taylor expansion so that $\epsilon(P) = \epsilon$ for a given error threshold ϵ. As expected, we see that for $u_{\max}L < 1$ the we do not need very high order expansions to obtain high precision, whereas for $u_{\max}L > 1$ the order required for the expansion scales linearly with $u_{\max}L$. The latter observation can be understood by developing the bound in Eq. (13) using Stirling's approximation for the factorial under the assumption $u_{\max}L, P \gg 1$:

$$0 \approx \frac{\log(\epsilon)}{P} \approx \frac{1}{P}\log\left(\frac{(u_{\max}L)^P}{2P!}\right) \approx \left(\log\frac{u_{\max}L}{P} + 1\right),$$

which implies the linear dependence $P \approx e u_{\max}L$.

This error analysis shows that the landscape can be represented locally (i.e. for $u_{\max}L \lesssim 1$, which is the case

if the pulses amplitude $u_{\max}$, total time T and/or maximum transition frequency $\omega_{\max}$ are sufficiently small) with a number of parameters which is polynomial in the number of time steps N, with a degree $P \lesssim 10$. As an example, Fig. 3 shows that it is possible to represent the landscape with an error below $\epsilon \lesssim 10^{-3}$ for $u_{\max}L < 1$ with a $P = 5$ representation, which contains $\mathcal{O}(N^5)$ coefficients or even more locally for $u_{\max}L \lesssim 0.25$ with a $P = 2$ representation, which only contains $\mathcal{O}(N^2)$ terms (corresponding to the gradient and Hessian of the landscape at the expansion point).

Such local representations provide relevant information during optimization and can be used to obtain noise-robust gradient estimates or to allow a direct jump to the minimum of the local landscape, speeding up convergence [16, 43]. In particular, for $P = 2$ the latter approach is equivalent to applying Newton's method [43], for which our results could help bound the convergence rate. Owing to its generality, the Taylor representation also serves as a performance baseline for any other local representation which exploits specific details of the system at hand [16].

In the case of bang-bang controls $s_\nu = \pm 1$, the Taylor representation is very similar to the one presented in [2], although in that work the landscape is related to the logarithm of the infidelity. Since the controls are chosen this way, the only allowed powers are $p_\nu = 0, 1$ and therefore the expansion is finite. The fact that the authors of [2] could successfully represent the landscape for small values of T using only the first few orders in their expansion is qualitatively similar to the picture that arises from our analysis. In our case we can roughly choose the timescale T^* that defines the crossover as

$$T^* = \frac{1}{u_{\max}\omega_{\max}}.$$

As we will see in Sec. VII, T^* can be interpreted as a Quantum Speed Limit, so for $T/T^* \ll 1$ the infidelity cannot reach values arbitrarily close to 0. But then the composition of J with the logarithm should only add a prefactor close to 1 to the bounds we discussed, leading to a qualitatively similar picture.

VI. KERNEL REPRESENTATION AND LANDSCAPE LEARNING

We now turn to the question regarding how it is possible, given a sample $\mathcal{D}_{\text{train}}$ of N_{train} control values $\boldsymbol{u}_i$ and the corresponding fidelities $J(\boldsymbol{u}_i) =: J_i$, namely

$$\mathcal{D}_{\text{train}} = \{(\boldsymbol{u}_i, J_i)\}_{i=1}^{N_{\text{train}}}$$

to efficiently predict the fidelity for new control values.

This problem can be stated in the language of machine learning as a supervised learning (or regression) task, where given a family of models of the landscape $J_{\boldsymbol{w}}(\boldsymbol{u})$ parametrized by the weights $\boldsymbol{w} \in \mathbb{C}^{N_{\text{weights}}}$, we look for the vector $\bar{\boldsymbol{w}}$ whose associated model best fits the points in the training data set. This is quantified by means of a loss function $\mathfrak{L}$, that we choose to be the sum of the square residuals with RIDGE regularization [2, 17, 27]:

$$\mathfrak{L}(\boldsymbol{w}, \boldsymbol{w}^\dagger, \mathcal{D}_{\text{train}}) = \sum_{i=1}^{N_{\text{train}}} |J_{\boldsymbol{w}}(\boldsymbol{u}_i) - J_i|^2 + \lambda_R \boldsymbol{w}^\dagger \boldsymbol{w},$$

so that $\bar{\boldsymbol{w}}$ will correspond to the minimum of $\mathfrak{L}$. If the considered model has a non-linear dependence on the parameters, which is the case for deep learning [3], one usually has to resort to gradient-based optimization to find $\bar{\boldsymbol{w}}$. In the case of a linear model, that is

$$J_{\boldsymbol{w}}(\boldsymbol{u}) = \boldsymbol{w}^\dagger \boldsymbol{\phi}(\boldsymbol{u}), \tag{14}$$

where the non-linear functions $\boldsymbol{\phi}(u) = (\phi_1(u), \phi_2(u), \cdots, \phi_{N_{\text{weights}}}(u))^T : \mathbb{R}^N \to \mathbb{C}^{N_{\text{weights}}}$ are usually called features, the problem reduces to linear regression, which can be solved algebraically. This is possible since in this case the loss reduces to a quadratic form in the weights:

$$\mathfrak{L}(\boldsymbol{w}, \boldsymbol{w}^\dagger) = \boldsymbol{w}^\dagger \boldsymbol{\Phi}^\dagger \boldsymbol{\Phi} \boldsymbol{w} + \boldsymbol{J}^\dagger \boldsymbol{J} - \boldsymbol{J}^\dagger \boldsymbol{\Phi} \boldsymbol{w} - \boldsymbol{w}^\dagger \boldsymbol{\Phi}^\dagger \boldsymbol{J} + \lambda_R \boldsymbol{w}^\dagger \boldsymbol{w},$$

where we have defined the so-called feature matrix $\boldsymbol{\Phi}$

$$\boldsymbol{\Phi}^\dagger = \begin{pmatrix} \boldsymbol{\phi}(\boldsymbol{u}_1) & \dots & \boldsymbol{\phi}(\boldsymbol{u}_{N_{\text{train}}}) \end{pmatrix}.$$

The stationary points $\boldsymbol{\nabla}_{\boldsymbol{w}^\dagger} \mathfrak{L} = \boldsymbol{0}$ can be found by solving the linear equation

$$\boldsymbol{w} = -\frac{1}{\lambda_R} \boldsymbol{\Phi}^\dagger (\boldsymbol{\Phi} \boldsymbol{w} - \boldsymbol{J}). \tag{15}$$

Since the loss is real, differentiating with respect to $\boldsymbol{w}$ gives rise to the Hermitian conjugate of this equation, which is also satisfied when this equation is satisfied. The solution can be written explicitly via matrix inversion:

$$\bar{\boldsymbol{w}} = \left(\boldsymbol{\Phi}^\dagger \boldsymbol{\Phi} + \lambda_R \mathbb{1}\right)^{-1} \boldsymbol{\Phi}^\dagger \boldsymbol{J}.$$

We see from this expression that the RIDGE parameter λ_R regularizes the inverse of the covariance matrix $\boldsymbol{\Phi}^\dagger \boldsymbol{\Phi} \in \mathbb{C}^{N_{\text{weights}} \times N_{\text{weights}}}$, avoiding potential problems arising from small eigenvalues of $\boldsymbol{\Phi}^\dagger \boldsymbol{\Phi}$, which can be caused by correlations in the data set.

In light of the Lie-Fourier representation we discussed in Sec. III, a natural choice for the features is a complex exponential of the form

$$\phi_{\boldsymbol{\omega}}(\boldsymbol{u}) = e^{-i\delta t \boldsymbol{\omega} \cdot \boldsymbol{u}}.$$

Since the frequencies fill up densely the hypercube $\boldsymbol{\omega} \in [-\omega_{\max}, \omega_{\max}]^N$, in the absence of a prior expectation on their distribution, we consider random Fourier features [17], which we sample with uniform probability in the frequency hypercube. As we will see later, the boundedness of the control region we are considering results in a certain freedom to choose the frequencies in the model,

so that they do not necessarily have to match the ones from the Lie-Fourier representation. This can potentially relieve us from having to deal with an infinite number of frequencies.

Another viable option is to use polynomial features of the form

$$\phi_{\boldsymbol{p}}(\boldsymbol{u}) = \prod_{\nu=1}^{N} (\boldsymbol{u} - \boldsymbol{u}_0)^{p_1} \cdots (\boldsymbol{u} - \boldsymbol{u}_0)^{p_N},$$

which give rise to a Taylor representation like the one we studied in Sec. V, provided that the degree vectors $\boldsymbol{p}$ are chosen appropriately. Based on our discussion of the error, we would expect this representation to efficiently learn the landscape at least for $u_{\max} L < 1$.

These two options both suffer from the same problem, namely, that in order to solve the problem numerically they require subselecting a finite number of representative features N_{weights} from an infinite-dimensional feature space. This space is in general needed in full to represent the landscape exactly, as we saw in detail in Sec. III for the Lie-Fourier representation and in Sec. V for the Taylor representation. But then, we always have to neglect features that might be important, which is potentially problematic in the absence of guiding principles on how to choose the ones we keep.

There is a way around this problem that allows us to effectively perform infinite-dimensional linear regression, by means of the so-called kernel trick [17, 27]. The starting point of this method is to introduce the column vector $\boldsymbol{a}$ as follows:

$$\boldsymbol{a} = -\frac{1}{\lambda_R}(\boldsymbol{\Phi}\boldsymbol{w} - \boldsymbol{J})$$

so that together with Eq. (15) we have

$$\boldsymbol{w}(\boldsymbol{a}) = \boldsymbol{\Phi}^\dagger \boldsymbol{a}.$$

We can now rewrite the loss as a function of $\boldsymbol{a}$ alone:

$$\mathfrak{L}(\boldsymbol{a}, \boldsymbol{a}^\dagger) = \boldsymbol{a}^\dagger \boldsymbol{K}\boldsymbol{K}\boldsymbol{a} - \boldsymbol{a}^\dagger \boldsymbol{K}\boldsymbol{J} - \boldsymbol{J}^\dagger \boldsymbol{K}\boldsymbol{a} + \boldsymbol{J}^\dagger \boldsymbol{J} + \lambda_R \boldsymbol{a}^\dagger \boldsymbol{K}\boldsymbol{a},$$

where we defined the kernel matrix $\boldsymbol{K} = \boldsymbol{\Phi}\boldsymbol{\Phi}^\dagger$, turning the original infinite-dimensional regression problem into a finite dimensional kernel regression. Since $\boldsymbol{K}^\dagger = \boldsymbol{K} \in \mathbb{C}^{N_{\text{train}} \times N_{\text{train}}}$, the dimensionality of this new regression problem is set by the number of data points N_{train} in the training dataset. Once again, we have to minimize the loss by solving $\boldsymbol{\nabla}_{\boldsymbol{a}^\dagger} \mathfrak{L} = \boldsymbol{0}$, which yields

$$\bar{\boldsymbol{a}} = (\boldsymbol{K} + \lambda_R \mathbb{1})^{-1} \boldsymbol{J}.$$

Finally, inference on a new data point is to be carried out by substitution inside Eq. (14)

$$J_{\boldsymbol{w}(\bar{\boldsymbol{a}})}(\boldsymbol{u}) = \bar{\boldsymbol{a}}^\dagger \boldsymbol{\Phi}\boldsymbol{\phi}(\boldsymbol{u}) = \sum_{i=1}^{N_{\text{train}}} \bar{a}_i^* \kappa(\boldsymbol{u}_i, \boldsymbol{u}),$$

where we defined the kernel function $\kappa(\boldsymbol{u}, \boldsymbol{u}') := \boldsymbol{\phi}^\dagger(\boldsymbol{u})\boldsymbol{\phi}(\boldsymbol{u}')$, which also appears in the matrix elements of the kernel matrix $K_{ij} = \kappa(\boldsymbol{u}_i, \boldsymbol{u}_j)$. Kernel methods like this one are categorized as instance-based learning methods, because the prediction for a new data point is obtained as a linear combinations of kernel functions evaluated on the training data points [3, 22, 23].

In the case of complex exponential features, the calculation of the kernel functions can be carried out analytically, and is particularly straightforward in the case of equally spaced frequencies as they appear in the Ising model we discussed in Sec. III D. In fact, in that case we have that

$$\kappa(\boldsymbol{u}, \boldsymbol{u}') = \boldsymbol{\phi}^\dagger(\boldsymbol{u})\boldsymbol{\phi}(\boldsymbol{u}') = \lim_{n\to\infty} \frac{1}{n_\Delta^N} \sum_{\boldsymbol{\omega} \in (\mathcal{S}_n^\Delta)^N} e^{-i\delta t \boldsymbol{\omega}\cdot(\boldsymbol{u}-\boldsymbol{u}')} = \lim_{k_{\max}\to\infty} \left(\frac{1}{2k_{\max}+1}\right)^N \sum_{\boldsymbol{k}=-\boldsymbol{k}_{\max}}^{\boldsymbol{k}_{\max}} e^{-i\frac{\omega_{\max}\delta t}{k_{\max}}\boldsymbol{k}\cdot(\boldsymbol{u}-\boldsymbol{u}')} =$$

$$= \prod_{i=1}^{N} \lim_{k_{\max}\to\infty} \frac{1}{2k_{\max}+1} \left(\sum_{k_i=0}^{k_{\max}} e^{-i\frac{\omega_{\max}\delta t}{k_{\max}} k_i (u_i - u_i')} + c.c. - 1 \right) = \prod_{i=1}^{N} \lim_{k_{\max}\to\infty} \frac{1}{2k_{\max}+1} \left(\frac{e^{-i\omega_{\max}\delta t(u_i-u_i')}}{1 - e^{+i\frac{\omega_{\max}\delta t}{k_{\max}}(u_i-u_i')}} + c.c. \right)$$

$$= \prod_{i=1}^{N} \lim_{k_{\max}\to\infty} \frac{\mathcal{O}(k_{\max})}{2k_{\max}+1} \left(\frac{e^{-i\omega_{\max}\delta t(u_i-u_i')}}{-i\omega_{\max}\delta t(u_i-u_i')} + c.c. \right) = \prod_{i=1}^{N} \frac{\sin\left[\omega_{\max}\delta t(u_i-u_i')\right]}{\omega_{\max}\delta t(u_i-u_i')} = \kappa(\boldsymbol{u}-\boldsymbol{u}'),$$

where we remind that $n_\Delta = \#\mathcal{S}_n^\Delta$ is the number of frequencies in the n-th order Lie-Fourier representation, and the sums over k_i are computed using the well-known formula for the geometric series. The fact that we obtain a (multidimensional) sinc kernel is consistent with the bandwidth-limited nature of the problem. Moreover, we find that for a data set $\mathcal{D}_{\text{train}}$ obtained by sampling an infinite cubic hyperlattice $\boldsymbol{u}_i \in \pi(\omega_{\max}\delta t)^{-1}\mathbb{Z}^N$ this instance of kernel regression reduces to the well known Whittaker-Shannon interpolation formula [44]. In fact,

since the sinc function reduces to a delta function on the hyperlattice points, we have that

$$\kappa(\pi(\omega_{\max}\delta t)^{-1}(\boldsymbol{p}-\boldsymbol{q}))=\delta_{\boldsymbol{p},\boldsymbol{q}}\quad\forall\boldsymbol{p},\boldsymbol{q}\in\mathbb{Z}^N,$$

therefore the kernel is the identity matrix $K_{ij}=\delta_{ij}$ and $\bar{\boldsymbol{a}}=\boldsymbol{J}$ by setting $\lambda_R=0$, which gives us the desired result

$$J_{\boldsymbol{w}(\bar{\boldsymbol{a}})}(\boldsymbol{u})=\sum_{i=1}^{N_{\text{train}}}J_i\kappa(\boldsymbol{u}-\boldsymbol{u}_i)=\sum_{\boldsymbol{p}\in\mathbb{Z}^N}J_i\kappa\left[\boldsymbol{u}-\frac{\pi\boldsymbol{p}}{\omega_{\max}\delta t}\right].$$

According to this kernel representation, the estimated landscape is a linear superposition of sinc kernels with a given wavelength. Moreover, in the case of the Whittaker-Shannon formula, the individual kernels represent the solution for a single sampling point, playing the same role of a Green function for a classical linear field theory. This suggests that, similarly to classical optics, the landscape has a maximum resolution related to the kernel's wavelength, below which details cannot be distinguished. We will make this argument more precise in the following section.

A. Numerical examples

We now benchmark the performance in the regression task of the various representations for the same quantum dynamical landscape we already studied numerically in Sec. III. The system we consider is the transverse field Ising model in Eq. (11), with no longitudinal field $h_z=0$ and the state transfer problem $|0_Q\rangle\mapsto|1_Q\rangle$. We generate a training dataset by randomly sampling a set of controls uniformly inside the hypercube $\boldsymbol{u}_i\in\mathcal{C}^N=[-u_{\max},u_{\max}]$, and then we compute the state transfer fidelity J_i for each one of the sampled controls. Then we train each regression model as described previously, and evaluate the prediction error on a new set data points $\mathcal{D}_{\text{test}}=\{(\boldsymbol{u}_i,J(\boldsymbol{u}_i))\}_{i=1}^{N_{\text{test}}}$, which we define as the root mean squared error:

$$\epsilon_{\text{rms}}=\sqrt{\frac{1}{N_{\text{test}}}\sum_{i=1}^{N_{\text{test}}}(J_i-\sum_{j=1}^{N_{\text{weights}}}w_j\phi_j(\boldsymbol{u}_i))^2}$$

In order to obtain a fair comparison between the three regression models, we need to take into account several aspects. First of all, while the number of features N_{weights} can be chosen freely for the Fourier and Taylor features, it is fixed to $N_{\text{weights}}=N_{\text{train}}$ in the case of kernel regression, as it is an instance-based method. For this reason, we proceed as follows: we split the training data set into a set that we use for the training itself and a validation data set which we use to evaluate the error ϵ_{rms}, repeat the training for a range of values of N_{weights}, and finally fix N_{weights} by choosing the value that gives the lowest validation error. Another important aspect is the choice of the RIDGE regularization parameter λ_R. We empirically found that keeping $N_{\text{weights}}<N_{\text{train}}$ greatly simplifies the analysis in this regard, since then the best results are typically obtained by fixing λ_R as small as it is permitted by numerical stability. The results of this analysis are presented in Fig. 4, where the prediction error ϵ_{rms} is plotted as a function of N_{train} for the various regression models.

The most significant finding is that the sinc kernel features appear to provide the best prediction performance if N_{train} is large enough. Even though its prediction error is larger than for the other feature maps when the training dataset is too small, the plots suggest that there is a threshold value for N_{train} above which the sinc kernel outperforms both the Taylor and Fourier feature maps. This threshold becomes larger as $N, u_{\max}$ increase in value. Meanwhile, for few training samples we see that the Taylor representation generally does best. In the intermediate range, a crossover takes place between Taylor being best for few time step layers, while Fourier works better for more time steps.

We also find qualitatively that the sinc kernel regression is generally more stable and can tolerate smaller values of λ_R compared to the other choices. Finally, we find that the inferior performance of the sinc kernel for low N_{train} can be greatly improved by decreasing the kernel bandwidth to $\omega_{ker}\leq\omega_{max}$, giving rise to

$$\kappa(\boldsymbol{u}-\boldsymbol{u}')=\prod_{i=1}^{N}\frac{\sin\left[\omega_{\text{ker}}\delta t(u_i-u_i')\right]}{\omega_{\text{ker}}\delta t(u_i-u_i')}.$$

The results that we obtain using this strategy are showcased in Fig. 5. There we see that a trade-off between precision on small and large data sets arises, so that ω_{ker} has to be tuned based on practical considerations regarding the problem and the data resources at hand.

B. Analysis of Fourier regression

Let us now go back to Fourier features regression and examine the problem more closely. A first insight that we can gain is that the regression problem for bounded sampled controls $\boldsymbol{u}_i\in\mathcal{C}^N$ does not require finding an exact representation of the landscape J over $\boldsymbol{u}\in\mathbb{R}^N$, for which we would indeed need an infinite number of frequencies. All we need to do is to find an approximation that is good enough inside the hypercube $\mathcal{C}^N$. This has important consequences as far as the Fourier spectrum of the approximations is concerned. In order to understand this point, we reformulate the loss we have chosen (we fix $\lambda_R=0$ for simplicity) as the Monte Carlo sampling of an integral:

$$\sum_{i=1}^{N_{\text{train}}}\frac{|J_{\boldsymbol{w}}(\boldsymbol{u}_i)-J_i|^2}{N_{\text{train}}}\approx\int_{\mathcal{C}^N}\frac{d^N\boldsymbol{u}}{(2u_{\max})^N}|J_{\boldsymbol{w}}(\boldsymbol{u})-J(\boldsymbol{u})|^2$$

where we assume the sampled controls to be drawn from the uniform probability distribution over $\mathcal{C}^N$. As we saw

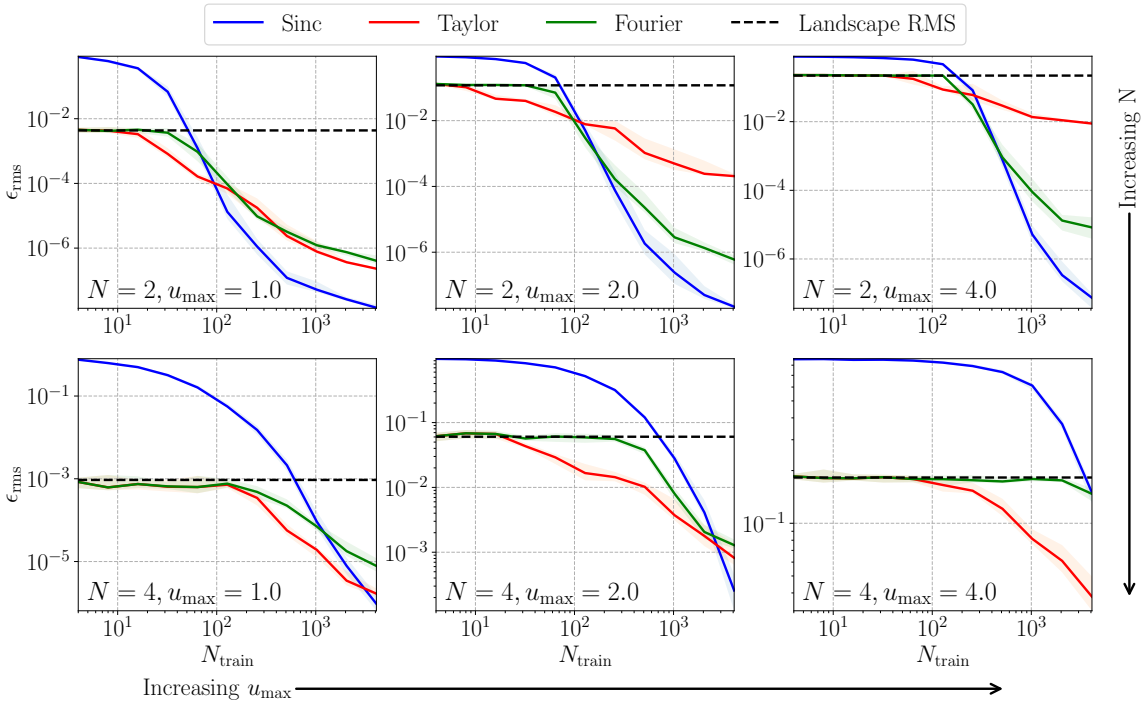

FIG. 4: Prediction performance $\epsilon_{\rm rms}$ of surrogate models as a function of training dataset size $N_{\rm train}$ for the system given by Eq. (11). The results shown in the plot are obtained for $Q = 5$ qubits, time $T = 1.0$, where the colors correspond to the choice of feature map. Each of the six plots corresponds to a value of the parameters $(N, u_{\rm max}) \in \{2, 4\} \times \{1.0, 2.0, 4.0\}$. For Taylor and Fourier features we have $\lambda_R = 10^{-6}$, and only the result for the optimal value of $N_{\rm weights} \leq N_{\rm train}$ is shown, while for the sinc kernel we have $\lambda_R = 10^{-12}$. The training datasets are sampled from a pool of 12672 controls, while the test datasets with $N_{\rm test} = 128$ are sampled from another pool of 3200 and each training/test is repeated 32 times. The solid lines correspond to the median of the prediction errors $\epsilon_{\rm rms}$ over the samples, while the shaded area corresponds to the $25 - 75$ percentile range. The dotted line shows the square root of the variance of the sampled landscape. Compared to the other feature maps, the sinc kernel model typically shows lower values of $\epsilon_{\rm rms}$ for large values of $N_{\rm train}$, while the opposite is true for the Taylor representation.

in Sec. III, J can be approximated with arbitrarily low error over $\mathcal{C}^N$ with its Lie-Fourier representation J_n for a large enough n. We now try to solve the regression problem for the integral loss by choosing the model $J_{\boldsymbol{w}}$ as a sum of Fourier components picked from a set of frequencies $\mathcal{E}$ (that we can suppose to be finite, bounded and satisfying the same symmetries as the Lie-Fourier spectrum $(\mathcal{S}_n^\Delta)^N$). The integral loss can then be written directly in term of the weights as follows:

$$\tilde{\mathfrak{L}}(\boldsymbol{w}, \boldsymbol{w}^\dagger) = \int_{\mathcal{C}^N} \frac{d^N \boldsymbol{u}}{(2u_{\rm max})^N} \left| \sum_{\boldsymbol{\alpha} \in \mathcal{E}} w_{\boldsymbol{\alpha}} e^{-i\delta t \boldsymbol{\alpha}\cdot\boldsymbol{u}} - \sum_{\boldsymbol{\omega} \in (\mathcal{S}_n^\Delta)^N} c_{\boldsymbol{\omega}} e^{-i\delta t \boldsymbol{\omega}\cdot\boldsymbol{u}} \right|^2 =$$

$$\int_{\mathcal{C}^N} \frac{d^N \boldsymbol{u}}{(2u_{\rm max})^N} \left[\sum_{\boldsymbol{\alpha},\boldsymbol{\alpha}' \in \mathcal{E}} w^*_{\boldsymbol{\alpha}'} w_{\boldsymbol{\alpha}} e^{i\delta t(\boldsymbol{\alpha}'-\boldsymbol{\alpha})\cdot\boldsymbol{u}} + \sum_{\boldsymbol{\omega},\boldsymbol{\omega}' \in (\mathcal{S}_n^\Delta)^N} c^*_{\boldsymbol{\omega}'} c_{\boldsymbol{\omega}} e^{i\delta t(\boldsymbol{\omega}'-\boldsymbol{\omega})\cdot\boldsymbol{u}} - \sum_{\boldsymbol{\alpha}\in\mathcal{E},\boldsymbol{\omega}\in(\mathcal{S}_n^\Delta)^N} [w^*_{\boldsymbol{\alpha}} c_{\boldsymbol{\omega}} e^{i\delta t(\boldsymbol{\alpha}-\boldsymbol{\omega})\cdot\boldsymbol{u}} + c^*_{\boldsymbol{\omega}} w_{\boldsymbol{\alpha}} e^{i\delta t(\boldsymbol{\omega}-\boldsymbol{\alpha})\cdot\boldsymbol{u}}] \right]$$

$$= \sum_{\boldsymbol{\alpha},\boldsymbol{\alpha}' \in \mathcal{E}} w^*_{\boldsymbol{\alpha}'} w_{\boldsymbol{\alpha}} \tilde{\kappa}(\boldsymbol{\alpha}' - \boldsymbol{\alpha}) + \sum_{\boldsymbol{\omega},\boldsymbol{\omega}' \in (\mathcal{S}_n^\Delta)^N} c^*_{\boldsymbol{\omega}'} c_{\boldsymbol{\omega}} \tilde{\kappa}(\boldsymbol{\omega}' - \boldsymbol{\omega}) - \sum_{\boldsymbol{\alpha}\in\mathcal{E},\boldsymbol{\omega}\in(\mathcal{S}_n^\Delta)^N} [w^*_{\boldsymbol{\alpha}} c_{\boldsymbol{\omega}} \tilde{\kappa}(\boldsymbol{\alpha} - \boldsymbol{\omega}) + c^*_{\boldsymbol{\omega}} w_{\boldsymbol{\alpha}} \tilde{\kappa}(\boldsymbol{\omega} - \boldsymbol{\alpha})] \quad (16)$$

where we defined the frequency space kernel $\tilde{\kappa}$ as

$$\tilde{\kappa}(\boldsymbol{\omega}) = \prod_{\nu=1}^{N} \frac{\sin \omega_\nu \delta t u_{\rm max}}{\omega_\nu \delta t u_{\rm max}}.$$

1. *Discrete frequencies approximation*

Since $\tilde{\kappa}$ has a finite resolution in Fourier space, the loss is to a first approximation only a function of a local

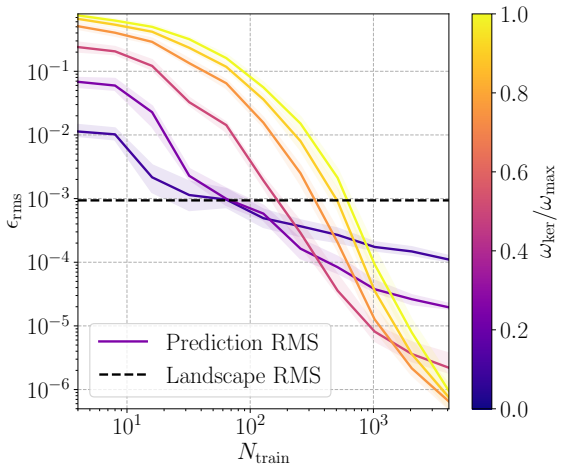

FIG. 5: Sinc kernel regression with reduced kernel bandwidth $\omega_{\rm ker} \leq \omega_{\rm max}$ on the system given by Eq. (11). The results shown in the plot are obtained for $Q = 5, T = 1.0, N = 4, u_{\rm max} = 1.0$ and the different colors (from blue to yellow) corresponding to $\omega_{\rm ker}/\omega_{\rm max} = 0.1, 0.3, 0.5, 0.8, 0.9, 1.0$. The training datasets are sampled from a pool of 12672 points and each training is repeated 32 times. The median (solid lines) and interquartile range (shaded area) of the prediction errors $\epsilon_{\rm rms}$ over the samples are shown. The dotted line shows the square root of the variance of the sampled landscape. Reducing $\omega_{\rm ker}$ considerably reduces prediction error for small training datasets, but also increases it for large datasets.

average of the weights w_α within a certain volume in frequency space. We can study the consequences of this fact by expanding the kernel function for small arguments:

$$\prod_{\nu=1}^{N} \frac{\sin \omega_\nu \delta t u_{\rm max}}{\omega_\nu \delta t u_{\rm max}} = 1 - \frac{(||\boldsymbol{\omega}||_2 \delta t u_{\rm max})^2}{6} + o((||\boldsymbol{\omega}||_2 \delta t u_{\rm max})^2)$$

This means that within a ball in frequency space defined by the Euclidean norm $||\cdot||_2$ and with radius $\Delta\omega$, we can approximate the kernel with a constant up to an error of order $(\Delta\omega\, \delta t\, u_{\rm max})^2$. Since the coefficients relating to frequencies closer than $\Delta\omega$ are then summed up in the loss function, the regression problem effectively depends only on a smaller subset of frequency modes than the original ones. In practice, this suggests that we can find approximate representations with fewer frequencies than the ones in the Lie-Fourier representation, as long as we are just interested in a bounded control region of interest. Estimating how many of these frequencies should be used is a hard problem because it is related to a N-dimensional sphere stacking problem [45].

2. *Flat landscape approximation*

It is perhaps even more interesting to take this reasoning to its extreme logical consequences and consider a regression model containing only the zero frequency mode $w_{\mathbf{0}}$. By substituting the small argument expansion for the kernel into Eq. (16) and neglecting the terms $o((\delta t u_{\rm max}||\boldsymbol{\omega}||_2)^2)$ [46], we obtain the following

$$\tilde{\mathfrak{L}}(w_{\mathbf{0}}, w_{\mathbf{0}}^*) = |w_{\mathbf{0}}|^2 + \sum_{\boldsymbol{\omega},\boldsymbol{\omega}' \in (\mathcal{S}_n^\Delta)^N} c_{\boldsymbol{\omega}'}^* c_{\boldsymbol{\omega}} (1 - \frac{\delta t^2 {u_{\rm max}}^2}{6} ||\boldsymbol{\omega}' - \boldsymbol{\omega}||_2^2)$$
$$- \sum_{\boldsymbol{\omega} \in (\mathcal{S}_n^\Delta)^N} (w_{\mathbf{0}}^* c_{\boldsymbol{\omega}} + c_{\boldsymbol{\omega}}^* w_{\mathbf{0}}) (1 - \frac{\delta t^2 {u_{\rm max}}^2}{6} ||\boldsymbol{\omega}||_2^2).$$

We can find the solution $\bar{w}_{\mathbf{0}}$ to the regression problem by solving the equation $\boldsymbol{\nabla}_{w^\dagger} \mathfrak{L} = \mathbf{0}$. To the zeroth order approximation the solution is given (up to a phase) by

$$\bar{w}_{\mathbf{0}} = \sum_{\boldsymbol{\omega} \in (\mathcal{S}_n^\Delta)^N} c_{\boldsymbol{\omega}} = J_n(\mathbf{0}),$$

as defined by Eq. (8). We can substitute again into the loss function to evaluate how well the constant model approximates the landscape. We obtain:

$$\mathfrak{L}(\bar{w}_{\mathbf{0}}, \bar{w}_{\mathbf{0}}^*) = \frac{\delta t^2 u_{\rm max}^2}{6} \sum_{\boldsymbol{\omega},\boldsymbol{\omega}' \in (\mathcal{S}_n^\Delta)^N} c_{\boldsymbol{\omega}'}^* c_{\boldsymbol{\omega}} (||\boldsymbol{\omega}||_2^2 + ||\boldsymbol{\omega}'||_2^2 - ||\boldsymbol{\omega}' - \boldsymbol{\omega}||_2^2)$$
$$\leq \frac{(\delta t u_{\rm max})^2}{6} \left| \sum_{\boldsymbol{\omega} \in (\mathcal{S}_n^\Delta)^N} c_{\boldsymbol{\omega}} \right|^2 2N\omega_{\rm max}^2 \leq \frac{(u_{\rm max} L)^2}{3N} \xrightarrow{N \to \infty} 0,$$

where in the last step we made use of the boundedness of the landscape $|J_n(\mathbf{0})|^2 \leq 1$.

This results implies a remarkable property of any dynamical landscape with finite time-energy budget L: as the number of controls N increase, the landscape becomes increasingly close to a flat landscape when distance between landscapes is measured using the L_2 norm. As we will see in the next Section, this is related to the appearance of (polynomial) barren plateaux. In relation to landscape learning, this suggests that the sum of square residuals for uniform finite samples is not a well defined loss function, unless N is fixed, which in practice forces us to introduce a cutoff in control pulse discretization (or, equivalently, in frequency) in order to obtain a well defined regression problem. Since numerical experiments necessarily deal with finite values of N, it is possible in practice to use the loss L_2 for regression, but these considerations cast a doubt on the relevance of scaling analysis for this kind of problems.

In order to surpass the difficulties arising from this artificial cutoff, it is most likely to be necessary to employ stronger notions of distance (e.g. using the sup-norm L_∞), non-uniform sampling strategies, or a combination of the two. Since the landscapes we are studying are, after all, objective functions to be optimized, sampling

could also be provided by an optimizer [3, 4]. This way, the focus of the problem would shift from learning the landscape itself to learning the landscape as it "appears" to the optimizer.

VII. LANDSCAPE METRICS

We now discuss the relevance of the results concerning quantum dynamical landscapes that we derived up to now in the context of optimization. Finding the controls that minimize the landscape $J(\boldsymbol{u}) = \langle \hat{O}(\boldsymbol{u}) \rangle$ is the problem we have to solve both in the context of optimal control (for which the landscape often corresponds to the infidelity, so that $\hat{O} = \hat{I} - |\chi\rangle\langle\chi|$) and VQA.

As we have seen in the previous section, the landscape can be represented by means of a kernel with a bandwidth which is set by the time and energy scales of the problem. Therefore, analogously to classical optics, the finite bandwidth of the landscape should prevent us from distinguishing details below a certain length scale l in control space, at least up to a certain tolerance δJ. But then the landscape can be discretized up to an error δJ by sampling it on points standing at a distance l from one another. In order to make these considerations more precise, we can use the formula from Lemma 6. Then, we can see that for two controls $\boldsymbol{u}, \boldsymbol{u'}$ whose infidelity differs by δJ, the following holds true

$$||\boldsymbol{u} - \boldsymbol{u'}||_1 \geq \frac{2\delta J}{\omega_{\max}\delta t} =: l. \tag{17}$$

As anticipated, the two controls have to be separated by a certain minimum distance l (which depends also on δJ), that has to be measured using the taxicab norm $||\cdot||_1$. This simple fact has multiple consequences on the landscape properties.

A. General Quantum Speed Limit

We first analyse the general structure of the Quantum Speed Limit (QSL), which can be defined as the minimum time T_{QSL} needed to perform a control task, such as state transfer to the target state $|\chi\rangle$. We can obtain this from the infidelity landscape assuming the value $J = 0$ at least once within the bounded control region of interest $\mathcal{C}^N = [-u_{\max}, u_{\max}]^N$. Even though we cannot say anything about the QSL for a single target state, we can put a lower bound on the time $T^{\perp}_{QSL}$ needed to reach both the original target state $|\chi\rangle$ and an orthogonal state $|\chi_\perp\rangle, \langle\chi|\chi_\perp\rangle = 0$. Although this definition of the QSL looks rather artificial, if the system exhibits full state controllability it will also satisfy this condition, so that $T^{full}_{QSL} \geq T^{\perp}_{QSL}$. The bound on $T^{\perp}_{QSL}$ (and therefore on T^{full}_{QSL}) can be derived as follows: Since by definition there are controls $\boldsymbol{u}^*, \boldsymbol{u}^*_\perp \in \mathcal{C}^N$ such that $U(\boldsymbol{u}^*)|\psi\rangle = |\chi\rangle$ and $U(\boldsymbol{u}^*_\perp)|\psi\rangle = |\chi_\perp\rangle$, the infidelity landscape will assume there the values $J(\boldsymbol{u}^*) = 1 - |\langle\chi|\chi\rangle|^2 = 0$ and $J(\boldsymbol{u}^*_\perp) = 1 - |\langle\chi|\chi_\perp\rangle|^2 = 1$. But then as we have already seen, Lemma 6 implies that

$$||\boldsymbol{u}^* - \boldsymbol{u}^*_\perp||_1 \geq |J(\boldsymbol{u}^*) - J(\boldsymbol{u}^*_\perp)|K^{-1} \geq \frac{2}{\omega_{\max}\delta t}$$

The two controls inside the hypercube $\mathcal{C}^N$ which are furthest away from each other stand at opposite corners, so that we have

$$||\boldsymbol{u}^* - \boldsymbol{u}^*_\perp||_1 \leq ||\boldsymbol{u}_{\max} - (-\boldsymbol{u}_{\max})||_1 = 2u_{\max}N,$$

which plugged back in the previous equation with $T = N\delta t$ gives the result:

$$T^{\perp}_{QSL} \geq \frac{1}{\omega_{\max}u_{\max}} = \frac{1}{u_{\max}|\lambda_{\max} - \lambda_{\min}|}, \tag{18}$$

where we remind that $\omega_{\max} = |\lambda_{\max} - \lambda_{\min}|$ is the maximum transition angular frequency in the control Hamiltonian ($\hbar = 1$). Note that since $T^{\perp}_{QSL}$ becomes vanishingly small as the bounds on the controls $u_{\max}$ are enlarged, this lower bound for the Quantum Speed Limit is clearly not tight in the rather common cases in which the amplitude of the drift Hamiltonian $\hat{H}_d$ constitutes the bottleneck in time for achieving controllability.

B. Trap separation and trap density

The notion of a minimum distance between distinguishable landscape points can also be used to infer the properties of the landscape around the local minima $\boldsymbol{u}^*$, which we define as global minima inside a neighbourhood

$$\mathcal{M} = \{\boldsymbol{u}^* \in \mathcal{C}^N | \ ||\boldsymbol{u} - \boldsymbol{u}^*||_1 \leq \epsilon \Rightarrow J(\boldsymbol{u}) \geq J(\boldsymbol{u}^*)\} \tag{19}$$

for some given $\epsilon > 0$.

As a visualization, we can picture the quantum dynamical landscape $J(\boldsymbol{u})$ as a topographic landscape with mountains and valleys, filled up with water up until a certain level $\bar{J}$ which is the same across the region we are considering. The water will split up into disconnected water pockets forming a collection of lakes, as pictured in Fig. 6. We can interpret the water level as the current infidelity value and the lakes would be the control regions we have to explore to find better controls. We define the depth of each lake as the difference in height between the water level and the lowest point of the landscape inside the lake (i.e., the local minimum). But then, we can use the result on Lipschitz continuity to relate the extent of the lakes to their depth, i.e., the improvement in infidelity that can be achieved by exploring it. By noticing that the local minima as we defined them are also critical points $\boldsymbol{\nabla} J(\boldsymbol{u}^*) = \boldsymbol{0}$, we can use the second version of the inequality in Lemma 6, that is

$$||\boldsymbol{u} - \boldsymbol{u}^\star||_1 \geq \sqrt{\frac{2\delta J}{(\omega_{\max}\delta t)^2}} = \sqrt{2\delta J}\frac{N}{L}, \tag{20}$$

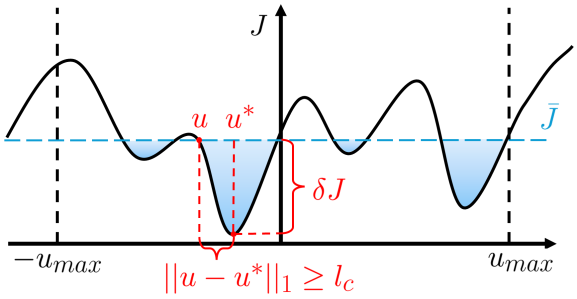

FIG. 6: The quantum dynamical landscape J as a topographical landscape filled with water up to $\bar{J}$. Climbing out of the water from a minimum of depth δJ requires moving in control space over a distance of at least l_c. Even though the size of a lake l_c diverges with the control space dimensionality N, the underwater control volume fraction shrinks to zero (see Sec. VII B).

which is a stronger condition than Eq. (17) when $2J < 1$, and where as before $L = T\omega_{\max}$. We can then define a new control length scale l_c

$$l_c = \beta \frac{N}{L} \tag{21}$$

where $\beta = \max\{\sqrt{2\delta J}, 2\delta J\}$. Then, given a lake of depth δJ, starting from its shore we need to travel a distance (measured using the taxicab norm $||\cdot||_1$) of at least l_c in order to get to the bottom of the lake $\boldsymbol{u}^*$ (the local minimum). In other terms, Eq. (21) implies that the connected neighbourhood of the local minimum where $|J(\boldsymbol{u}) - J(\boldsymbol{u}^*)| < \delta J$ (corresponding to the surface of the lake in our picture) contains a taxicab ball $\mathcal{T}^{(N)}_{l_c}(\boldsymbol{u}^\star)$ of radius l_c centered around $\boldsymbol{u}^\star$:

$$\mathcal{T}^{(N)}_{l_c}(\boldsymbol{u}^\star) = \{\boldsymbol{u} \in \mathbb{R}^N : ||\boldsymbol{u} - \boldsymbol{u}^\star||_1 \leq l_c\}.$$

This means that to every local minimum of depth δJ is associated a volume in control space of at least

$$\mathrm{Vol}\, \mathcal{T}^N_{l_c} = \frac{(2l_c)^N}{N!} = \left(\frac{2\beta}{L}\right)^N \frac{N^N}{N!}.$$

As we cannot fit more than $(\mathrm{Vol}\, \mathcal{T}^N_{l_c})^{-1}$ of these balls in the unit volume, this gives us an upper bound to their average density within the bulk volume $(2u_{max})^N$, this way neglecting surface effects determined by finite bounds on the control region of interest. We can study the large N behaviour of the density bound by means of the Stirling formula:

$$(\mathrm{Vol}\, \mathcal{T}^N_{l_c})^{-1} = \exp\left(-N \log\left(e\frac{2\beta}{L}\right) + \mathcal{O}(\log N)\right), \tag{22}$$

so that the maximum density is exponentially suppressed if $2e\beta > L$ and grows exponentially for $2e\beta < L$ while the marginal case $2e\beta L^{-1} = 1$ requires further inspection of the remainder $\mathcal{O}(\log N)$. On one hand, this suggests a way to bound the number of minima in a given volume but on the other hand, since the radius l_c increases as N is scaled up, Eq. (22) is only reliable for unbounded problems $u_{\max} = \infty$, which is a typical assumption in VQA settings [47].

C. Landscape ruggedness

Another relevant property related to landscape minima is the ruggedness, which instead is a local measure of their sharpness. It can be defined as the average diagonal element of the Hessian evaluated at the minima [3]

$$\rho = \frac{1}{N\#\mathcal{M}} \sum_{\boldsymbol{u}\in\mathcal{M}} \sum_{\nu=1}^{N} \partial^2_\nu J(\boldsymbol{u}),$$

where $\mathcal{M}$ is defined by Eq. (19). We can readily give an upper bound for this quantity by using the derivatives bound from Lemma 5 for $P = 2$:

$$|\rho| \leq \frac{1}{N\#\mathcal{M}} \sum_{\boldsymbol{u}\in\mathcal{M}} \sum_{\nu=1}^{N} |\partial^2_\nu J(\boldsymbol{u})| \leq \frac{(\omega_{\max}\delta t)^2}{2}.$$

It is worth noting that for fixed total time T, ρ vanishes as N increases. Since the Hessian $\boldsymbol{H}$ is the leading contribution to the Taylor expansion of the landscape near a local minimum $\boldsymbol{\nabla} J(\boldsymbol{u}_0) = \boldsymbol{0}$, a large ruggedness implies a low robustness of the solution with respect to small perturbations. Using again Lemma 6, we can upper bound this error. Given a small control deviation $||\boldsymbol{u} - \boldsymbol{u}^\star||_1 \omega_{\max}\delta t \ll 1$, the stronger version of the inequality for critical points gives rise to

$$|J(\boldsymbol{u}) - J(\boldsymbol{u}^\star)| \leq \frac{(\omega_{\max}\delta t)^2}{2} ||\boldsymbol{u} - \boldsymbol{u}^\star||^2_1,$$

which, unlike ρ, allows to quantitatively relate control errors to fidelity variations.

D. Barren plateaux

Shifting the focus away from local minima, we can now consider a metric which is often studied in the context of landscape optimization (especially for VQA applications), namely the variance of the gradient over the controls, defined as follows

$$\mathrm{Var}_{\boldsymbol{u}\in\mathcal{C}^N}[\partial_\nu J(\boldsymbol{u})] = \mathbb{E}_{\boldsymbol{u}\in\mathcal{C}^N}[|\partial_\nu J|^2] - |\mathbb{E}_{\boldsymbol{u}\in\mathcal{C}^N}[\partial_\nu J]|^2$$

where expectation values are integral averages evaluated over the control hypercube $\mathcal{C}^N$ via

$$\mathbb{E}_{\boldsymbol{u}\in\mathcal{C}^N}[g] = \left(\frac{1}{2u_{\max}}\right)^N \int_{\mathcal{C}^N} c^N \boldsymbol{u}\; g(\boldsymbol{u}).$$

An exponential suppression of the latter with respect to some resource (number of qubits, circuit depth, etc.) is customarily referred to as the problem of barren plateaux (or, alternatively, of vanishing gradients). In the literature these averages are often computed over unbounded controls $\boldsymbol{u} \in \mathbb{R}^N$, which here corresponds to taking the limit $u_{\max} \to \infty$. Although this assumption simplifies the derivations, in practice the available controls are always bounded by experimental constraints. For this reason we try here to prove as much as possible for the case of bounded controls.

The Lie-Fourier representation of the landscape allows us to directly relate the variances of any derivative of J_n to the representation coefficients. This way we can prove upper bounds which are also valid for the true landscape J:

Lemma 8 (Variance over bounded controls)**.** *For any integer $n \geq 1$, $P \geq 0$ and $1 \leq \nu_1, \dots, \nu_P \leq N$ we have:*

$$\underset{\boldsymbol{u}\in\mathcal{C}^N}{\mathrm{Var}}\left[\left(\prod_{p=1}^{P}\partial_{\nu_p}\right)J_n\right] = \delta t^{2P}\sum_{\boldsymbol{\omega},\boldsymbol{\omega'}\in(\mathcal{S}_n^{\Delta})^N\setminus\{\mathbf{0}\}} c^*_{\boldsymbol{\omega}}c_{\boldsymbol{\omega'}}\Big(\prod_{p=1}^{P}\omega_{\nu_p}\omega'_{\nu_p}\Big)\big(\tilde{\kappa}(\boldsymbol{\omega}-\boldsymbol{\omega'}) - \tilde{\kappa}(\boldsymbol{\omega})\tilde{\kappa}(\boldsymbol{\omega'})\big), \quad \tilde{\kappa}(\boldsymbol{\omega}) = \prod_{\nu=1}^{N}\frac{\sin\delta t\omega_\nu u_{\max}}{\delta t\omega_\nu u_{\max}}$$

Moreover, the following upper bounds hold:

$$\underset{\boldsymbol{u}\in\mathcal{C}^N}{\mathrm{Var}}\left[\left(\prod_{p=1}^{P}\partial_{\nu_p}\right)J_n, J\right] \leq \min\left[\frac{L^{2P}}{4N^{2P}}, \frac{(L^{P+1}u_{\max})^2}{3N^{2P+1}} + o\left(\frac{L^{2P+2}}{N^{2P+1}}\right)\right].$$

Proof. See App. C for a proof valid in the case of multiple controls and a generic observable $\hat{O}$. □

It is interesting to note that, similarly to what was already observed in Sec. VI B, the variance over the hypercube $\mathcal{C}^N$ could be expressed by means of a quadratic form of the Lie-Fourier coefficients featuring a modified version of the sinc kernel $\tilde{\kappa}(\boldsymbol{\omega})$ where the bandwidth $\delta t u_{\max}$ is now set by the size of the integration region. Since the formula holds in general for any set of frequencies and coefficients, it also proves that this quadratic form is semi-positive definite.

The upper bound in Lemma 8 shows that the variance for bounded controls increases with the time-energy budget L and it decreases with the number of controls or circuit layers N, the dependence being polynomial in both cases, while in the same limit being exponentially suppressed with the order of the derivatives P. Since these upper bounds are independent on the details of the spectrum (other than $\omega_{\max}$), they can be understood as constraints on the best-case scenario for barren plateaux across quantum systems as set by time, depth and bandwidth limitation for their Lie-Fourier expansion.

The upper bound for the variance implies that even for $P = 0$ (which corresponds to not taking any derivatives) this quantity goes to zero for large N if $L = T\omega_{\max}$ is kept constant. As a consequence, the landscape becomes flatter and the control region $\mathcal{R}_{\overline{\delta J}}(N)$ where J is smaller than its average by more than a finite amount $\overline{\delta J} > 0$, which we define as

$$\mathcal{R}_{\overline{\delta J}}(N) = \{\boldsymbol{u} \in \mathcal{C}^N \text{ s.t. } \mathbb{E}_{\boldsymbol{u}\in\mathcal{C}^N}[J] - J(\boldsymbol{u}) > \overline{\delta J}\},$$

shrinks to zero in volume compared to the control hypercube. In fact we have that:

$$\begin{aligned}\underset{\boldsymbol{u}\in\mathcal{C}^N}{\mathrm{Var}}\, J &= \int_{\mathcal{C}^N}\frac{d^N\boldsymbol{u}}{\mathrm{Vol}\,\mathcal{C}^N}|J(\boldsymbol{u}) - \mathbb{E}_{\boldsymbol{u}\in\mathcal{C}^N}[J]|^2 \geq \\ &\int_{\mathcal{C}^N}\frac{d^N\boldsymbol{u}}{\mathrm{Vol}\,\mathcal{C}^N}|J(\boldsymbol{u}) - \mathbb{E}_{\boldsymbol{u}\in\mathcal{C}^N}[J]|^2\Theta(\mathbb{E}_{\boldsymbol{u}\in\mathcal{C}^N}[J] - J(\boldsymbol{u}) - \overline{\delta J}) \\ &\geq \frac{(\overline{\delta J})^2}{\mathrm{Vol}\,\mathcal{C}^N}\int_{\mathcal{C}^N} d^N\boldsymbol{u}\;\Theta(\mathbb{E}_{\boldsymbol{u}\in\mathcal{C}^N}[J] - J(\boldsymbol{u}) - \overline{\delta J}) \\ &= (\overline{\delta J})^2\frac{\mathrm{Vol}\,\mathcal{R}_{\overline{\delta J}}(N)}{\mathrm{Vol}\,\mathcal{C}^N}.\end{aligned}$$

But then by making use of the asymptotic upper bound from Lemma 8 for $P = 0$, we conclude that

$$\frac{\mathrm{Vol}\,\mathcal{R}_{\overline{\delta J}}(N)}{\mathrm{Vol}\,\mathcal{C}^N} \leq \mathcal{O}\left(\frac{1}{N}\right) \xrightarrow{N\to\infty} 0.$$

This fact is closely related to the discussion in Sec. VI B and also clarifies what happens to the volume surrounding the minima that we discussed previously (the overall lake surface). In fact, if we choose the water level to be below the landscape average $\bar{J} < \mathbb{E}_{\boldsymbol{u}\in\mathcal{C}^N}[J]$ and fix $\overline{\delta J} = \mathbb{E}_{\boldsymbol{u}\in\mathcal{C}^N}[J] - \bar{J}$, then $\mathcal{R}_{\overline{\delta J}}(N)$ is the part of the control landscape underwater. The fact that the volume fraction associated to this control region shrinks to zero as $N \to \infty$ could appear somewhat surprising, given that the minimum distance $l_c \propto N$ we have to travel to get from shore to bottom diverges. But the two facts are not necessarily mutually exclusive, as the balls $\mathcal{T}^N_{l_c}$ will eventually not completely fit inside $\mathcal{C}^N$ as N increases, and the changes in volume determined by the growth in dimensionality are hard to picture intuitively.

Unfortunately we cannot extend all the results for the true landscape J from Lemma 8 to unbounded controls

by just taking the limit $u_{\max} \to \infty$, because the limits in the Lie product expansion order n and control region size $u_{\max}$ cannot be exchanged. Instead, if we work with a finite n, i.e. in the typical VQA/PQC setting, where the circuit has finite depth and therefore the Fourier representation is finite and exact, then that is possible. As noticed before, the $n = 1$ case correspond physically to an interleaved circuit, and the frequencies in the Fourier representation $\mathcal{S}_1^\Delta = \mathcal{S}^\Delta$ correspond only to the differences in control Hamiltonian eigenvalues. Keeping this distinction in mind, we can prove similar results for unbounded controls.

Lemma 9 (Variance over unbounded controls)**.** *For any integer $n \geq 1$, $P \geq 0$ and $1 \leq \nu_1, \dots, \nu_P \leq N$ we have:*

$$\underset{\boldsymbol{u}\in\mathbb{R}^N}{\mathrm{Var}}\left[\left(\prod_{p=1}^{P}\partial_{\nu_p}\right)J_n\right] = \delta t^{2P}\sum_{\boldsymbol{\omega}\in(\mathcal{S}_n^\Delta)^N\setminus\{\mathbf{0}\}}|c_{\boldsymbol{\omega}}|^2\prod_{p=1}^{P}\omega_{\nu_p}^2$$

Moreover, the following bounds hold:

$$\underset{\boldsymbol{u}\in\mathbb{R}^N}{\mathrm{Var}}\left[\left(\prod_{p=1}^{P}\partial_{\nu_p}\right)J_n, J\right] \leq \left(\frac{(\delta t\omega_{\max})^P}{2}\right)^2,$$

$$\sum_{\nu_1,\dots,\nu_P}\underset{\boldsymbol{u}\in\mathbb{R}^N}{\mathrm{Var}}\left[\left(\prod_{p=1}^{P}\partial_{\nu_p}\right)J_n\right] \geq \frac{\Delta J^2\delta t^{2P}}{4\sum_{\boldsymbol{\omega}\in(\mathcal{S}_n^\Delta)^N\setminus\{\mathbf{0}\}}\frac{1}{||\boldsymbol{\omega}||_2^{2P}}}$$

where we defined ΔJ as the maximum variation in J_n,

$$\Delta J := \sup_{\boldsymbol{u}\in\mathbb{R}^N} J_n(\boldsymbol{u}) - \inf_{\boldsymbol{u}\in\mathbb{R}^N} J_n(\boldsymbol{u}).$$

Proof. See App. C for a proof in the case of multiple controls and a generic observable $\hat{O}$. □

While the upper bound holds for both the Lie-Fourier representations J_n and the landscape J itself, the lower bound only holds for finite n. The latter can be derived thanks to the simpler structure of the kernel in the case of unbounded controls, but we do not rule out that it might be possible to derive an equivalent result for bounded controls. More specifically, the case $P = 1$ (i.e. the gradient variance) gives us a worst-case scenario estimate for the barren plateaux even across different systems and different initial and target states (or observables). We can derive a simplified, looser bound in terms of the smallest non-zero frequency $\omega_{\min} \in \mathcal{S}_n^\Delta$, which assures $\forall\boldsymbol{\omega} \in (\mathcal{S}_n^\Delta)^N\setminus\{\mathbf{0}\}$, $||\boldsymbol{\omega}||_2^2 \geq \omega_{\min}^2$, so that

$$\begin{aligned}\sum_{\nu=1}^{N}\underset{\boldsymbol{u}\in\mathbb{R}^N}{\mathrm{Var}}\left[\partial_\nu J_n\right] &\geq \frac{\Delta J^2\delta t^2}{4\sum_{\boldsymbol{\omega}\in(\mathcal{S}_n^\Delta)^N\setminus\{\mathbf{0}\}}\frac{1}{||\boldsymbol{\omega}||_2^2}}\\ &\geq \frac{\Delta J^2\delta t^2\omega_{\min}^2}{4((\#\mathcal{S}_n^\Delta)^N-1)},\end{aligned}$$

which recovers the exponential suppression in circuit depth N that is typical in the case of barren plateaux. In the case $n = 1$ of an interleaved circuit, $\mathcal{S}_1^\Delta = \mathcal{S}^\Delta$ can be of order $\mathcal{O}(D^2 = 2^{2Q})$ if $\hat{H}_c$ is non-degenerate, but as we saw in Sec. III D it can be much smaller in physically relevant systems thanks to degeneracies. In the case of the Q-qubits Ising model (with or without longitudinal field), we have $\mathcal{S}^\Delta \sim \mathcal{O}(Q^2)$, so that even in the worst case scenario the gradient variance is only exponentially suppressed in the circuit depth N, but not in the number of qubits Q. This is in contrast to the case of a non-degenerate $\hat{H}_c$, where this analysis does not rule out the existence of fixed depth barren plateaux when Q increases. As a conclusive remark, we note that the dependence of gradient variance suppression on the growth of the dynamical Lie algebra [7] implies that the latter must play a decisive role in the distribution of the Fourier coefficients, deciding how far away the system will be from this bound.

VIII. APPLICATIONS TO OPTIMIZER DESIGN

The considerations we made in the previous Section are also relevant for the design of optimization algorithms. First of all, the length scale l we derived from Lipschitz continuity constrains how local the search needs to be.

If we want to obtain an improvement of at least δJ to our current best control $\boldsymbol{u}$, Eq. (17) assures us that such an improvement cannot be found within distance l. This is especially relevant to in-situ optimization, such as black-box optimizers like DCRAB [39, 48] and SOMA [49]. For example, taking into account shot noise or other uncertainties, it may be meaningless to try to pick points closer than l. For instances of DCRAB based on a simplex search, the length scale can specifically be used to constrain the size of the simplex, for example at the start of the optimization. Alternatively, given the finite resolution of arbitrary waveform generators, one can constrain the minimum vertical resolution so as to attain a given change in the cost function landscape.

Going further still, one can consider a general parameterization of the controls given by a linear transformation of the piece-wise constant basis as prescribed by Ref. [24], such that

$$u(t) = \sum_{i=1}^{N_c} r_i(t)v_i.$$

As we have seen in Corollary 1, the derivative properties that we proved for the landscape $J(\boldsymbol{u})$ are still true for the parametrized landscape $\tilde{J}(\boldsymbol{v})$ provided that the basis functions are properly normalized (which we are always free to achieve). Hence, the notion of the length scale l also applies to the new control space.

The convergence rate of the optimizers can also be affected by characterizing landscape properties, especially for model-based control. Gradient optimizers [10, 50] in particular will have to estimate step sizes which can be aided by notions of minimum distance as above, as well as bounds on first and second derivatives given by Lemma 5.

Similarly quantum Newton search [51] and quasi-Newton search [10, 43] can also benefit from bounds on the Hessian. Likewise, the minimum length scale can be helpful to inform stopping conditions, especially in the presence of shot noise or decoherence.

The nature of the landscape is also especially important for sampling based methods such as those using machine learning, e.g. the aforementioned meta-learning optimizer SOMA [49], since it can help to set a minimum distance between search points. This minimum distance is especially relevant to discrete point sampling, e.g. within a Monte-Carlo tree search, as used in global reinforcement learning [52], or in grid search methods such as DIRECT [53]. In fact, the Lipschitz continuity of the landscape calls for a wider adoption of global optimizers that are specific to this class of functions, improving convergence with respect to more naive approaches [53–56].

As an illustrative example, we can apply the upper bound for the Lipschitz constant to the DIRECT algorithm [53]. Appropriate bounds for extensions of this algorithm [22, 55, 56] can be derived in a similar way. Since DIRECT samples J on a rectangular grid, we can assume $\boldsymbol{u'} - \boldsymbol{u} = h\boldsymbol{e}_\nu$ for some canonical basis vector $\boldsymbol{e}_\nu$ and $h > 0$ [57]. But then we have the following result

$$|J(\boldsymbol{u}) - J(\boldsymbol{u'})| \leq \frac{\omega_{\max}\delta t}{2}||\boldsymbol{u'} - \boldsymbol{u}||_1 \leq \frac{\omega_{\max}\delta t h}{2}.$$

This means that we can use the upper bound $K \leq \omega_{\max}\delta t/2$ within DIRECT as an additional condition during the selection of potentially optimal hyperrectangles to decrease the number of function evaluations.

IX. CONCLUSIONS

In this paper we derived from first principles and under very general assumptions the main properties of three different feature map representations of a quantum dynamical landscape. The latter is the expectation value of an observable over the output state of a generalized Parametrized Quantum Circuit, which is equivalent to a controlled quantum system with stepwise-constant controls, and generalizes Variational Quantum Algorithm circuits. The feature map representations are approximations of the landscape, which we use to study the properties of the landscape itself, and as physics-informed models for supervised learning, with the aim of informing the development of methods for quantum cost function optimization.

First, we obtained a Fourier representation [25] by means of a Lie-Trotter approximation of the dynamics. We showed that the resulting frequency spectrum fills up densely a finite hypercube, with size given by the maximum transition frequency in the control Hamiltonian. We proved analytically some important properties of the representation coefficients which are related to boundedness and discrete symmetries, which are useful for the further development of simulation and learning algorithms [20]. We found numerically in the case of the Ising Model that owing to the high degree of symmetry of the Hamiltonian and initial and target states, the spectrum is a stepwise-continuous function, opening up the possibility of further compression by using e.g. polynomial bases in frequency space. These numerical results were obtained thanks to an algorithm that can be applied to any model with equally spaced control eigenvalues. This allows further landscape exploration of Pauli controlled systems, which are a common Ansatz for ease of experimental implementation and theoretical investigation [16, 20, 22, 58].

We then showed that the bounds in absolute value and bandwidth of the landscape cause it to be a Lipschitz continuous function. This means that there is a global maximum ratio between change in cost function and traveled length in control space. We related this property to a local upper bound on the error of a Taylor expansion, which is therefore an efficient representation when the overall time-energy budget is limited. Local models are often useful for optimization, and this result can inform further study in this regard [16, 43].

Since the dense Fourier spectrum represents a challenge for supervised learning with a finite dimensional feature space, we derived analytically an equivalent kernel regression problem that gives rise to the sinc kernel representation. Here, the landscape is represented as a linear superposition of finite bandwidth kernels in a way that generalizes the Green's function expansion from linear optics. In numerical benchmarks against the other two representations, the sinc kernel showed to be more efficient if the training data set is large enough, while its inferior performance on small data sets can be greatly improved by reducing the kernel bandwidth. We also commented on the limits of random sampling and of the L_2 distance to distinguish different landscapes, arguing that stronger notions of distance, such as L_∞, or optimization-driven sampling strategies [3, 4, 52] are necessary to obtain a meaningful regression problem when the number of time steps N is not fixed or is too large.

We then discussed the consequences of our findings on some landscape metrics which are relevant for optimization. In particular, we showed that the aforementioned Lipschitz constant implies a minimum granted robustness of local optima, together with a relation between the volume of an unexplored region in control space and the improvement in the cost function value we can obtain by exploring that region. We also related the Fourier representation to the variance of the landscape and its derivatives over the controls. This allowed us to prove a set of upper bounds that constrain the best case scenario for this quantities, purely based on time, energy and depth limitations and largely independently on the specific Hamiltonian. This way, we showed that for constant final time T and bounded controls, the non-trivial regions of the landscape have a vanishingly small vol-

ume compared to a fixed control region of interest as the number of control parameters increases. Overall, owing to the generality of the properties and the bounds we found, we deem them to be useful as universal baselines for the considered metrics in more specific cases.

Finally, we explored how these landscape properties inform the design and tuning of optimizers. In this sense, the generalization of our proofs to parametrized landscapes is instrumental, allowing for the application to popular Quantum Optimal Control algorithms such as GRAPE [10, 50], DCRAB [48], Krotov [59], SOMA [49], and SPINACH [43], where our results inform the choice of stopping conditions and the estimate of convergence ratios. They also provide estimates for appropriate stepping sizes, bounds on the Hessian, and inform the choice of vertical resolution in the controls, which can all greatly benefit hyperparameter choices in such algorithms. We also argued in favour of Lipschitz-aware optimizers such as DIRECT [22, 53], for which we derive a more specific Lipschitz constant. The information about distance between sampled points is especially relevant to global optimizers and sampling based strategies such as machine learning, for example dictating the branching in Monte-Carlo tree searches [52]. That is, this can be used to decrease the number of calls to the quantum circuit by discarding sampling in areas which cannot contain better quantum cost function values because of Lipschitz continuity.

ACKNOWLEDGMENTS

We thank Marin Bukov, Nicolò Beato and Mogens Dalgaard for all the insightful conversations, and especially Mogens Dalgaard for sharing his code for the simulation and optimal control of spin chains. We also thank Jan Reuter, Phila Rembold, Robert Zeier, Matthias Müller, Matteo Rizzi, Markus Schmitt, Michael Schilling and Francesco Preti for related discussions. This work was funded by Horizon Europe programme HORIZON-CL4-2022-QUANTUM-02-SGA via the project 101113690 (PASQuanS2.1), programme HORIZON-CL4-2022-QUANTUM-01-SGA via project 101113946 (OpenSuperQPlus100), and by Horizon Europe programme (HORIZON-CL4-2021-DIGITALEMERGING-02-10) Grant Agreement 101080085 QCFD, and it was supported from the Jülich Supercomputing Center through the JUWELS and JURECA clusters.

Appendix A: Symbols, notation and useful formulas

a. *Sets and vectors*

We indicate sets with calligraphic capital letters like $\mathcal{H}, \mathcal{F}, \mathcal{D}$. Given a set $\mathcal{F}$, we use the exponential notation $\mathcal{F}^N$ with $N \in \mathbb{N}$ for the N-fold cartesian product of the set with itself and the symbol $\#\mathcal{F}$ for its cardinality.

Throughout the paper we often switch between vector and coordinate notation. We use boldface lower case letters for (column) vectors $\boldsymbol{v} \in \mathbb{C}^N$, always implying that $\boldsymbol{v} = (v_1, \dots, v_N)^T$, and boldface upper case letters for matrices $\boldsymbol{A} \in \mathbb{C}^{M\times N}$, with

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & \dots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{M1} & \dots & a_{MN} \end{pmatrix} = \begin{pmatrix} \boldsymbol{a}_1 & \dots & \boldsymbol{a}_N \end{pmatrix}$$

We use vectorized shorthand notations for fixed scalar quantities, e.g.:

$$\mathbf{1} = (1, \dots, 1)^T, \quad \boldsymbol{\omega}_{\max} = (\omega_{\max}, \dots, \omega_{\max})^T.$$

When dealing with sums over a vector index $\boldsymbol{k} \in \mathbb{Z}^N$, where each element $k_i \in \mathbb{Z}$ of the vector spans the integers between $a_i \le k_i \le b_i$ we sometimes use the following shorthand notation:

$$\sum_{\boldsymbol{k}=\boldsymbol{a}}^{\boldsymbol{b}} := \sum_{k_1=a_1}^{b_1} \cdots \sum_{k_N=a_N}^{b_N} .$$

b. *Norms and balls*

We use the following notation for norms of vectors in $\mathbb{C}^N$:

$$||\boldsymbol{v}||_1 = \sum_{i=1}^{N} |v_i|$$

$$||\boldsymbol{v}||_2 = \sqrt{\sum_{i=1}^{N} |v_i|^2}$$

$$||\boldsymbol{v}||_\infty = \max_{i=1,\dots,N} |v_i|$$

and equivalent definitions hold for matrices. We indicate with $\{\boldsymbol{e}_i\}_{i=1,\dots,N}$ the canonical basis. We call $\mathcal{C}_A^{(N)}(\boldsymbol{s}_0)$ the real ball defined by the sup-norm $||\cdot||_\infty$ of radius A centered around $\boldsymbol{s}_0$ (i.e. the translated hypercube):

$$\mathcal{C}_A^{(N)}(\boldsymbol{s}_0) = \{\boldsymbol{s} \in \mathbb{R}^N : ||\boldsymbol{s} - \boldsymbol{s}_0||_\infty \le A\}$$

and $\mathcal{T}_A^{(N)}(\boldsymbol{s}_0)$ the real ball defined by the L_1 (also known as "taxicab") norm $||\cdot||_1$ of radius A centered around $\boldsymbol{s}_0$:

$$\mathcal{T}_A^{(N)}(\boldsymbol{s}_0) = \{\boldsymbol{s} \in \mathbb{R}^N : ||\boldsymbol{s} - \boldsymbol{s}_0||_1 \le A\}.$$

In both cases, we drop the $\boldsymbol{s}_0$ argument when considering a ball centered at the origin $\boldsymbol{s}_0 = \mathbf{0}$.

We use the braket notation for quantum states $|\psi\rangle, |\chi\rangle$ and operators $\hat{U}$. Since we work with finite dimensional systems with Hilbert space $\mathcal{H} \simeq \mathbb{C}^D$, we can choose a finite orthonormal basis $\{\boldsymbol{\beta}_i\}_{i=1,\dots,D}$ with $\langle\beta_i|\beta_j\rangle = \delta_{ij}$,

$\sum_{i=1}^{d} |\beta_i\rangle \langle\beta_i| = \mathrm{id}(\mathcal{H}) =: \hat{I}$ and represent quantum operators as matrices with (capital letter) entries U_{ij}

$$U_{ij} = \langle\beta_i| \hat{U} |\beta_j\rangle , \ \hat{U} = \sum_{ij} U_{ij} |\beta_i\rangle \langle\beta_j|$$

and quantum states as vectors with entries ψ_i

$$\psi_i = \langle\beta_i|\psi\rangle , \ |\psi\rangle = \sum_{i=1}^{D} \psi_i |\beta_i\rangle .$$

When dealing with product Hilbert spaces $\mathcal{H}^{\otimes Q}$, given $|\psi\rangle \in \mathcal{H}$ we use the notation

$$|\psi_Q\rangle := |\psi\rangle^{\otimes Q} .$$

We indicate with

$$||\hat{O}||_\infty = \max_{|\psi\rangle \in \mathcal{H}\setminus\{|0\rangle\}} \sqrt{\frac{\langle \hat{O}\psi|\hat{O}\psi\rangle}{\langle\psi|\psi\rangle}}$$

the sup-norm of the operator $\hat{O}$. The sup-norm is subadditive and sub-multiplicative, and coincides with the absolute value of the largest singular value of O_{ij}. We use the convention $\hbar = 1$ throughout the text.

c. *Linear expansion of exponentials*

Since we are going to use this result inside several proofs, we show here a formula concerning the linear expansion of the exponential. Given a (finite dimensional) operator $\hat{X} \in \mathbb{C}^{D\times D}$, we have

$$||e^{\hat{X}} - \hat{I}||_\infty = ||\sum_{n=1}^{\infty} \frac{\hat{X}^n}{n!}||_\infty \leq \sum_{n=1}^{\infty} \frac{||\hat{X}||_\infty^n}{n!} = e^{||\hat{X}||_\infty} - 1,$$

$$||e^{\hat{X}} - \hat{I} - \hat{X}||_\infty = ||\sum_{n=2}^{\infty} \frac{\hat{X}^n}{n!}||_\infty \leq$$

$$\leq \sum_{n=2}^{\infty} \frac{||\hat{X}||_\infty^n}{n!} = e^{||\hat{X}||_\infty} - 1 - ||\hat{X}||_\infty,$$

where we made use of the subadditivity and submultiplicativity of the sup-norm $||\cdot||_\infty$. Because of the mean value theorem, we have for $x \geq 0 \in \mathbb{R}, \exists y \in [0, x]$

$$e^x = 1 + xe^y$$

which implies

$$e^x - 1 - x = x(e^y - 1).$$

By applying again the same theorem we have $\exists z \in [0, y]$ such that

$$e^x - 1 - x = x(ye^z) \leq x^2 e^x.$$

Using these inequalities we can conclude:

$$||e^{\hat{X}} - \hat{I}||_\infty \leq e^{||\hat{X}||_\infty} - 1 \leq ||\hat{X}||_\infty e^{||\hat{X}||_\infty}, \qquad \text{(A1)}$$

$$||e^{\hat{X}} - \hat{I} - \hat{X}||_\infty \leq e^{||\hat{X}||_\infty} - 1 - ||\hat{X}||_\infty$$

$$\leq ||\hat{X}||_\infty^2 e^{||\hat{X}||_\infty}. \qquad \text{(A2)}$$

Appendix B: Reduction to standard form

We consider the time evolution of a quantum system of finite dimension D under the time dependent Hamiltonian

$$\hat{H}(t) = \hat{H}^{(0)} + \sum_{\mu=1}^{M} \hat{H}^{(\mu)} u_\mu(t)$$

where the controls $u_\mu(t)$ are bound to an interval $u_\mu \in [u_\mu^{\min}, u_\mu^{\max}]$. In general, only M' out of M control Hamiltonians $\hat{H}^{(\mu)}$ are linearly independent (let us say they are $\mu = 1, \ldots, M'$). Then we can write the remaining $M - M'$ terms as a linear combination of the first M' terms

$$\hat{H}(t) = \hat{H}^{(0)} + \sum_{\mu'=1}^{M'} \hat{H}^{(\mu')} u_{\mu'}(t) + \sum_{\mu=M'+1}^{M} \hat{H}^{(\mu)} u_\mu(t)$$

$$= \hat{H}^{(0)} + \sum_{\mu'=1}^{M'} \hat{H}^{(\mu')} u_{\mu'}(t) + \sum_{\mu=M'+1}^{M} \sum_{\mu'=1}^{M'} a_\mu^{(\mu')} \hat{H}^{(\mu')} u_\mu(t) =$$

$$= \hat{H}^{(0)} + \sum_{\mu'=1}^{M'} \underbrace{\left(u_{\mu'}(t) + \sum_{\mu=M'+1}^{M} a_\mu^{(\mu')} u_\mu(t) \right)}_{u'_{\mu'}(t)} \hat{H}^{(\mu')},$$

where the linearly independent controls $u'_\mu(t)$ are bound to another interval $u'_\mu \in [u_\mu^{'min}, u_\mu^{'max}]$. Since $\hat{H}(t)$ only depends on $u'_\mu(t)$, we can suppose without loss of generality that the control Hamiltonians are linearly independent from the start and drop the apex notation.

We will now show that by a further reparametrization we can also fix the energy scales across different controls to a common value $\omega_{\max}$ and (optionally) center the control bounds around the origin. Let us define the maximum transition frequency $\omega_{\max}^{(\mu)}$ in the control Hamiltonian $\hat{H}^{(\mu)}$ as

$$\omega_{\max}^{(\mu)} = |\lambda_{\max}^{(\mu)} - \lambda_{\min}^{(\mu)}|$$

where $\lambda_{\max}^{(\mu)}(\lambda_{\min}^{(\mu)})$ is the maximum (minimum) eigenvalue of $\hat{H}^{(\mu)}$. We define new controls $\tilde{u}_\mu(t)$ as

$$u_\mu(t) = m_\mu + \frac{\omega_{\max}}{\omega_{\max}^{(\mu)}} \tilde{u}_\mu(t), \quad m_\mu = \frac{u_\mu^{\min} + u_\mu^{\max}}{2}$$

Then we can rewrite the Hamiltonian as

$$\hat{H}(t) = \hat{H}^{(0)} + \sum_{\mu=1}^{M} \hat{H}^{(\mu)} u_\mu(t) =$$

$$\hat{H}^{(0)} + \sum_{\mu=1}^{M} \hat{H}^{(\mu)} m_\mu + \sum_{\mu=1}^{M} \hat{H}^{(\mu)} \frac{\omega_{\max}}{\omega_{\max}^{(\mu)}} \tilde{u}_\mu(t) =$$

$$\hat{\tilde{H}}^{(0)} + \sum_{\mu=1}^{M} \hat{\tilde{H}}^{(\mu)} \tilde{u}_\mu(t)$$

so that now we have $\forall \mu\ \tilde{\omega}^{(\mu)}_{\max} = \omega_{\max}$ and the rescaled controls $\tilde{u}_\mu(t)$ are bound to the symmetric interval $[-\widetilde{\Delta u}_\mu/2, \widetilde{\Delta u}_\mu/2]$ given by

$$\widetilde{\Delta u}_\mu = \frac{\omega^{(\mu)}_{\max}}{\omega_{\max}}(u^{\max}_\mu - u^{\min}_\mu)$$

Although the price to pay lies in modifying the drift Hamiltonian $\hat{H}^{(0)}$, we notice that loosening the bounds symmetrically does not change m_μ nor the drift term:

$$u^{\min}_\mu \mapsto u^{\min}_\mu - \delta, \quad u^{\max}_\mu \mapsto u^{\max}_\mu + \delta,$$
$$\widetilde{\Delta u}_\mu \mapsto \widetilde{\Delta u}_\mu + \frac{\omega^{(\mu)}_{\max}}{\omega_{\max}}\delta, \quad m_\mu \mapsto m_\mu.$$

We can then study the problem as the control bounds are changed for fixed drift by simply changing $\widetilde{\Delta u}_\mu$. Finally, we notice that adding a term δI proportional to the identity to the Hamiltonian only changes the overall phase of the time evolution, and therefore does not affect observation values, hence the landscape. But then we can use this freedom to shift the controls $\hat{H}^{(\mu)}$ so that $|\lambda^{(\mu)}_{\max}| = ||\hat{H}^{(\mu)}||_\infty = \omega_{\max}/2$.

Appendix C: Multiple controls

We now generalize the Lie-Fourier representation that was discussed in the main text to the case of multiple controls. More specifically, we consider the following Hamiltonian

$$\hat{H}(t) = \hat{H}^{(0)} + \sum_{\mu=1}^{M} \hat{H}^{(\mu)} u_\mu(t),$$

where we assume without loss of generality that $u_\mu(t) \in [-u^{(\mu)}_{\max}, u^{(\mu)}_{\max}]$ and $\omega^{(\mu)}_{\max} = \omega_{\max}$ (see Appendix B for details), while we keep the same time discretization convention as in the single control case. We will call $u_{\max} = \max_\mu u^{(\mu)}_{\max}$. When discretizing the control pulses $u_{\mu\nu} := u_\mu(t_\nu)$, we use the following matrix notation:

$$\boldsymbol{U} = \begin{pmatrix} u_{11} & \dots & u_{1N} \\ \vdots & \ddots & \vdots \\ u_{M1} & \dots & u_{MN} \end{pmatrix} = \begin{pmatrix} \boldsymbol{u}_1 & \dots & \boldsymbol{u}_N \end{pmatrix}$$

where now the control vector $\boldsymbol{u}_\nu \in \mathbb{R}^M$ represents the value of the M controls at the ν-th timestep. Correspondingly, the control region of interest for the discretized pulse is

$$\boldsymbol{U} \in \mathcal{C} := \prod_{\mu,\nu=1}^{M,N} [-u^{(\mu)}_{\max}, u^{(\mu)}_{\max}], \ \mathrm{Vol}(\mathcal{C}) = \prod_{\mu,\nu=1}^{M,N} 2u^{(\mu)}_{\max}$$

We then start by expanding the timestep unitary $\hat{U}(\boldsymbol{u})$, defined by

$$\hat{U}(\boldsymbol{u}) = e^{-i\delta t(H^{(0)} + \sum_{\mu=1}^{M} u_\mu H^{(\mu)})}$$

as a sum of complex exponentials by means of the Lie-Trotter product formula Eq. (4). In order to achieve that, we express the control Hamiltonians in their eigenbasis $\hat{H}^{(\mu)} = \hat{V}^{(\mu)\dagger}\hat{\Lambda}^{(\mu)}\hat{V}^{(\mu)}$, and absorb the change of basis in the terms $\hat{W}$ which do not depend on the controls

$$\begin{aligned}\hat{U}_n(\boldsymbol{u}) &= (e^{-\frac{i\delta t}{n}\hat{H}^{(0)}} e^{-\frac{i\delta t}{n}u_1\hat{H}^{(1)}} \cdots e^{-\frac{i\delta t}{n}u_M\hat{H}^{(M)}})^n \\ &= \hat{V}^{(M)\dagger} \underbrace{\hat{V}^{(M)} e^{-\frac{i\delta t}{n}\hat{H}^{(0)}} \hat{V}^{(1)\dagger}}_{\hat{W}^{(1)}(n^{-1}\delta t)} e^{-\frac{i\delta t}{n}u_1\hat{\Lambda}^{(1)}} \underbrace{\hat{V}^{(1)}\hat{V}^{(2)\dagger}}_{\hat{W}^{(2)}} \cdots \underbrace{\hat{V}^{(M-1)}\hat{V}^{(M)\dagger}}_{\hat{W}^{(M)}} e^{-\frac{i\delta t}{n}u_M\hat{\Lambda}^{(M)}} \times \\ &\quad \times \underbrace{\hat{V}^{(M)} e^{-\frac{i\delta t}{n}\hat{H}^{(0)}} \hat{V}^{(1)\dagger}}_{\hat{W}^{(1)}(n^{-1}\delta t)} e^{-\frac{i\delta t}{n}u_1\hat{\Lambda}^{(1)}} \cdots \hat{V}^{(M)}.\end{aligned}$$

By generalizing the notation from the $M = 1$ case we can define the M multiindices $\boldsymbol{J} = (\boldsymbol{j}^{(1)}, \dots, \boldsymbol{j}^{(M)})$ and write

$$[\hat{U}_n(\boldsymbol{u})]_{ik} = (\sum_{\boldsymbol{j}^{(1)}\dots\boldsymbol{j}^{(M)}l} e^{-i\delta t\sum_{\mu=1}^{M} u_\mu(\boldsymbol{q}_{\boldsymbol{j}^{(\mu)}}\cdot\boldsymbol{\lambda}^{(\mu)})} V^{(M)\dagger}_{il} W^{(1)}_{lj_1^{(1)}}(n^{-1}\delta t) W^{(2)}_{j_1^{(1)}j_1^{(2)}}\cdots W^{(M)}_{j_1^{(M-1)}j_1^{(M)}}\cdots$$
$$\cdots W^{(1)}_{j_{n-1}^{(M)}j_n^{(1)}}(n^{-1}\delta t) W^{(2)}_{j_n^{(1)}j_n^{(2)}}\cdots W^{(M)}_{j_n^{(M-1)}j_n^{(M)}} V_{j_n^{(M)}k})$$
$$= \sum_{\boldsymbol{J}\in[D]^{n\times M}} e^{-i\delta t\sum_{\mu=1}^{M} u_\mu\omega_{\boldsymbol{j}^{(\mu)}}} A^{\boldsymbol{J}}_{ik}(n,\delta t) = \sum_{\boldsymbol{\omega}\in\mathcal{S}_n} e^{-i\delta t\sum_{\mu=1}^{M} u_\mu\omega_\mu} B^{\boldsymbol{\omega}}_{ik}(n,\delta t) = \sum_{\boldsymbol{\omega}\in\mathcal{S}_n} e^{-i\delta t\boldsymbol{\omega}\cdot\boldsymbol{u}} B^{\boldsymbol{\omega}}_{ik}(n,\delta t),$$

where we defined $\boldsymbol{\omega} = (\omega_1,\dots,\omega_M)^T \in \mathcal{S}_n$ and we generalized the symbol $\mathcal{S}_n := \mathcal{S}_n^{(1)} \times \cdots \times \mathcal{S}_n^{(M)}$ compared to the single control case. Anyways the two definitions coincide for $M = 1$. Now by stacking the N time-step unitaries we can build up the full unitary $\hat{U}_n(\boldsymbol{U})$. We can define the multiindices $\boldsymbol{J^{(\nu)}} = (\boldsymbol{j}^{(1\nu)},\dots,\boldsymbol{j}^{(M\nu)})$

$$\hat{U}_n(\boldsymbol{U}) = \hat{U}_n(\boldsymbol{u}_N)\cdots\hat{U}_n(\boldsymbol{u}_1)$$
$$= \sum_{\boldsymbol{J}^{(1)}\dots\boldsymbol{J}^{(N)}} e^{-i\delta t\sum_{\mu\nu} u_{\mu\nu}\omega_{\boldsymbol{j}^{(\mu\nu)}}} \hat{A}^{\boldsymbol{J}^{(N)}}\cdots\hat{A}^{\boldsymbol{J}^{(1)}}$$
$$= \sum_{\boldsymbol{\omega}^{(1)}\dots\boldsymbol{\omega}^{(N)}} e^{-i\delta t\sum_{\mu\nu} u_{\mu\nu}\omega_{\mu\nu}} \hat{B}^{\boldsymbol{\omega}^{(N)}}\cdots\hat{B}^{\boldsymbol{\omega}^{(1)}}$$
$$= \sum_{\boldsymbol{\Omega}\in\mathcal{S}_n^N} e^{-i\delta t\mathrm{Tr}(\boldsymbol{\Omega}^T\boldsymbol{U})}\hat{B}^{\boldsymbol{\Omega}}(n,\delta t)$$

where we defined the frequency matrix $\boldsymbol{\Omega}$

$$\boldsymbol{\Omega} = \begin{pmatrix} \omega_{11} & \dots & \omega_{1N} \\ \vdots & \ddots & \vdots \\ \omega_{M1} & \dots & \omega_{MN} \end{pmatrix}$$

which is an element of the set $\mathcal{S}_n^N = \prod_{\mu\nu=1}^{MN}\mathcal{S}_n^{(\mu)}$. By plugging this expression for the unitary operator back into Eq. (2) we finally find the Lie-Fourier representation J_n of the quantum dynamical landscape J associated to the observable $\hat{O}$

$$J_n(\boldsymbol{U}) = \langle\psi|\hat{U}^\dagger(\boldsymbol{U})\hat{O}\hat{U}(\boldsymbol{U})|\psi\rangle = \langle\psi|\left(\sum_{\boldsymbol{\Omega''}\in\mathcal{S}_n^N} e^{i\delta t\mathrm{Tr}(\boldsymbol{\Omega''}^T\boldsymbol{U})}\hat{B}^{\boldsymbol{\Omega''}\dagger}(n,\delta t)\right)\hat{O}\left(\sum_{\boldsymbol{\Omega'}\in\mathcal{S}_n^N} e^{-i\delta t\mathrm{Tr}(\boldsymbol{\Omega'}^T\boldsymbol{U})}\hat{B}^{\boldsymbol{\Omega'}}(n,\delta t)\right)|\psi\rangle =$$
$$= \sum_{\boldsymbol{\Omega'},\boldsymbol{\Omega''}\in\mathcal{S}_n^N}\langle\psi|\hat{B}^{\boldsymbol{\Omega''}\dagger}(n,\delta t)\hat{O}\hat{B}^{\boldsymbol{\Omega'}}(n,\delta t)|\psi\rangle e^{-i\delta t\mathrm{Tr}((\boldsymbol{\Omega'}-\boldsymbol{\Omega''})^T\boldsymbol{U})} = \sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^\Delta)^N} c_{\boldsymbol{\Omega}}(n,\delta t)e^{-i\delta t\mathrm{Tr}(\boldsymbol{\Omega}^T\boldsymbol{U})}$$

where $(\mathcal{S}_n^\Delta)^N = \prod_{\mu\nu=1}^{MN}\mathcal{S}_n^{(\mu)\Delta}$ and $\mathcal{S}_n^{(\mu)\Delta}$ is again the set of frequency differences within the Fourier spectrum of the timestep unitaries, for each control

$$\mathcal{S}_n^{(\mu)\Delta} = \{\omega = \omega' - \omega''|\omega',\omega''\in\mathcal{S}_n^{(\mu)}\},$$

and the coefficients of the expansion $c_{\boldsymbol{\Omega}}$ are given by

$$c_{\boldsymbol{\Omega}}(n,\delta t) = \sum_{\boldsymbol{\Omega'},\boldsymbol{\Omega''}\in\mathcal{S}_n^N}\delta_{\Omega,\Omega'-\Omega''}\langle\psi|\hat{B}^{\boldsymbol{\Omega''}\dagger}\hat{O}\hat{B}^{\boldsymbol{\Omega'}}|\psi\rangle$$
$$= \sum_{\boldsymbol{\Omega'}\in\mathcal{S}_n^N}\langle\psi|\hat{B}^{(\boldsymbol{\Omega'}-\boldsymbol{\Omega})\dagger}\hat{O}\hat{B}^{\boldsymbol{\Omega'}}|\psi\rangle$$

1. Proofs valid for multiple controls

Lemma 3 (L_2 Boundedness of the coefficients)**.**

$$\forall n \sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^\Delta)^N}|c_{\boldsymbol{\Omega}}(n,\delta t)|^2 \le ||\hat{O}||_\infty^2$$

Proof. We first notice the following:

$$\mathbb{E}_{\boldsymbol{U}\in\mathcal{C}}[|J_n|^2] = \frac{1}{\mathrm{Vol}(\mathcal{C})}\int_{\mathcal{C}} d^{MN}\boldsymbol{U}\ J_n^*(\boldsymbol{U})J_n(\boldsymbol{U})$$
$$= \sum_{\boldsymbol{\Omega},\boldsymbol{\Omega'}\in(\mathcal{S}_n^\Delta)^N} c^*_{\boldsymbol{\Omega}}c_{\boldsymbol{\Omega'}}\prod_{\mu,\nu=1}^{M,N}\int_{-u_{\max}^{(\mu)}}^{u_{\max}^{(\mu)}}\frac{du}{2u_{\max}^{(\mu)}}\ e^{i\delta t(\omega'_{\mu\nu}-\omega_{\mu\nu})u}$$
$$= \sum_{\boldsymbol{\Omega},\boldsymbol{\Omega'}\in(\mathcal{S}_n^\Delta)^N} c^*_{\boldsymbol{\Omega}}c_{\boldsymbol{\Omega'}}\prod_{\mu,\nu=1}^{M,N}\frac{\sin\delta t(\omega'_{\mu\nu}-\omega_{\mu\nu})u_{\max}^{(\mu)}}{\delta t(\omega'_{\mu\nu}-\omega_{\mu\nu})u_{\max}^{(\mu)}}$$

Now we fix $\forall\mu\ u_{\max}^{(\mu)} = u_{\max}$ and take the following limit:

$$\lim_{u_{\max}\to\infty}\mathbb{E}_{\boldsymbol{U}\in\mathcal{C}_{u_{\max}}}[|J_n|^2] = \sum_{\boldsymbol{\Omega},\boldsymbol{\Omega'}\in(\mathcal{S}_n^\Delta)^N} c^*_{\boldsymbol{\Omega}}c_{\boldsymbol{\Omega'}}\prod_{\mu,\nu=1}^{M,N}\delta_{\omega_{\mu\nu},\omega'_{\mu\nu}}$$
$$= \sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^\Delta)^N}|c_{\boldsymbol{\Omega}}|^2.$$

But then we can make use of the following inequality:

$$|J_n(\boldsymbol{U})| = |\langle\psi|\hat{U}_n^\dagger(\boldsymbol{U})\hat{O}\hat{U}_n(\boldsymbol{U})|\psi\rangle|$$
$$\leq ||\hat{U}_n^\dagger(\boldsymbol{U})\hat{O}\hat{U}_n(\boldsymbol{U})||_\infty \leq ||\hat{O}||_\infty$$

where we made use of the sub-multiplicativity of the sup-norm and of the fact that for any unitary $||\hat{U}||_\infty = 1$. This implies a bound on the expectation value as

$$\mathbb{E}_{\boldsymbol{U}\in\mathcal{C}_{u_{\max}}}[|J_n|^2] \leq \frac{1}{\mathrm{Vol}(\mathcal{C}_{u_{\max}})}\int_{\mathcal{C}_{u_{\max}}} d^{MN}\boldsymbol{U}\ ||\hat{O}||_\infty^2 = ||\hat{O}||_\infty^2.$$

This bound will therefore be also valid in the limit $u_{\max}\to\infty$, allowing us to conclude. The result in the main text follows by fixing $M = 1$ and noticing that $||\,|\chi\rangle\langle\chi|\,||_\infty = 1$. □

Lemma 11 (Uniform convergence over compact sets)**.** *J_n converges uniformly to J over $\boldsymbol{U}\in[-u_{\max},u_{\max}]^{M\times N}$ for any $u_{\max}\geq 0$, and the same is true for the derivatives of any order:*

$$\sup_{\boldsymbol{U}\in[-u_{\max},u_{\max}]^{M\times N}}\left|\prod_{p=1}^{P}\partial_{\mu_p\nu_p}(J(\boldsymbol{U})-J_n(\boldsymbol{U}))\right|\xrightarrow{n\to\infty}0$$

Proof. Let us call $\mathcal{B}_{u_{\max}} = \{z\in\mathbb{C} : |z| < u_{\max}\}$, $\bar{\mathcal{B}}_{u_{\max}} = \{z\in\mathbb{C} : |z|\leq u_{\max}\}$ and consider the holomorphic extensions of J_n, J. It is easy to see that they are given by

$$J_n(\boldsymbol{U}) := \langle\psi|\hat{U}_n^\dagger(\boldsymbol{U}^*)\hat{O}\hat{U}_n(\boldsymbol{U})|\psi\rangle$$
$$J(\boldsymbol{U}) := \langle\psi|\hat{U}^\dagger(\boldsymbol{U}^*)\hat{O}\hat{U}(\boldsymbol{U})|\psi\rangle$$

for any $\boldsymbol{U}\in\mathbb{C}^{M\times N}$. These functions are holomorphic as they are combinations of exponentials, finite sums and products, and they are only functions of $\boldsymbol{U}$ (and not of $\boldsymbol{U}^*$). The fact that the Lie product expansion of $\hat{U}(\boldsymbol{u})$ converges uniformly over $\boldsymbol{u}\in\bar{\mathcal{B}}_{u_{\max}}^M$ is a slight modification of the classic result due to Lie. Following e.g. [28], we can define

$$\hat{S}_n = \exp\left(-\frac{i\delta t}{n}\hat{H}(\boldsymbol{u})\right),\quad \hat{T}_n = e^{-\frac{i\delta t}{n}\hat{H}^{(0)}}\prod_{\mu=1}^{M}e^{-\frac{i\delta t}{n}u_\mu\hat{H}^{(\mu)}},$$

so that we have $\hat{U}(\boldsymbol{u}) = \hat{S}_n^n, \hat{U}_n(\boldsymbol{u}) = \hat{T}_n^n$ and

$$||\hat{U}(\boldsymbol{u})-\hat{U}_n(\boldsymbol{u})||_\infty \leq n||\hat{S}_n-\hat{T}_n||e^\zeta$$

where

$$\zeta := \delta t\left(||\hat{H}^{(0)}||_\infty + \frac{u_{\max}M\omega_{\max}}{2}\right).$$

We can estimate $||\hat{S}_n-\hat{T}_n||$ using the triangular inequality

$$||\hat{S}_n-\hat{T}_n|| \leq ||\hat{S}_n-\hat{I}-\frac{i\delta t}{n}\hat{H}(\boldsymbol{u})|| + ||\hat{T}_n-\hat{I}-\frac{i\delta t}{n}\hat{H}(\boldsymbol{u})||$$

and the result on exponential approximations shown in Appendix A, which gives the result

$$||\hat{U}(\boldsymbol{u})-\hat{U}_n(\boldsymbol{u})||_\infty \leq n\left(2e^{\frac{\zeta}{n}}\frac{\zeta^2}{n^2}\right)e^\zeta \xrightarrow{n\to\infty} 0$$

This property trivially extends to finite products, in fact given two parametrized operator sequences $\hat{X}_n(\boldsymbol{u}), \hat{Y}_n(\boldsymbol{u'})$ with $\boldsymbol{u},\boldsymbol{u'}\in\bar{\mathcal{B}}_{u_{\max}}^M$, we have

$$||\hat{X}_n(\boldsymbol{u})\hat{Y}_n(\boldsymbol{u'})-\hat{X}(\boldsymbol{u})\hat{Y}(\boldsymbol{u'})||_\infty \leq$$
$$||\hat{X}_n(\boldsymbol{u})||_\infty||\hat{Y}_n(\boldsymbol{u'})-\hat{Y}(\boldsymbol{u'})||_\infty +$$
$$+||\hat{X}_n(\boldsymbol{u})-\hat{X}(\boldsymbol{u})||_\infty||\hat{Y}(\boldsymbol{u'})||_\infty.$$

But then if the two sequences converge uniformly and are uniformly bounded,

$$\sup_{\boldsymbol{u}\in\bar{\mathcal{B}}_{u_{\max}}^M}||\hat{X}_n(\boldsymbol{u})-\hat{X}(\boldsymbol{u})||\xrightarrow{n\to\infty}0,$$
$$\sup_{\boldsymbol{u}\in\bar{\mathcal{B}}_{u_{\max}}^M}||\hat{Y}_n(\boldsymbol{u})-\hat{Y}(\boldsymbol{u})||\xrightarrow{n\to\infty}0,$$
$$\sup_{\boldsymbol{u}\in\bar{\mathcal{B}}_{u_{\max}}^M}||\hat{X}_n(\boldsymbol{u})||\leq X_{\max},\ \sup_{\boldsymbol{u}\in\bar{\mathcal{B}}_{u_{\max}}^M}||\hat{Y}_n(\boldsymbol{u})||\leq Y_{\max}$$

which is satisfied by the timestep unitaries, as we have

$$||\hat{U}(\boldsymbol{u})||_\infty, ||\hat{U}_n(\boldsymbol{u})||_\infty \leq e^\zeta,$$

then also the product converges uniformly. But then also the approximants J_n converge uniformly to J, in fact

$$\sup_{\boldsymbol{U}\in\bar{\mathcal{B}}_{u_{\max}}^{M\times N}}|J_n(\boldsymbol{U})-J(\boldsymbol{U})|$$
$$= \sup_{\boldsymbol{U}\in\bar{\mathcal{B}}_{u_{\max}}^{M\times N}}|\langle\psi|\hat{U}_n^\dagger(\boldsymbol{U}^*)\hat{O}\hat{U}_n(\boldsymbol{U})-\hat{U}^\dagger(\boldsymbol{U}^*)\hat{O}\hat{U}(\boldsymbol{U})|\psi\rangle|$$
$$\leq \sup_{\boldsymbol{U}\in\bar{\mathcal{B}}_{u_{\max}}^{M\times N}}||\hat{U}_n^\dagger(\boldsymbol{U}^*)\hat{O}\hat{U}_n(\boldsymbol{U})-\hat{U}^\dagger(\boldsymbol{U}^*)\hat{O}\hat{U}(\boldsymbol{U})||_\infty$$
$$\leq \epsilon_n \xrightarrow{n\to\infty} 0.$$

for some infinitesimal sequence $\epsilon_n \geq 0$. Now we just need to show that this notion of convergence extends to the derivatives $\partial_{11}^{p_{11}} \cdots \partial_{MN}^{p_{MN}} J_n(\boldsymbol{U})$. Since as we have already seen J_n, J are holomorphic functions, we can write the partial derivatives as contour integrals by means of Cauchy's theorem [42], from which follows easily the uniform convergence on all compacts of the partial derivatives. In fact, for all $\boldsymbol{U} \in \mathcal{B}_{u_{\max}}$ we have

$$
\begin{aligned}
|\partial_{11}^{p_{11}} \cdots \partial_{MN}^{p_{MN}} (J - J_n)(\boldsymbol{U})| &= \left| \frac{p_{11}! \cdots p_{MN}!}{(2\pi i)^{MN}} \int_{\partial \bar{\mathcal{B}}_{u_{\max}}} du'_{11} \int_{\partial \bar{\mathcal{B}}_{u_{\max}}} du'_{MN} \frac{J(\boldsymbol{U}') - J_n(\boldsymbol{U}')}{(u'_{11} - u_{11})^{p_{11}+1} \cdots (u'_{MN} - u_{MN})^{p_{MN}+1}} \right| \\
&\leq \frac{p_{11}! \cdots p_{MN}!}{(2\pi)^{MN}} \int_0^{2\pi} d\phi'_{11} u_{\max} \int_0^{2\pi} d\phi'_{MN} u_{\max} \sup_{\boldsymbol{U}' \in \partial \bar{\mathcal{B}}_{u_{\max}}^{M \times N}} \left(\frac{|J(\boldsymbol{U}') - J_n(\boldsymbol{U}')|}{|u'_{11} - u_{11}|^{p_{11}+1} \cdots |u'_{MN} - u_{MN}|^{p_{MN}+1}} \right) \\
&\leq \epsilon_n \frac{p_{11}! \cdots p_{MN}!}{(2\pi)^{MN}} \frac{2\pi u_{\max}}{a_{11}^{p_{11}+1}} \cdots \frac{2\pi u_{\max}}{a_{MN}^{p_{MN}+1}} \xrightarrow{n \to \infty} 0,
\end{aligned}
$$

where we defined $a_{\mu\nu} = \inf_{u' \in \partial \bar{\mathcal{B}}_{u_{\max}}} |u_{\mu\nu} - u'_{\mu\nu}| > 0$ which is strictly larger than zero, as $u_{\mu\nu}$ lies in the open ball of radius $u_{\max}$, while $u'_{\mu\nu}$ lies on its boundary, the circle of radius $u_{\max}$. So we proved uniform convergence also for the derivatives on the sets $\mathcal{B}_{u_{\max}}^{M \times N}$ for any $u_{\max} \geq 0$. But then the same result clearly holds for the restrictions to real controls $\boldsymbol{U} \in [-u_{\max}, u_{\max}]^{M \times N} \subset \mathcal{B}_{u_{\max}+\delta}^{M \times N}$ for $\delta > 0$, since the complex derivatives reduce in this case to the usual real derivatives. □

We can prove a generalized version of the result presented in the main text concerning the partial derivatives of the landscape J. We assume without loss of generality that $\omega_{\max}^{(\mu)} = \omega_{\max}$ (see Appendix B for details). The result is first shown to be valid for each of the approximants J_n and then uniform convergence is used to conclude that the result is also valid for J.

Lemma 5 (Boundedness of the derivatives)**.** *The derivatives of any order $P \geq 1$ of $J(\boldsymbol{U})$ are bounded by*

$$
\left| \prod_{p=1}^{P} \partial_{\mu_p \nu_p} J(\boldsymbol{U}) \right| \leq (\delta t \omega_{\max})^P \frac{\Delta O}{2}
$$

where $\Delta O = (\max_{|\psi\rangle \in \mathcal{H}} - \min_{|\psi\rangle \in \mathcal{H}}) \langle \psi | \hat{O} | \psi \rangle$.

Proof. Since J_n is the expectation value of $\hat{O}$ on a quantum state generated by the action of a unitary on the initial state, namely

$$
J_n(\boldsymbol{U}) = \langle \psi | \hat{U}_n^\dagger(\boldsymbol{U}) \hat{O} \hat{U}_n(\boldsymbol{U}) | \psi \rangle
$$

we have

$$
\forall \boldsymbol{U}, n \quad \min_{|\psi\rangle \in \mathcal{H}} \langle \psi | \hat{O} | \psi \rangle \leq J_n(\boldsymbol{U}) \leq \max_{|\psi\rangle \in \mathcal{H}} \langle \psi | \hat{O} | \psi \rangle .
$$

But then if we define

$$
\bar{O} = \frac{(\max_{|\psi\rangle \in \mathcal{H}} + \min_{|\psi\rangle \in \mathcal{H}}) \langle \psi | \hat{O} | \psi \rangle}{2}
$$

the function $\tilde{J}_n(\boldsymbol{U}) = J_n(\boldsymbol{U}) - \bar{O}$ has the same derivatives of order $P \geq 1$ as J_n and satisfies $\tilde{J}_n(\boldsymbol{U}) \in [-\Delta O/2, \Delta O/2]$.

For any given $n, \boldsymbol{U}, \mu, \nu$ we can define the restriction

$$
F(u) = \tilde{J}_n(u_{11}, \ldots, \underbrace{u}_{\mu\nu}, \ldots, u_{MN}),
$$

so that by the definition of partial derivative we get

$$
\frac{d}{du} F(u) = \partial_{\mu\nu} J_n(\boldsymbol{U}).
$$

Clearly, since F is a restriction of $\tilde{J}_n$, we also have $F(u) \in [-F_{\max}, F_{\max}]$, with $F_{\max} = \Delta O/2$.

By rearranging the order of the frequency sums in the Lie-Fourier representation J_n, we can see that F is a linear combination of complex exponentials in the variable u:

$$
F(u) = \sum_{\boldsymbol{\Omega} \in (\mathcal{S}_n^\Delta)^N} c_{\boldsymbol{\Omega}}(n, \delta t) e^{-i\delta t \mathrm{Tr}(\boldsymbol{\Omega}^T \boldsymbol{U})} \Big|_{u_{\mu\nu} = u} = \sum_{\omega_{\mu\nu} \in \mathcal{S}_n^{(\mu)\Delta}} \underbrace{\sum_{\omega_{11} \in \mathcal{S}_n^{(1)\Delta}} \cdots \sum_{\omega_{MN} \in \mathcal{S}_n^{(M)\Delta}} c_{\boldsymbol{\Omega}} e^{-i\delta t \sum_{\alpha\beta \neq \mu\nu} \omega_{\alpha\beta} u_{\alpha\beta}}}_{c'_{\omega_{\mu\nu}}(n, \delta t, u_{\alpha\beta \neq \mu\nu})} e^{-i\delta t \omega_{\mu\nu} u}
$$

This implies several properties: first of all $\forall n, \boldsymbol{U}$ F is a holomorphic function when we extend the variable to complex values $u \in \mathbb{C}$. F is bandwidth limited, namely $\omega_{\mu\nu} \in [-\omega_{\max}, \omega_{\max}]$. It also means that $F = \mathcal{O}(e^{a|s|})$ with $a \leq a_{\max} = \delta t \omega_{\max}$.

Then, the conditions for Bernstein's Inequality as proved in [60], Vol.2 p.276 are all satisfied, which in our case implies that $\forall n, \boldsymbol{U}$

$$|\partial_{\mu\nu} J_n(\boldsymbol{U})| = \left| \frac{d}{du} F(u) \right| \leq a_{\max} F_{\max} = \delta t \omega_{\max} \frac{\Delta O}{2}.$$

An alternative proof (although of a version of the inequality which is looser by a factor of 2) can be found in [61], Sec. 2.3.8. So the case $P = 1$ is settled. If now we suppose the bound

$$\left| \prod_{p=1}^{P-1} \partial_{\mu_p \nu_p} J_n(\boldsymbol{U}) \right| \leq (\delta t \omega_{\max})^{P-1} \frac{\Delta O}{2}$$

to be true, we can now show by the same argument as before that it also holds for P. In fact, by defining the restriction $F(u)$ as

$$F(u) = \prod_{p=1}^{P-1} \partial_{\mu_p \nu_p} \tilde{J}_n(u_{11}, \ldots, \underbrace{u}_{\mu\nu}, \ldots, u_{MN}),$$

we have that $F(u)$ satisfies all the conditions as in the $P = 1$ case, with $a_{\max} = \delta t \omega_{\max}$ and $F_{\max} = (\delta t \omega_{\max})^{P-1} \Delta O/2$. But then by Bernstein's Inequality we have again $\forall n, \boldsymbol{U}$

$$\left| \prod_{p=1}^{P} \partial_{\mu_p \nu_p} J_n(\boldsymbol{U}) \right| = \left| \frac{d}{du} F(u) \right| \leq a_{\max} F_{\max} = \delta t \omega_{\max} (\delta t \omega_{\max})^{P-1} \frac{\Delta O}{2}.$$

We can then conclude by induction that the bound on the derivatives of J_n is true for all $P \geq 1$.

We now show that the same bounds hold for the function J itself. Now, since the bounds apply to the partial derivatives $\forall n$ and are constant, the same bounds will also hold in the limit of large n, and this limit coincides with the partial derivatives of J.

The case treated in the main text is easily obtained by setting $M = 1$ and noting that in this case $E = \max_{|\psi\rangle \in \mathcal{H}} |\langle\psi| (|\chi\rangle\langle\chi|) |\psi\rangle| = |\langle\chi|\chi\rangle|^2 = 1$. □

We point out that while we derived this bound on the derivatives as a consequence of the Fourier representation, it is likely not the only way to prove a similar result. Other proofs might be possible e.g. by using the standard formula [50]

$$\frac{d}{ds} e^{A+sB}\big|_{s=0} = e^A \int_0^1 e^{At} B e^{-At} dt$$

or a variation thereof.

Under the same assumptions of M multiple controls and generic observable $\hat{O}$ we can also prove the result concerning the variance of the first order derivatives.

Lemma 6 (Lipschitz continuity)**.** *The function $J(\boldsymbol{u})$ is Lipschitz continuous, that is*

$$|J(\boldsymbol{U}) - J(\boldsymbol{U'})| \leq K ||\boldsymbol{U} - \boldsymbol{U'}||_1$$

where the Lipschitz constant $K \geq 0$ is upper bounded by $\omega_{\max} \delta t \Delta O/2$. Moreover, if $\boldsymbol{U}$ is a critical point, so that $\forall \mu, \nu \ \partial_{\mu\nu} J(\boldsymbol{U}) = 0$, then the following inequality holds

$$|J(\boldsymbol{U}) - J(\boldsymbol{U'})| \leq K_c ||\boldsymbol{U} - \boldsymbol{U'}||_1^2.$$

where $0 \leq K_c \leq (\omega_{\max} \delta t)^2 \Delta O/2$.

Proof. As we already discussed, $J : \mathbb{R}^{M \times N} \to \mathbb{R}$ is a continuously differentiable function (as it is the restriction to real controls of a holomorphic function) and has bounded partial derivatives. Let us consider its restriction to the straight line connecting $\boldsymbol{U}$ to $\boldsymbol{U'}$

$$\tilde{J} : \mathbb{R} \to \mathbb{R}, \tilde{J}(l) = J(\boldsymbol{U}(l)) = J((\boldsymbol{U} - \boldsymbol{U'})l + \boldsymbol{U'}),$$

so that $\tilde{J}(1) = J(\boldsymbol{U})$ and $\tilde{J}(0) = J(\boldsymbol{U'})$, and the restricted function is also continuously differentiable, with derivatives given by the chain rule. Because of the mean value theorem, we have

$$\tilde{J}(1) - \tilde{J}(0) = \frac{d\tilde{J}}{dl}\bigg|_{l=c} = \sum_{\mu,\nu=1}^{M,N} \partial_{\mu\nu} J(\boldsymbol{U}(c)) (\boldsymbol{U} - \boldsymbol{U'})_{\mu\nu}$$

for some $0 \leq c \leq 1$. But then by using the bound from Lemma 5 we have the result

$$|J(\boldsymbol{U}) - J(\boldsymbol{U'})| \leq \sum_{\mu,\nu=1}^{M,N} |\partial_{\mu\nu} J(\boldsymbol{U}(c))| |\boldsymbol{U} - \boldsymbol{U'}|_{\mu\nu} \leq \frac{\omega_{\max} \delta t \Delta O}{2} ||\boldsymbol{U} - \boldsymbol{U'}||_1.$$

If $\boldsymbol{U}$ is a critical point $\forall \mu, \nu \ \partial_{\mu\nu} J(\boldsymbol{U}) = 0$, we have that

$$\frac{d\tilde{J}}{dl}\bigg|_{l=0} = 0,$$

so we can apply the mean value theorem twice, obtaining:

$$\tilde{J}(1) - \tilde{J}(0) = \frac{d\tilde{J}}{dl}\bigg|_{l=c} = \frac{d^2\tilde{J}}{dl^2}\bigg|_{l=d} d = d \sum_{\mu,\nu=1}^{M,N} \sum_{\mu',\nu'=1}^{M,N} \partial_{\mu\nu} \partial_{\mu'\nu'} J(\boldsymbol{U}(d)) (\boldsymbol{U} - \boldsymbol{U'})_{\mu\nu} (\boldsymbol{U} - \boldsymbol{U'})_{\mu'\nu'}$$

for some $0 \leq d \leq c$. But then by using the bound from Lemma 5 we have the result

$$|J(\boldsymbol{U}) - J(\boldsymbol{U'})| \leq \sum_{\mu,\nu=1}^{M,N} \sum_{\mu',\nu'=1}^{M,N} |\partial_{\mu\nu} \partial_{\mu'\nu'} J(\boldsymbol{U}(d))| |\boldsymbol{U} - \boldsymbol{U'}|_{\mu\nu} |\boldsymbol{U} - \boldsymbol{U'}|_{\mu'\nu'} \leq \frac{(\omega_{\max} \delta t)^2 \Delta O}{2} ||\boldsymbol{U} - \boldsymbol{U'}||_1^2.$$

By setting $M = 1$ and $\Delta O = 1$, consistently with a fidelity landscape for a single control, we obtain the result from the main text. □

Corollary 1 (Linear parametrizations of controls)**.** *Let $\boldsymbol{U}(\boldsymbol{V})$ be a linear parametrization of the controls given by $u_{\mu\nu} = \sum_{\mu'\nu'} r_{\mu\nu\mu'\nu'} v_{\mu'\nu'}$, where $\mu' = 1, \dots, M_c$ and $\nu' = 1, \dots, N_c$, which is normalized as follows:*

$$\forall \mu', \nu' \quad \sum_{\mu=1}^{M_c} \sum_{\nu=1}^{N_c} |r_{\mu\nu\mu'\nu'}| = 1.$$

Then, the partial derivatives of the parametrized landscape $\tilde{J}(\boldsymbol{V}) = J(\boldsymbol{U}(\boldsymbol{V}))$ obey the bounds

$$\left| \prod_{p=1}^{P} \tilde{\partial}_{\mu'_p\nu'_p} \tilde{J}(\boldsymbol{V}) \right| \leq (\delta t \omega_{\max})^P \frac{\Delta O}{2} \prod_{p=1}^{P} \sum_{\mu_p,\nu_p=1}^{M_c,N_c} |r_{\mu_p\nu_p\mu'_p\nu'_p}| = (\delta t \omega_{\max})^P \frac{\Delta O}{2}.$$

where $\tilde{\partial}_{\mu'\nu'} := \frac{\partial}{\partial v_{\mu'\nu'}}$ and $\partial_{\mu\nu} := \frac{\partial}{\partial u_{\mu\nu}}$.

Proof. Let us start with a preliminary step by computing

$$\tilde{\partial}_{\mu'\nu'} u_{\mu\nu}(\boldsymbol{V}) = \tilde{\partial}_{\mu'\nu'} \sum_{\alpha=1}^{M_c} \sum_{\beta=1}^{N_c} r_{\mu\nu\alpha\beta} v_{\alpha\beta} = r_{\mu\nu\mu'\nu'},$$

with the higher order derivatives are zero because $\boldsymbol{U}(\boldsymbol{V})$ is a linear function. For this reason, by applying the chain rule for differentiation p times and using the bounds from Lemma 5 we obtain the first result:

$$\begin{aligned} &\left| \prod_{p=1}^{P} \tilde{\partial}_{\mu'_p\nu'_p} \tilde{J}(\boldsymbol{V}) \right| = \\ &= \left| \prod_{p=1}^{P} \left(\sum_{\mu_p=1}^{M_c} \sum_{\nu_p=1}^{N_c} r_{\mu_p\nu_p\mu'_p\nu'_p} \partial_{\mu_p\nu_p} \right) J(\boldsymbol{U}(\boldsymbol{V})) \right| \\ &\leq (\delta t \omega_{\max})^P \frac{\Delta O}{2} \prod_{p=1}^{P} \left(\sum_{\mu_p=1}^{M_c} \sum_{\nu_p=1}^{N_c} |r_{\mu_p\nu_p\mu'_p\nu'_p}| \right). \end{aligned}$$

If then we require the condition

$$\forall \mu', \nu' \quad \sum_{\mu=1}^{M_c} \sum_{\nu=1}^{N_c} |r_{\mu\nu\mu'\nu'}| = 1,$$

we have the same bound as the one used in the proof of Lemma 6, that is

$$\left| \prod_{p=1}^{P} \tilde{\partial}_{\mu'_p\nu'_p} \tilde{J}(\boldsymbol{V}) \right| \leq (\delta t \omega_{\max})^P \frac{\Delta O}{2}$$

Therefore, using trivially the same reasoning as the one presented in that proof we can obtain the claimed result about Lipschitz continuity with the same upper bound for K. By setting $M = 1$ and $\Delta O = 1$ we recover the result from the main text. □

Lemma 7 (Taylor approximation error in the Lagrange form)**.**

$$|J(\boldsymbol{U}) - J_P(\boldsymbol{U})| \leq \frac{\Delta O}{2} \frac{(\delta t \omega_{\max})^{P+1}}{(P+1)!} ||\boldsymbol{U} - \boldsymbol{U}_0||_1^{P+1}$$

Proof. Let us define the P-order Taylor representation of f as follows:

$$J_P(\boldsymbol{U}) = \sum_{p=0}^{P} \sum_{\mu_1\nu_1 \dots \mu_p\nu_p} \frac{\partial_{\mu_1\nu_1} \dots \partial_{\mu_p\nu_p} J(\boldsymbol{U}_0)}{p!} \prod_{i=1}^{p} (\boldsymbol{U} - \boldsymbol{U}_0)_{\mu_i\nu_i}.$$

The result can then be proved using similar tools as the ones used in Lemma 6. Let us first define the restrictions $\tilde{J}(l), \tilde{J}_P(l)$ as done previously, using the parametrization

$$l \in [0,1] \mapsto \boldsymbol{U}(l) = (\boldsymbol{U} - \boldsymbol{U}_0) l + \boldsymbol{U}_0.$$

It is easy to check that the derivatives of $\tilde{J}, \tilde{J}_P$ evaluated at $l = 0$ are the same up to order P:

$$\begin{aligned} \left. \frac{d^p \tilde{J}_P}{dl^p} \right|_{l=0} &= \sum_{\mu_1\nu_1 \dots \mu_p\nu_p} \partial_{\mu_1\nu_1} \dots \partial_{\mu_p\nu_p} J(\boldsymbol{U}_0) \prod_{i=1}^{p} (\boldsymbol{U} - \boldsymbol{U}_0)_{\mu_i\nu_i} \\ &= \left. \frac{d^p \tilde{J}}{dl^p} \right|_{l=0} \quad \forall\ 0 \leq p \leq P. \end{aligned}$$

But then we can apply the mean value theorem to the remainder $r(l) := \tilde{J}(l) - \tilde{J}_P(l)$:

$$\frac{d^P r(l)}{dl^P} = \left. \frac{d^{P+1} r}{dl^{P+1}} \right|_{l=c} l = \left. \frac{d^{P+1} \tilde{J}}{dl^{P+1}} \right|_{l=c} l$$

where $0 \leq c \leq l$. By integrating both sides P times we get:

$$\begin{aligned} \int_0^l dl_1 \cdots \int_0^{l_{P-1}} dl_P \frac{d^P r(l_P)}{dl^P} &= r(l) \\ &= \left. \frac{d^{P+1} \tilde{J}}{dl^{P+1}} \right|_{l=c} \int_0^l dl_1 \cdots \int_0^{l_{P-1}} dl_P l_P \\ &= \left. \frac{d^{P+1} \tilde{J}}{dl^{P+1}} \right|_{l=c} \frac{l^{(P+1)}}{(P+1)!}. \end{aligned}$$

By developing the derivative and evaluating the expression at $l = 1$ we get:

$$J(\boldsymbol{U}) - J_P(\boldsymbol{U}) = \sum_{\mu_1\nu_1 \dots \mu_{P+1}\nu_{P+1}} \frac{\partial_{\mu_1\nu_1} \dots \partial_{\mu_{P+1}\nu_{P+1}} J(\boldsymbol{U})}{(P+1)!} \prod_{i=1}^{P+1} (\boldsymbol{U} - \boldsymbol{U}_0)_{\mu_i\nu_i}$$

from which the claimed result follows readily by applying the derivative bound found in Lemma 5. □

Lemma 8 (Variance over bounded controls)**.** *Let* $\mathcal{C} = \prod_{\mu,\nu=1}^{M,N}[-u_{\max}^{(\mu)}, u_{\max}^{(\mu)}]$. *Then for any for any* $n \geq 1$, $P \geq 0$ *and* $1 \leq \mu_1, \dots, \mu_P \leq M$, $1 \leq \nu_1, \dots, \nu_P \leq N$ *we have:*

$$\operatorname*{Var}_{\boldsymbol{U}\in\mathcal{C}}\left[\left(\prod_{p=1}^{P}\partial_{\mu_p\nu_p}\right)J_n\right] = \delta t^{2P}\sum_{\boldsymbol{\Omega},\boldsymbol{\Omega'}\in(\mathcal{S}_n^{\Delta})^N\setminus\{\mathbf{0}\}} c^*_{\boldsymbol{\Omega}}c_{\boldsymbol{\Omega'}}\Big(\prod_{p=1}^{P}\omega_{\mu_p\nu_p}\omega'_{\mu_p\nu_p}\Big)\big(\tilde{\kappa}(\boldsymbol{\Omega}-\boldsymbol{\Omega'})-\tilde{\kappa}(\boldsymbol{\Omega})\tilde{\kappa}(\boldsymbol{\Omega'})\big), \quad \tilde{\kappa}(\boldsymbol{\Omega}) = \prod_{\mu,\nu=1}^{M,N}\frac{\sin\delta t\omega_{\mu\nu}u_{\max}^{(\mu)}}{\delta t\omega_{\mu\nu}u_{\max}^{(\mu)}}$$

Moreover, the following bounds hold:

$$\operatorname*{Var}_{\boldsymbol{U}\in\mathcal{C}}\left[\left(\prod_{p=1}^{P}\partial_{\mu_p\nu_p}\right)J_n, J\right] \leq \min\left[\frac{L^{2P}\Delta O^2}{4N^{2P}}, \frac{M(\Delta O L^{P+1}u_{\max})^2}{3N^{2P+1}} + o\left(\frac{L^{2P+2}}{N^{2P+1}}\right)\right].$$

Proof. This proof is made up of two parts: first, we show how to derive the representation of the variance in the claim, and then we use it to prove the upper bounds.

We remind that given a (real) function F, we define its variance over a set $\mathcal{C} \subset \mathbb{R}^{M\times N}$ as

$$\operatorname*{Var}_{\boldsymbol{U}\in\mathcal{C}}[F] = \mathbb{E}_{\boldsymbol{U}\in\mathcal{C}}[|F|^2] - |\mathbb{E}_{\boldsymbol{U}\in\mathcal{C}}[F]|^2, \tag{C1}$$

where

$$\mathbb{E}_{\boldsymbol{U}\in\mathcal{C}}[F] = \int_{\mathcal{C}}\frac{d^{MN}\boldsymbol{U}}{\mathrm{Vol}(\mathcal{C})}\,F(\boldsymbol{U})$$

Let us start by computing the expectation value of the derivatives:

$$\mathbb{E}_{\boldsymbol{U}\in\mathcal{C}}\left[\left(\prod_{p=1}^{P}\partial_{\mu_p\nu_p}J_n\right)\right] = \int_{\mathcal{C}}\frac{d^{MN}\boldsymbol{U}}{\mathrm{Vol}(\mathcal{C})}\left(\prod_{p=1}^{P}\partial_{\mu_p\nu_p}\right)J_n(\boldsymbol{U})$$

$$= \sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^{\Delta})^N} c_{\boldsymbol{\Omega}}\left(\prod_{p=1}^{P}-i\delta t\omega_{\mu_p\nu_p}\right)\prod_{\mu,\nu=1}^{M,N}\frac{1}{2u_{\max}^{(\mu)}}\int_{-u_{\max}^{(\mu)}}^{u_{\max}^{(\mu)}}du\; e^{-i\delta t\omega_{\mu\nu}u}$$

$$= \sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^{\Delta})^N} c_{\boldsymbol{\Omega}}\left(\prod_{p=1}^{P}-i\delta t\omega_{\mu_p\nu_p}\right)\prod_{\mu,\nu=1}^{M,N}\frac{\sin\delta t\omega_{\mu\nu}u_{\max}^{(\mu)}}{\delta t\omega_{\mu\nu}u_{\max}^{(\mu)}}. \tag{C2}$$

We can then do the same for the square modulus:

$$\mathbb{E}_{\boldsymbol{U}\in\mathcal{C}}\left[\left|\left(\prod_{p=1}^{P}\partial_{\mu_p\nu_p}\right)J_n\right|^2\right]$$

$$= \int_{\mathcal{C}}\frac{d^{MN}\boldsymbol{U}}{\mathrm{Vol}(\mathcal{C})}\left(\prod_{p=1}^{P}\partial_{\mu_p\nu_p}\right)J_n^*(\boldsymbol{U})\left(\prod_{p=1}^{P}\partial_{\mu_p\nu_p}\right)J_n(\boldsymbol{U})$$

$$= \delta t^{2P}\sum_{\boldsymbol{\Omega},\boldsymbol{\Omega'}\in(\mathcal{S}_n^{\Delta})^N} c^*_{\boldsymbol{\Omega}}c_{\boldsymbol{\Omega'}}\left(\prod_{p=1}^{P}\omega_{\mu_p\nu_p}\omega'_{\mu_p\nu_p}\right)\tilde{\kappa}(\boldsymbol{\Omega'}-\boldsymbol{\Omega}).$$

By substituting these results inside Eq. (C1) and by noticing that

$$\tilde{\kappa}(\mathbf{0}-\boldsymbol{\Omega'}) - \tilde{\kappa}(\mathbf{0})\tilde{\kappa}(\boldsymbol{\Omega'}) = 0$$

we can restrict all sums to non-zero frequencies in $(\mathcal{S}_n^{\Delta})^N\setminus\{\mathbf{0}\}$, which gives us the result in the claim. Once again, we obtain the case treated in the main text just by setting $M = 1$.

We now briefly comment on the convergence of expectation values and variances over bounded sets. Given a sequence F_n which is uniformly bounded $\forall \boldsymbol{U} \in \mathcal{C}$ $|F_n(\boldsymbol{U})| \leq F_{\max}$ and converges uniformly to F over a bounded set $\mathcal{C}$ (with non-zero volume), we have that also its expectation value over $\mathcal{C}$ will converge:

$$\begin{aligned}|\mathbb{E}_{\boldsymbol{U}\in\mathcal{C}}[F_n] - \mathbb{E}_{\boldsymbol{U}\in\mathcal{C}}[F]| &= \\ &= \frac{1}{\mathrm{Vol}(\mathcal{C})}\left|\int_{\mathcal{C}}d^{MN}\boldsymbol{U}F_n(\boldsymbol{U}) - F(\boldsymbol{U})\right| \\ &\leq \frac{1}{\mathrm{Vol}(\mathcal{C})}\int_{\mathcal{C}}d^{MN}\boldsymbol{U}|F_n(\boldsymbol{U}) - F(\boldsymbol{U})| \\ &\leq \sup_{\boldsymbol{U}\in\mathcal{C}}|F_n(\boldsymbol{U}) - F(\boldsymbol{U})| \xrightarrow{n\to\infty} 0.\end{aligned}$$

Moreover, if we have two such bounded converging sequences F_n, G_n then also their product converges uniformly and is bounded $|F_nG_n| \leq F_{\max}G_{\max}$, as can be seen from the following inequality:

$$\begin{aligned}|F_nG_n - FG| &= |F_nG_n - FG_n + FG_n - FG| \\ &\leq |F_n - F||G_n| + |G_n - G||F|.\end{aligned}$$

Because of this fact it is easy to see that also the variance (being defined in terms of expectation values of powers of F_n) converges as expected:

$$\begin{aligned}\lim_{n\to\infty}\operatorname*{Var}_{\boldsymbol{U}\in\mathcal{C}}[F_n] = \lim_{n\to\infty}\mathbb{E}_{\boldsymbol{U}\in\mathcal{C}}[F_n^2] - \lim_{n\to\infty}|\mathbb{E}_{\boldsymbol{U}\in\mathcal{C}}[F_n]|^2 \\ \mathbb{E}_{\boldsymbol{U}\in\mathcal{C}}[F^2] - |\mathbb{E}_{\boldsymbol{U}\in\mathcal{C}}[F]|^2 = \operatorname*{Var}_{\boldsymbol{U}\in\mathcal{C}}[F]\end{aligned}$$

This shows that the variance and expectation values of the Lie-Fourier representation of the landscape J_n and of its derivatives over $\mathcal{C}$ converge to the corresponding quantities computed for the true landscape J. It should be stressed that this convergence result does not necessarily hold for unbounded controls $\boldsymbol{U} \in \mathbb{R}^{M\times N}$, since in that case we do not have any guarantee of uniform convergence of J_n to J.

We now show how to obtain the upper bound in the claim. The first part of the bound is an obvious consequence of the bound of the derivatives from Lemma 5, which holds for both $F = J, J_n$. As already discussed before, by defining $\tilde{F}(\boldsymbol{U}) = F(\boldsymbol{U}) - \bar{O}$ we have $\forall \boldsymbol{U} \in \mathbb{R}^{M\times N}$, $\tilde{F}(\boldsymbol{U}) \in [-\Delta O/2, \Delta O/2]$, so that the bound

$$|\partial_{\mu\nu}\tilde{F}| \leq \frac{(\delta t\omega_{\max})^P \Delta O}{2}$$

holds also for $P = 0$. Moreover, the variance is not affected by uniform shifts, as

$$\begin{aligned}\operatorname*{Var}_{\boldsymbol{U}\in\mathcal{C}}[\tilde{F}] &= \mathbb{E}_{\boldsymbol{U}\in\mathcal{C}}[(F - \bar{O} - \mathbb{E}_{\boldsymbol{U}\in\mathcal{C}}[F] + \mathbb{E}_{\boldsymbol{U}\in\mathcal{C}}[\bar{O}])^2] \\ &= \mathbb{E}_{\boldsymbol{U}\in\mathcal{C}}[(F - \mathbb{E}_{\boldsymbol{U}\in\mathcal{C}}[F])^2] = \operatorname*{Var}_{\boldsymbol{U}\in\mathcal{C}}[F].\end{aligned}$$

But then we have

$$\begin{aligned}&\operatorname*{Var}_{\boldsymbol{U}\in\mathcal{C}}\left[\left(\prod_{p=1}^{P}\partial_{\mu_p\nu_p}\tilde{F}\right)\right] \leq \mathbb{E}_{\boldsymbol{U}\in\mathcal{C}}\left[\left|\left(\prod_{p=1}^{P}\partial_{\mu_p\nu_p}\tilde{F}\right)\right|^2\right] \\ &= \int_{\mathcal{C}}\frac{d^{MN}\boldsymbol{U}}{\operatorname{Vol}(\mathcal{C})}\left|\left(\prod_{p=1}^{P}\partial_{\mu_p\nu_p}\tilde{F}(\boldsymbol{U})\right)\right|^2 \leq \left(\frac{(T\omega_{\max})^P\Delta O}{2N^P}\right)^2.\end{aligned}$$

Concerning the second part of the bound, we first prove the result for finite n, starting by finding the asymptotic expansion of the kernel $\tilde{\kappa}(\boldsymbol{\Omega})$:

$$\begin{aligned}\log(\tilde{\kappa}(\boldsymbol{\Omega})) &= \sum_{\mu,\nu=1}^{M,N}\log\left(\frac{\sin\delta t\omega_{\mu\nu}u^{(\mu)}_{\max}}{\delta t\omega_{\mu\nu}u^{(\mu)}_{\max}}\right) = \\ &\sum_{\mu,\nu=1}^{M,N}\left(\frac{\sin\delta t\omega_{\mu\nu}u^{(\mu)}_{\max}}{\delta t\omega_{\mu\nu}u^{(\mu)}_{\max}} - 1\right) + o\left(\frac{\sin\delta t\omega_{\mu\nu}u^{(\mu)}_{\max}}{\delta t\omega_{\mu\nu}u^{(\mu)}_{\max}} - 1\right) \\ &= \sum_{\mu,\nu=1}^{M,N} 1 - \frac{(\delta t\omega_{\mu\nu}u^{(\mu)}_{\max})^2}{6} - 1 + o((\delta t\omega_{\mu\nu}u^{(\mu)}_{\max})^2) \\ &= \sum_{\mu,\nu=1}^{M,N} -\frac{(\delta t\omega_{\mu\nu}u^{(\mu)}_{\max})^2}{6} + o((\delta t\omega_{\mu\nu}u^{(\mu)}_{\max})^2) \quad \text{(C3)}\end{aligned}$$

Now we take the exponential of both sides and obtain the result:

$$\begin{aligned}\tilde{\kappa}(\boldsymbol{\Omega}) &= \exp\left(\sum_{\mu,\nu=1}^{M,N} -\frac{(\delta t\omega_{\mu\nu}u^{(\mu)}_{\max})^2}{6} + o((\delta t\omega_{\mu\nu}u^{(\mu)}_{\max})^2\right) \\ &= 1 - \frac{1}{6}\sum_{\mu,\nu=1}^{M,N}(\delta t\omega_{\mu\nu}u^{(\mu)}_{\max})^2 + o((\delta t\omega_{\mu\nu}u^{(\mu)}_{\max})^2).\end{aligned}$$

In order to obtain the claim, we make use of the representation of the variance we found previously and substitute the asymptotic expansion:

$$\begin{aligned}\operatorname*{Var}_{\boldsymbol{U}\in\mathcal{C}}\left[\left(\prod_{p=1}^{P}\partial_{\mu_p\nu_p}\right)J_n\right] &= \delta t^{2P}\sum_{\boldsymbol{\Omega},\boldsymbol{\Omega'}\in(\mathcal{S}_n^\Delta)^N\setminus\{\boldsymbol{0}\}} c^*_{\boldsymbol{\Omega}}c_{\boldsymbol{\Omega'}}\left(\prod_{p=1}^{P}\omega_{\mu_p\nu_p}\omega'_{\mu_p\nu_p}\right)(\tilde{\kappa}(\boldsymbol{\Omega}-\boldsymbol{\Omega'}) - \tilde{\kappa}(\boldsymbol{\Omega})\tilde{\kappa}(\boldsymbol{\Omega'})) \\ &= \sum_{\boldsymbol{\Omega},\boldsymbol{\Omega'}\in(\mathcal{S}_n^\Delta)^N\setminus\{\boldsymbol{0}\}} c^*_{\boldsymbol{\Omega}}c_{\boldsymbol{\Omega'}}\left(\delta t^{2P}\prod_{p=1}^{P}\omega_{\mu_p\nu_p}\omega'_{\mu_p\nu_p}\right)\sum_{\mu,\nu=1}^{M,N}[\frac{(\delta tu^{(\mu)}_{\max})^2}{6}(\omega^2_{\mu\nu} + \omega'^2_{\mu\nu} - (\omega_{\mu\nu} - \omega'_{\mu\nu})^2) + \\ &\quad + o((\delta t\omega_{\mu\nu}u^{(\mu)}_{\max})^2) + o((\delta t\omega'_{\mu\nu}u^{(\mu)}_{\max})^2) + o((\delta t(\omega_{\mu\nu} - \omega'_{\mu\nu})u^{(\mu)}_{\max})^2))] \\ &\leq \left|\sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^\Delta)^N\setminus\{\boldsymbol{0}\}} c^*_{\boldsymbol{\Omega}}\right|\left|\sum_{\boldsymbol{\Omega'}\in(\mathcal{S}_n^\Delta)^N\setminus\{\boldsymbol{0}\}} c_{\boldsymbol{\Omega'}}\right| MN\left(\frac{((\delta t\omega_{\max})^{P+1}u_{\max})^2}{3} + o(((\delta t\omega_{\max})^{P+1}u_{\max})^2)\right).\end{aligned}$$

In order to upper bound the term containing the sum of Lie-Fourier coefficients, we first notice that

$$\mathbb{E}_{\boldsymbol{U}\in\mathbb{R}^{M\times N}}[J_n] = c_{\boldsymbol{0}}.$$

To show that, we fix $\forall\mu\ u^{(\mu)}_{\max} = u_{\max}$ inside Eq. (C2) and take the limit $u_{\max}\to\infty$:

$$\begin{aligned}&\lim_{u_{\max}\to\infty}\mathbb{E}_{\boldsymbol{U}\in\mathcal{C}_{u_{\max}}}[\prod_{p=1}^{P}\partial_{\mu_p\nu_p}J_n] \\ &= \lim_{u_{\max}\to\infty}\sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^\Delta)^N} c_{\boldsymbol{\Omega}}\left(\prod_{p=1}^{P}-i\delta t\omega_{\mu_p\nu_p}\right)\prod_{\mu,\nu=1}^{M,N}\delta_{\omega_{\mu\nu},0} \\ &= \begin{cases}c_{\boldsymbol{0}}, & P = 0,\\ 0, & P > 0.\end{cases}\end{aligned}$$

We also recall from the discussion in Lemma 5 that

$$\forall \boldsymbol{U}, n \quad \min_{|\psi\rangle\in\mathcal{H}} \langle\psi|\hat{O}|\psi\rangle \leq J_n(\boldsymbol{U}) \leq \max_{|\psi\rangle\in\mathcal{H}} \langle\psi|\hat{O}|\psi\rangle .$$

By linearity of the expectation value this implies that

$$\min_{|\psi\rangle\in\mathcal{H}} \langle\psi|\hat{O}|\psi\rangle \;\leq\; \mathbb{E}_{\boldsymbol{U}\in\mathbb{R}^{M\times N}}[J_n] \;\leq\; \max_{|\psi\rangle\in\mathcal{H}} \langle\psi|\hat{O}|\psi\rangle$$

from which easily follows that

$$\begin{aligned}\left|\sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^{\Delta})^N\setminus\{\boldsymbol{0}\}} c_{\boldsymbol{\Omega}}\right| &= |J_n(0) - \mathbb{E}_{\boldsymbol{U}\in\mathbb{R}^{M\times N}}[J_n]| \\ &\leq \max_{|\psi\rangle\in\mathcal{H}} \langle\psi|\hat{O}|\psi\rangle - \min_{|\psi\rangle\in\mathcal{H}} \langle\psi|\hat{O}|\psi\rangle = \Delta O.\end{aligned}$$

By substituting this result in the chain of inequalities for the variance we obtain the claim. By setting $\Delta O = 1$ and $M = 1$ we obtain the result from the main text. The result for J comes trivially from taking the limit on both sides, which is possible thanks to the uniform convergence of variance and expectation values we discussed previously in this proof. □

Lemma 9 (Variance over unbounded controls)**.** *For any $n \geq 1$, $P \geq 0$ and $1 \leq \mu_1, \dots, \mu_P \leq M$, $1 \leq \nu_1, \dots, \nu_P \leq N$ we have:*

$$\operatorname*{Var}_{\boldsymbol{U}\in\mathbb{R}^{M\times N}}\left[\left(\prod_{p=1}^{P}\partial_{\mu_p\nu_p}\right)J_n\right] = \delta t^{2P}\sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^{\Delta})^N\setminus\{\boldsymbol{0}\}}|c_{\boldsymbol{\Omega}}|^2\prod_{p=1}^{P}\omega_{\mu_p\nu_p}^2$$

Moreover, the following bounds hold:

$$\operatorname*{Var}_{\boldsymbol{U}\in\mathbb{R}^{M\times N}}\left[\left(\prod_{p=1}^{P}\partial_{\mu_p\nu_p}\right)J_n, J\right] \leq \left(\frac{(\delta t\omega_{\max})^P\Delta O}{2}\right)^2, \quad \sum_{\mu_1,\nu_1,\dots,\mu_P,\nu_P}\operatorname*{Var}_{\boldsymbol{U}\in\mathbb{R}^{M\times N}}\left[\left(\prod_{p=1}^{P}\partial_{\mu_p\nu_p}\right)J_n\right] \geq \frac{\Delta J^2\delta t^{2P}}{4\sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^{\Delta})^N\setminus\{\boldsymbol{0}\}}\frac{1}{||\boldsymbol{\Omega}||_2^{2P}}}$$

where we defined ΔJ as the maximum variation in J_n

$$\Delta J := \sup_{\boldsymbol{U}\in\mathbb{R}^{M\times N}} J_n(\boldsymbol{U}) - \inf_{\boldsymbol{U}\in\mathbb{R}^{M\times N}} J_n(\boldsymbol{U}).$$

Proof. Let us define $\mathcal{C}_{u_{\max}} := [-u_{\max}, u_{\max}]^{M\times N}$. The integrals that define the expectation values of a function F over $\mathbb{R}^{M\times N}$ can then be obtained as limits for $u_{\max}\to\infty$ of integrals over $\mathcal{C}_{u_{\max}}$:

$$\operatorname*{Var}_{\boldsymbol{U}\in\mathbb{R}^{M\times N}}[F] = \lim_{u_{\max}\to\infty}\mathbb{E}_{\boldsymbol{U}\in\mathcal{C}}[|F|^2] - \lim_{u_{\max}\to\infty}|\mathbb{E}_{\boldsymbol{U}\in\mathcal{C}}[F]|^2. \tag{C4}$$

Following the steps already take in the proof of Lemma 8, we start with the expectation value of J_n, which was there shown to give rise to

$$\lim_{u_{\max}\to\infty}\mathbb{E}_{\boldsymbol{U}\in\mathcal{C}_{u_{\max}}}\left[\left(\prod_{p=1}^{P}\partial_{\mu_p\nu_p}\right)J_n\right] = \begin{cases} c_{\boldsymbol{0}}, & P=0,\\ 0, & P>0.\end{cases}$$

We can then do the same for the square modulus:

$$\begin{aligned}&\lim_{u_{\max}\to\infty}\mathbb{E}_{\boldsymbol{U}\in\mathcal{C}_{u_{\max}}}\left[\left|\left(\prod_{p=1}^{P}\partial_{\mu_p\nu_p}\right)J_n\right|^2\right] \\ &= \delta t^{2P}\sum_{\boldsymbol{\Omega},\boldsymbol{\Omega'}\in(\mathcal{S}_n^{\Delta})^N} c^*_{\boldsymbol{\Omega}}c_{\boldsymbol{\Omega'}}\Big(\prod_{p=1}^{P}\omega_{\mu_p\nu_p}\omega'_{\mu_p\nu_p}\Big)\prod_{\mu,\nu=1}^{M,N}\delta_{\omega_{\mu\nu},\omega'_{\mu\nu}} \\ &= \delta t^{2P}\sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^{\Delta})^N}|c_{\boldsymbol{\Omega}}|^2\prod_{p=1}^{P}\omega_{\mu_p\nu_p}^2.\end{aligned}$$

By substituting these results inside Eq. (C4) and by noticing that

$$\tilde{\kappa}(\boldsymbol{0}-\boldsymbol{\Omega'}) - \tilde{\kappa}(\boldsymbol{0})\tilde{\kappa}(\boldsymbol{\Omega'}) = 0$$

we can restrict all sums to non-zero frequencies in $(\mathcal{S}_n^{\Delta})^N\setminus\{\boldsymbol{0}\}$, which gives us the result in the claim.

Concerning the upper bound in the claim, it is a direct consequence of the corresponding bound for finite $u_{\max}$ that we proved in Lemma 8. Since the bound holds for

any $u_{\max}$ it also holds for $u_{\max} \to \infty$ because of the continuity property of the limit.

The lower bound is not as trivial to derive, and only holds for J_n (and not J, since the limits in $u_{\max}$ and n do not necessarily commute). Using the representation of the variance we derived previously in this Lemma, we have

$$\begin{aligned}\sum_{\mu_1,\nu_1,\dots,\mu_P,\nu_P} \operatorname*{Var}_{\boldsymbol{U}\in\mathbb{R}^{M\times N}} \left[\left(\prod_{p=1}^{P}\partial_{\mu_p\nu_p}\right) J_n\right] \\ = \delta t^{2P} \sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^\Delta)^N\setminus\{\mathbf{0}\}} |c_{\boldsymbol{\Omega}}|^2 \prod_{p=1}^{P}\sum_{\mu_p,\nu_p=1}^{M,N} \omega_{\mu_p\nu_p}^2 \\ = \delta t^{2P} \sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^\Delta)^N\setminus\{\mathbf{0}\}} |c_{\boldsymbol{\Omega}}|^2 ||\boldsymbol{\Omega}||_2^{2P}\end{aligned}$$

where the $\boldsymbol{\Omega} = \mathbf{0}$ term can be omitted from the sum as it does not contribute to the variance. We then notice the following:

$$|J_n(\boldsymbol{U}) - c_{\mathbf{0}}| = \left|\sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^\Delta)^N\setminus\{\mathbf{0}\}} c_{\boldsymbol{\Omega}} e^{-i\delta t \mathrm{Tr}(\boldsymbol{\Omega}^T\boldsymbol{U})}\right| \leq \sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^\Delta)^N\setminus\{\mathbf{0}\}} |c_{\boldsymbol{\Omega}}|$$

therefore, by considering separately the two inequalities given by the absolute value, namely

$$-\sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^\Delta)^N\setminus\{\mathbf{0}\}} |c_{\boldsymbol{\Omega}}| \leq J_n(\boldsymbol{U}) - c_{\mathbf{0}} \leq \sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^\Delta)^N\setminus\{\mathbf{0}\}} |c_{\boldsymbol{\Omega}}|$$

and noting they are true also for any $\boldsymbol{U} \in \mathbb{R}^{M\times N}$, they will hold respectively for the infimum and the supremum:

$$\begin{aligned}\sup_{\boldsymbol{U}\in\mathbb{R}^{M\times N}} J_n(\boldsymbol{U}) - c_{\mathbf{0}} \leq \sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^\Delta)^N\setminus\{\mathbf{0}\}} |c_{\boldsymbol{\Omega}}| \\ \inf_{\boldsymbol{U}\in\mathbb{R}^{M\times N}} J_n(\boldsymbol{U}) - c_{\mathbf{0}} \geq -\sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^\Delta)^N\setminus\{\mathbf{0}\}} |c_{\boldsymbol{\Omega}}|.\end{aligned}$$

By subtracting the second equation from the first one we obtain that

$$\sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^\Delta)^N\setminus\{\mathbf{0}\}} |c_{\boldsymbol{\Omega}}| \geq \frac{\Delta J}{2} \tag{C5}$$

where we defined the maximum variation ΔJ as in the claim. The lower bound can then be obtained as the global minimum of the function

$$V(c_{\boldsymbol{\Omega}}, c_{\boldsymbol{\Omega}}^*) := \delta t^{2P} \sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^\Delta)^N\setminus\{\mathbf{0}\}} |c_{\boldsymbol{\Omega}}|^2 ||\boldsymbol{\Omega}||_2^{2P}$$

over $c_{\boldsymbol{\Omega}}$ subject to the constraint given by Eq. (C5).

Let us first study the corresponding problem with equality constraints

$$\sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^\Delta)^N\setminus\{\mathbf{0}\}} |c_{\boldsymbol{\Omega}}| = \frac{\Delta J}{2} \tag{C6}$$

We can use the method of Lagrange multipliers by finding the stationary point of the cost function L

$$L(c_{\boldsymbol{\Omega}}, c_{\boldsymbol{\Omega}}^*) = \sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^\Delta)^N\setminus\{\mathbf{0}\}} \delta t^{2P} |c_{\boldsymbol{\Omega}}|^2 ||\boldsymbol{\Omega}||_2^{2P} - \lambda |c_{\boldsymbol{\Omega}}|.$$

Since L is not differentiable whenever any $c_{\boldsymbol{\Omega}} = 0$, we will also have to examine these irregular points along with the stationary points. Thus, the candidate minima are given by

$$\begin{aligned}\forall \boldsymbol{\Omega} \in (\mathcal{S}_n^\Delta)^N\setminus\{\mathbf{0}\}\ \frac{\partial L}{\partial c_{\boldsymbol{\Omega}}^*} = c_{\boldsymbol{\Omega}}\left(\delta t^{2P}||\boldsymbol{\Omega}||_2^{2P} - \frac{\lambda}{2|c_{\boldsymbol{\Omega}}|}\right) = 0 \\ \implies \forall \boldsymbol{\Omega} \in (\mathcal{S}_n^\Delta)^N\setminus\{\mathbf{0}\}\ c_{\boldsymbol{\Omega}} = 0 \vee |c_{\boldsymbol{\Omega}}| = \frac{\lambda}{2\delta t^{2P}||\boldsymbol{\Omega}||_2^{2P}}.\end{aligned}$$

While almost any subset of frequencies can have zero coefficients, not all of them can be be zero, otherwise Eq. (C6) would not be satisfied. We can then define $\mathcal{F} \neq \emptyset$ as the set of frequencies with non-zero coefficients $c_{\boldsymbol{\Omega}} \neq 0$, and fix λ using the equality Eq. (C6) as

$$\lambda = \frac{\Delta J \delta t^{2P}}{\sum_{\boldsymbol{\Omega}\in\mathcal{F}\setminus\{\mathbf{0}\}} \frac{1}{||\boldsymbol{\Omega}||_2^{2P}}}. \tag{C7}$$

We can now substitute this result in the expression for the variance:

$$\begin{aligned}\sum_{\mu_1,\nu_1,\dots,\mu_P,\nu_P} \operatorname*{Var}_{\boldsymbol{U}\in\mathbb{R}^{M\times N}} \left[\left(\prod_{p=1}^{P}\partial_{\mu_p\nu_p}\right) J_n\right] \\ = \delta t^{2P} \sum_{\boldsymbol{\Omega}\in\mathcal{F}\setminus\{\mathbf{0}\}} |c_{\boldsymbol{\Omega}}|^2 ||\boldsymbol{\Omega}||_2^{2P} \\ = \frac{\lambda^2}{4\delta t^{2P}} \sum_{\boldsymbol{\Omega}\in\mathcal{F}\setminus\{\mathbf{0}\}} \frac{1}{||\boldsymbol{\Omega}||_2^{2P}} = \frac{\Delta J^2 \delta t^{2P}}{4\sum_{\boldsymbol{\Omega}\in\mathcal{F}\setminus\{\mathbf{0}\}} \frac{1}{||\boldsymbol{\Omega}||_2^{2P}}},\end{aligned}$$

where in the last step we made use of Eq. (C7). As the denominator is a sum of positive quantities, it is easy to see that the minimum over the equality constraint Eq. (C6) is obtained when all frequencies are non-zero $\mathcal{F} = (\mathcal{S}_n^\Delta)^N$. On the other hand the maximum is obtained when all frequencies are zero except the largest one, for which $||\boldsymbol{\Omega}||_2^2 = \omega_{max}^{2MN}$, and $V_{max} = \Delta J^2 \delta t^{2P} \omega_{max}^{2MNP}/4$. To see that the minimum over the equality constraint Eq. (C6) is also the minimum over the inequality constraint Eq. (C5), we can reason as follows. Let us consider the following sets

$$\begin{aligned}\mathcal{E}_- = \{c_{\boldsymbol{\Omega}} \in \mathbb{C}^{n_\Delta^N - 1} \mid \sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^\Delta)^N\setminus\{\mathbf{0}\}} |c_{\boldsymbol{\Omega}}| \geq \frac{\Delta J}{2}\}, \\ \mathcal{E}_+ = \{c_{\boldsymbol{\Omega}} \in \mathbb{C}^{n_\Delta^N - 1} \mid \delta t^{2P} \sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^\Delta)^N\setminus\{\mathbf{0}\}} |c_{\boldsymbol{\Omega}}|^2 ||\boldsymbol{\Omega}||_2^{2P} \leq M\}.\end{aligned}$$

where we used the notation $n_\Delta = \#\mathcal{S}_n^\Delta$. Since these two sets are solutions to continuous inequalities they are both closed, and $\mathcal{E}_+$ is bounded as $\mathcal{E}_+ \subset \mathcal{E}_b$, with

$$\mathcal{E}_b = \{c_{\boldsymbol{\Omega}} \in \mathbb{C}^{n_\Delta^N - 1} \mid \sum_{\boldsymbol{\Omega}\in(\mathcal{S}_n^\Delta)^N\setminus\{\mathbf{0}\}} |c_{\boldsymbol{\Omega}}|^2 \leq \frac{M}{(\delta t \omega_{min})^{2P}}\}.$$

It follows that $\mathcal{E} = \mathcal{E}_+ \cap \mathcal{E}_-$ is a closed and bounded set. As such, it admits a minimum (and a maximum) of the continuous function V (this is a well known consequence of the Bolzano-Weierstrass theorem). Now, we have

$$\mathcal{E}_- = (\mathcal{E}_- \cap \mathcal{E}_+) \cup (\mathcal{E}_- \cap (\mathbb{C}^{n_\Delta^N - 1} \setminus \mathcal{E}_+)) = \mathcal{E} \cup \mathcal{E}'.$$

Since by definition $V > M$ over $\mathcal{E}'$, V cannot assume its minimum value inside $\mathcal{E}'$ (unless $\mathcal{E} = \emptyset$), since it assumes smaller values $V \leq M$ inside $\mathcal{E}$. The set $\mathcal{E}$ can be further decomposed into $\mathcal{E} = \mathcal{E}_{int} \cup \partial\mathcal{E}_- \cup \partial\mathcal{E}_+$, with

$$\partial\mathcal{E}_- = \{c_{\boldsymbol{\Omega}} \in \mathbb{C}^{n_\Delta^N - 1} \mid \sum_{\boldsymbol{\Omega} \in (\mathcal{S}_n^\Delta)^N \setminus \{\mathbf{0}\}} |c_{\boldsymbol{\Omega}}| = \frac{\Delta J}{2}\}$$

$$\partial\mathcal{E}_+ = \{c_{\boldsymbol{\Omega}} \in \mathbb{C}^{n_\Delta^N - 1} \mid \delta t^{2P} \sum_{\boldsymbol{\Omega} \in (\mathcal{S}_n^\Delta)^N \setminus \{\mathbf{0}\}} |c_{\boldsymbol{\Omega}}|^2 ||\boldsymbol{\Omega}||_2^{2P} = M\}$$

$$\mathcal{E}_{int} = \{c_{\boldsymbol{\Omega}} \in \mathbb{C}^{n_\Delta^N - 1} \mid \sum_{\boldsymbol{\Omega} \in (\mathcal{S}_n^\Delta)^N \setminus \{\mathbf{0}\}} |c_{\boldsymbol{\Omega}}| > \frac{\Delta J}{2} \wedge$$
$$\delta t^{2P} \sum_{\boldsymbol{\Omega} \in (\mathcal{S}_n^\Delta)^N \setminus \{\mathbf{0}\}} |c_{\boldsymbol{\Omega}}|^2 ||\boldsymbol{\Omega}||_2^{2P} < M\}$$

If $M > V_{max}$, V cannot assume its minimum value inside $\partial\mathcal{E}_+$, as then there would be points on $\partial\mathcal{E}_-$ where V would be smaller. That cannot even happen at an internal point belonging to the open set $\mathcal{E}_{int}$, as then it would be a stationary point $\boldsymbol{\nabla} V = \mathbf{0}$, while the only stationary point of V is $c_{\boldsymbol{\Omega}} = 0$, which does not belong to $\mathcal{E}$. The only remaining option is that V assumes its minimum value on $\partial\mathcal{E}_-$. This implies that the stationary point we have found with the method of Lagrange multipliers is the minimum of V over $\mathcal{E}_-$, that is the minimum of the variance subject to Eq. (C5). The results cited in the main text follow by fixing $M = 1$, $\Delta O = 1$. □

Appendix D: Proofs valid only for one control M=1

Lemma 2 (Boundedness of the coefficients)**.**

$$\exists r \in \mathbb{R},\ s.t.\ \ \forall n,\ \sum_{\boldsymbol{\omega} \in (\mathcal{S}_n^\Delta)^N} |c_{\boldsymbol{\omega}}(n, \delta t)| \leq r$$

Proof. We start by noticing that the following inequality holds:

$$\forall i, j\ \ 0 \leq |e^{\frac{\hat{A}}{n}} - \hat{I}|_{ij} \leq ||e^{\frac{\hat{A}}{n}} - \hat{I}||_\infty \leq \frac{||\hat{A}||_\infty}{n} e^{\frac{||\hat{A}||_\infty}{n}},$$

where we employed Eq. (A1). So, since $||\hat{V}||_\infty = 1$, by defining

$$R(n) := \frac{||\delta t \hat{H}_d||_\infty}{n} e^{\frac{||\delta t \hat{H}_d||_\infty}{n}},$$

we can write an inequality for W_{ij}:

$|W_{ij}| = |\delta_{ij} - \delta_{ij} + W_{ij}| \leq \delta_{ij} + |\hat{W} - \hat{I}|_{ij} \leq \delta_{ij} + R(n)$. Then, we plug this inequality into the sum of the moduli of the Fourier coefficients:

$$\sum_{ik\boldsymbol{j}} |\tilde{A}_{ik}^{\boldsymbol{j}}| \leq \sum_{ik\boldsymbol{j}} |W_{ij_n}| \cdots |W_{j_2 j_1}| |\delta_{j_1 k}|$$
$$\leq \sum_{ij_1 \ldots j_n} (\delta_{ij_n} + R(n)) \cdots (\delta_{j_2 j_1} + R(n)) =$$
$$= D \sum_{m=0}^{n} \binom{n}{m} R(n)^m D^m$$
$$= D \sum_{m=0}^{n} \binom{n}{m} \left(\frac{D\delta t ||\hat{H}_d||_\infty}{n} \right)^m e^{\frac{m\delta t ||\hat{H}_d||_\infty}{n}} =$$
$$D \left(1 + \frac{D\delta t ||\hat{H}_d||_\infty}{n} e^{\frac{\delta t ||\hat{H}_d||_\infty}{n}} \right)^n \xrightarrow{n \to \infty} D e^{D\delta t ||\hat{H}_d||_\infty}.$$

The passage from the sum over $\boldsymbol{j}$ to the sum over m is performed by noticing that the expansion of the product in the second line gives rise to terms which consists in $m+1$ independent chains of Kronecker deltas (which can also be trivial), each one of which gives rise to a d factor when summed up over its indices.

Because of this inequalities, the positive sequence $\sum_{ik\boldsymbol{j}} |\tilde{A}_{ik}^{\boldsymbol{j}}|$ is upper-bounded by a (positive) converging sequence. But a positive converging sequence has a finite maximum A. This result implies the boundedness of the $\tilde{B}_{ik}^{\omega}$ coefficients

$$\sum_{ik} \sum_{\omega \in \mathcal{S}_n} |\tilde{B}_{ik}^{\omega}| = \sum_{ik} \sum_{\omega \in \mathcal{S}_n} |\sum_{\boldsymbol{j} \in \mathcal{J}_\omega} \tilde{A}_{ik}^{\boldsymbol{j}}| \leq$$
$$\leq \sum_{ik} \sum_{\omega \in \mathcal{S}_n} \sum_{\boldsymbol{j} \in \mathcal{J}_\omega} |\tilde{A}_{ik}^{\boldsymbol{j}}| = \sum_{ik\boldsymbol{j}} |\tilde{A}_{ik}^{\boldsymbol{j}}| \leq A,$$

and of the coefficients of the fidelity itself:

$$\sum_{\boldsymbol{\omega}\in(\mathcal{S}_n^\Delta)^N}|c_{\boldsymbol{\omega}}| = \sum_{\boldsymbol{\omega}\in(\mathcal{S}_n^\Delta)^N}\left|\sum_{\boldsymbol{\omega'}\boldsymbol{\omega''}\in\mathcal{F}}\delta_{\omega,\omega'-\omega''}\,\langle\tilde{\psi}|\,\tilde{B}^{\boldsymbol{\omega''}\dagger}\,|\tilde{\chi}\rangle\,\langle\tilde{\chi}|\,\tilde{B}^{\boldsymbol{\omega'}}\,|\tilde{\psi}\rangle\right| \leq \sum_{\boldsymbol{\omega}\in(\mathcal{S}_n^\Delta)^N}\sum_{\boldsymbol{\omega'}\boldsymbol{\omega''}\in\mathcal{F}}\delta_{\omega,\omega'-\omega''}|\,\langle\tilde{\psi}|\,\tilde{B}^{\boldsymbol{\omega''}\dagger}\,|\tilde{\chi}\rangle\,\langle\tilde{\chi}|\,\tilde{B}^{\boldsymbol{\omega'}}\,|\tilde{\psi}\rangle\,|$$

$$= \left(\sum_{\boldsymbol{\omega'}\in\mathcal{F}}|\,\langle\tilde{\chi}|\,\tilde{B}^{\boldsymbol{\omega'}}\,|\tilde{\psi}\rangle\,|\right)^2 \leq \left(\underbrace{\max_{ij}|\tilde{\chi}_i||\tilde{\psi}_j|}_{\leq 1}\sum_{ij}\sum_{\boldsymbol{\omega'}\in\mathcal{F}}|\tilde{B}^{\boldsymbol{\Omega'}}|_{ij}\right)^2 \leq \left(\sum_{ij}\sum_{\omega^{(1)}\ldots\omega^{(N)}}|\tilde{B}^{\omega^{(N)}}\cdots\tilde{B}^{\omega^{(1)}}|_{ij}\right)^2 \leq$$

$$\leq \left(\sum_{ij}\sum_{\omega\in\mathcal{S}_n}|\tilde{B}^{\omega}_{ij}|\right)^{2N} \leq A^{2N}.$$

which concludes the proof with $r = A^{2N}$. □

Lemma 4 (Symmetries and selection rules)**.** *See main text.*

Proof. Since $[\hat{H}_d, \hat{\Gamma}] = [\hat{H}_c, \hat{\Gamma}] = 0$, we can choose $\hat{V}$ in Eq. (6) so that both $\hat{\Lambda} = \hat{V}\hat{H}_c\hat{V}^\dagger$ and $\hat{\tilde{\Gamma}} = \hat{V}\hat{\Gamma}\hat{V}^\dagger$ have diagonal matrix representations:

$$\langle i|\,\hat{\tilde{\Gamma}}\,|k\rangle = \delta_{ik}\gamma_{g(i)}$$

where $g(i)$ maps the index $i \in \{1,\ldots,D\}$ to the corresponding symmetry sector index $g \in \{1,\ldots,G\}$. But since also $[\hat{\tilde{H}}_d, \hat{\tilde{\Gamma}}] = 0$, then

$$\langle i|\,\hat{\tilde{H}}_d\,|k\rangle = \delta_{g(i)g(k)}[\hat{\tilde{H}}_d^{(g(i))}]_{ik}$$

must be block diagonal, with blocks $\hat{\tilde{H}}_d^{(g)}$ corresponding to the degenerate eigenspaces of $\hat{\tilde{\Gamma}}$. We name $\mathcal{I}_g$ the set of index values belonging to the g-th block:

$$\mathcal{I}_g = \{i \in \{1,\ldots,D\} \mid g(i) = g\}$$

This in turn implies that $\hat{W} = \hat{V}e^{-\frac{i\delta t}{n}\hat{H}_d}\hat{V}^\dagger = e^{-\frac{i\delta t}{n}\hat{\tilde{H}}_d}$ has block diagonal matrix elements in the same fashion. But then

$$\tilde{A}^{\boldsymbol{j}}_{ik} = W_{ij_n}\cdots W_{j_2j_1}\delta_{j_1k} \neq 0$$

only when all the indices $i, k, j_1, \ldots, j_n \in \mathcal{I}_g$ belong to the same block. Correspondingly, if we consider products like

$$\hat{\tilde{A}}^{\boldsymbol{J}} := \hat{\tilde{A}}^{\boldsymbol{j}^{(N)}}\cdots\hat{\tilde{A}}^{\boldsymbol{j}^{(1)}},$$

where $\boldsymbol{J}$ is the multiindex

$$\boldsymbol{J} = \begin{pmatrix}\boldsymbol{j}^{(1)} & \ldots & \boldsymbol{j}^{(N)}\end{pmatrix},$$

we have $\tilde{A}^{\boldsymbol{J}}_{ik} \neq 0$ only when $i, k \in \mathcal{I}_g$ and $\boldsymbol{j}^{(1)}, \ldots, \boldsymbol{j}^{(N)} \in \mathcal{I}_g^n$ belong to the same block. This determines selection rules for the frequencies in the spectrum $\mathcal{S}_n$ of the time step operator $\hat{U}_n(u)$

$$[\hat{\tilde{U}}_n(u)]_{ik} = \sum_{\boldsymbol{j}\in[D]^n} e^{-i\delta t u\omega_{\boldsymbol{j}}}\tilde{A}^{\boldsymbol{j}}_{ik} = \sum_{\omega\in\mathcal{S}_n} e^{-i\delta t u\omega}\tilde{B}^{\omega}_{ik},$$

since then the only combinations of eigenvalues that are allowed are the ones of eigenvalues within the same symmetry sector, so that the spectrum becomes

$${}^{(\Gamma)}\mathcal{S}_n = \bigcup_{g=1}^{G}\mathcal{S}_n^{(g)}$$

$$\mathcal{S}_n^{(g)} = \{\omega_{\boldsymbol{j}} = \frac{1}{n}(\lambda_{j_1} + \cdots + \lambda_{j_n}) \mid \boldsymbol{j} \in \mathcal{I}_g^n\}.$$

Concerning $\tilde{B}^{\omega}_{ik}$, in general we can only say that they are block diagonal in the indices i, k in the same way as W_{ik} and $\tilde{A}^{\boldsymbol{j}}_{ik}$, but the frequency ω does not select uniquely the block.

For similar reasons, when it comes to the full time evolution operator $\tilde{U}_n(\boldsymbol{u})$

$$\hat{\tilde{U}}_n(\boldsymbol{u}) = \sum_{\boldsymbol{j}^{(1)}\ldots\boldsymbol{j}^{(N)}} e^{-i\delta t\sum_\nu u_\nu\omega_{\boldsymbol{j}^{(\nu)}}}\hat{\tilde{A}}^{\boldsymbol{j}^{(N)}}\cdots\hat{\tilde{A}}^{\boldsymbol{j}^{(1)}}$$

$$= \sum_{\boldsymbol{\omega}\in{}^{(\Gamma,\mathrm{N})}\mathcal{S}_n} e^{-i\delta t\boldsymbol{\omega}\cdot\boldsymbol{u}}\hat{B}^{\boldsymbol{\omega}}$$

the only combinations of frequencies for which the coefficients are non-zero are in the set

$${}^{(\Gamma,\mathrm{N})}\mathcal{S}_n = \bigcup_{g=1}^{G}(\mathcal{S}_n^{(g)})^N$$

We can now use this information to write a decomposition for the Lie-Fourier representation of the matrix

element

$$\begin{aligned}\langle\chi|\,\hat{U}_n(\boldsymbol{u})\,|\psi\rangle &= \langle\tilde{\chi}|\,\hat{\tilde{U}}_n(\boldsymbol{u})\,|\tilde{\psi}\rangle \\ &= \sum_{i,k}\sum_{\boldsymbol{J}} e^{-i\delta t\sum_\nu u_\nu\omega_{j^{(\nu)}}}\tilde{\chi}_i\tilde{A}^{\boldsymbol{J}}_{ik}\tilde{\psi}_k \\ &= \sum_{g=1}^{G}\sum_{i,k\in\mathcal{I}_g}\sum_{\boldsymbol{J}\in\mathcal{I}_g^{n\times N}} e^{-i\delta t\sum_\nu u_\nu\omega_{j^{(\nu)}}}\tilde{\chi}_i\tilde{A}^{\boldsymbol{J}}_{ik}\tilde{\psi}_k \\ &= \sum_{g=1}^{G}\sum_{i,k\in\mathcal{I}_g}\sum_{\boldsymbol{\omega}\in(\mathcal{S}_n^{(g)})^N} e^{-i\delta t\boldsymbol{\omega}\cdot\boldsymbol{u}}\tilde{\chi}_i\tilde{B}^{\boldsymbol{\omega}}_{ik}\tilde{\psi}_k \\ &= \sum_{g=1}^{G}\sum_{\boldsymbol{\omega}\in(\mathcal{S}_n^{(g)})^N} e^{-i\delta t\boldsymbol{\omega}\cdot\boldsymbol{u}}b^{(g)}_{\boldsymbol{\omega}}\end{aligned}$$

where $b^{(g)}_{\boldsymbol{\omega}}\neq 0$ only if $g\in\mathcal{G}$, defined as

$$\mathcal{G}=\{g=1,\dots,G\mid\exists i,k\in\mathcal{I}_g,\tilde{\chi}_i\tilde{\psi}_k\neq 0\},\tag{D1}$$

that is, $b^{(g)}_{\boldsymbol{\omega}}$ is non-zero only if both initial and target state have a non-zero overlap with the eigenstates generating the g-th symmetry sector. In fact we have

$$\begin{aligned}\tilde{\psi}_k=\langle k|\,\hat{V}\,|\psi\rangle=\langle k|\sum_{i=1}^{D}|i\rangle\,\langle\gamma_{g(i)},\lambda_i^{(g(i))}|\psi\rangle= \\ =\langle\gamma_{g(k)},\lambda_k^{(g(k))}|\psi\rangle\neq 0\end{aligned}$$

and similarly for $\tilde{\chi}$. But this is equivalent to asking $\hat{P}_g\,|\psi\rangle\neq 0$ and $\hat{P}_g\,|\chi\rangle\neq 0$ as in the claim. Therefore, we can conclude

$$\langle\chi|\,\hat{U}_n(\boldsymbol{u})\,|\psi\rangle=\sum_{g\in\mathcal{G}}\sum_{\boldsymbol{\omega}\in(\mathcal{S}_n^{(g)})^N}e^{-i\delta t\boldsymbol{\omega}\cdot\boldsymbol{u}}b^{(g)}_{\boldsymbol{\omega}}=\sum_{\boldsymbol{\omega}\in{}^{(\Gamma,\mathrm{N})}\mathcal{S}_n}e^{-i\delta t\boldsymbol{\omega}\cdot\boldsymbol{u}}b_{\boldsymbol{\omega}}.$$

The spectrum of the fidelity is constructed as usual by taking all possible frequency differences, also across different symmetry sectors, by means of Eq. (8), giving rise to the frequency set ${}^{(\Gamma,\mathrm{N})}\mathcal{S}_n^{\Delta}$. □

Appendix E: Details about Ising model landscapes

1. Computing Lie-Fourier coefficients with the Discrete Fourier Transform (DFT)

When the single-timestep spectrum $\mathcal{S}_n^{\Delta}$ of the Lie-Fourier representation J_n of the landscape only contains evenly spaced frequencies, that is

$$\mathcal{S}_n^{\Delta}=\{\omega=\omega_{\max}\frac{k}{k_{\max}}\mid k=-k_{\max},\dots,k_{\max}\},$$

its coefficients can be computed numerically by means of the DFT. In fact, we can write

$$J_n(\boldsymbol{u})=\sum_{\boldsymbol{k}=-\boldsymbol{k}_{\max}}^{\boldsymbol{k}_{\max}}c_{\boldsymbol{k}}e^{-i\frac{\delta t\omega_{\max}}{k_{\max}}\boldsymbol{k}\cdot\boldsymbol{u}},\tag{E1}$$

where we used the shorthand $c_{\boldsymbol{k}}=c_{\boldsymbol{\omega}(\boldsymbol{k})}$ and the multi-index notation for $\boldsymbol{k}\in\mathbb{Z}^N$. Let us now multiply both sides of Eq. (E1) by $\exp(-i\varphi(\boldsymbol{u}))$, with

$$\varphi(\boldsymbol{u})=\delta t\omega_{\max}(\boldsymbol{u}\cdot\boldsymbol{1})$$

and evaluate the expression on a square hyperlattice given by

$$\boldsymbol{u_j}=\frac{2\pi\boldsymbol{j}}{\delta t\omega_{\max}}\frac{k_{\max}}{n_\Delta},\tag{E2}$$

where $\boldsymbol{j}\in\mathbb{Z}^N$ is an integer multiindex whose elements each range from 0 to $n_\Delta-1$. Then, we obtain the left hand side expressed as the N-dimensional DFT of the coefficients

$$e^{-i\varphi(\boldsymbol{u_j})}J_n(\boldsymbol{u_j})=\sum_{\boldsymbol{k}=\boldsymbol{0}}^{\boldsymbol{n_\Delta-1}}c_{\boldsymbol{k}-\boldsymbol{k}_{\max}}e^{-i\frac{2\pi}{n_\Delta}\boldsymbol{k}\cdot\boldsymbol{j}}$$

which means that the coefficients themselves can be obtained using the inverse DFT:

$$c_{\boldsymbol{k}-\boldsymbol{k}_{\max}}=\frac{1}{n_\Delta^N}\sum_{\boldsymbol{j}=\boldsymbol{0}}^{\boldsymbol{n_\Delta-1}}e^{-i\varphi(\boldsymbol{u_j})}J_n(\boldsymbol{u_j})e^{i\frac{2\pi}{n_\Delta}\boldsymbol{k}\cdot\boldsymbol{j}}.$$

We note that all the Fourier components, and hence the functions J_n, are invariant under the transformation $\boldsymbol{u}\mapsto\boldsymbol{u}+2\pi k_{\max}(\delta t\omega_{\max})^{-1}\boldsymbol{m}$, with $\boldsymbol{m}\in\mathbb{Z}^N$. This means that they are all periodic with period $T_n=2\pi k_{\max}(\delta t\omega_{\max})^{-1}$ (or an integer multiple thereof) along each dimension. Because of the orthogonality of these Fourier components in the $L_2[0,T_n]^N$ sense, we can be sure that the coefficients given by the inverse DFT and the ones from the Lie-Fourier expansion are the same.

The DFT and its inverse can be computed efficiently for a sample of M points using the Fast Fourier Transform algorithm in $O(M\log M)$ flops, but the sample defined by Eq. (E2) contains $M=n_\Delta^N$ points, which prevents us from pushing the computation to large numbers of time steps N.

2. Time symmetry in the real case

Whenever both the Hamiltonian $\hat{H}$ and the states $|\psi\rangle,|\chi\rangle$ defining the state transfer problem only have real matrix elements and overlaps in the same basis, the landscape

$$J(\delta t;\boldsymbol{u})=|\,\langle\chi|\,\hat{U}(\delta t;u_N)\cdots\hat{U}(\delta t;u_1)\,|\psi\rangle\,|^2,$$

where $\hat{U}(\delta t;u)=\exp\Big(-i\delta t\hat{H}(u)\Big)$ is symmetric under change of sign of time

$$J(\delta t)=J(-\delta t).$$

In order to see this, let us first express the overlap in index form:

$$\langle\chi|\,\hat{U}(\delta t;u_N)\cdots\hat{U}(\delta t;u_1)\,|\psi\rangle = \sum_{\boldsymbol{j}}\chi_{j_1}U_{j_1j_2}(\delta t;u_1)\cdots U_{j_Nj_{N+1}}(\delta t;u_1)\psi_{j_{N+1}}$$

While working in this basis, the Hamiltonians at the different timesteps are going to be real and symmetric matrices, hence they can be diagonalized by means of real orthogonal matrices $V\in\mathbb{R}^{D\times D}, V^TV=\mathbb{1}$:

$$U_{ik}(\delta t;u)=V_{ij}(u)e^{-i\delta t\Lambda_j(u)}V^T_{jk}(u).$$

By expressing this way every unitary appearing in the overlap, it is easy to see that it can be expressed as a real combination of Fourier components (which are not the same as the ones in the Lie-Fourier representation in the main text):

$$\begin{aligned}\langle\chi|\,\hat{U}(\delta t;u_N)\cdots\hat{U}(\delta t;u_1)\,|\psi\rangle &= \\ &= \sum_{\boldsymbol{j},\boldsymbol{k}}\chi_{j_1}V_{j_1k_1}(u)e^{-i\delta t\Lambda_{k_1}(u)}V^T_{k_1j_2}(u)\cdots \\ \cdots V_{j_Nk_N}(u)e^{-i\delta t\Lambda_{k_N}(u)}V^T_{k_Nj_{N+1}}(u)\psi_{j_{N+1}} &= \sum_{\boldsymbol{j},\boldsymbol{k}}r_{\boldsymbol{jk}}e^{-i\delta t\omega_{\boldsymbol{k}}}.\end{aligned}$$

But then the square absolute value is time symmetric as expected:

$$\begin{aligned}J(\delta t)=\left|\sum_{\boldsymbol{j},\boldsymbol{k}}r_{\boldsymbol{jk}}e^{-i\delta t\omega_{\boldsymbol{k}}}\right|^2 &= \left|\left(\sum_{\boldsymbol{j},\boldsymbol{k}}r_{\boldsymbol{jk}}e^{-i\delta t\omega_{\boldsymbol{k}}}\right)^*\right|^2 \\ &\left|\sum_{\boldsymbol{j},\boldsymbol{k}}r_{\boldsymbol{jk}}e^{+i\delta t\omega_{\boldsymbol{k}}}\right|^2 = J(-\delta t).\end{aligned}$$

[1] M. Bukov, A. G. R. Day, P. Weinberg, A. Polkovnikov, P. Mehta, and D. Sels, Phys. Rev. A **97**, 052114 (2018).
[2] A. G. R. Day, M. Bukov, P. Weinberg, P. Mehta, and D. Sels, Phys. Rev. Lett. **122**, 020601 (2019).
[3] M. Dalgaard, F. Motzoi, and J. Sherson, Phys. Rev. A **105**, 012402 (2022).
[4] N. Beato, P. Patil, and M. Bukov, Towards a theory of phase transitions in quantum control landscapes (2024), arXiv:2408.11110 [quant-ph].
[5] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Nature Communications **9**, 4812 (2018).
[6] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, PRX Quantum **3**, 010313 (2022).
[7] M. Larocca, P. Czarnik, K. Sharma, G. Muraleedharan, P. J. Coles, and M. Cerezo, Quantum **6**, 824 (2022).
[8] H. A. Rabitz, M. M. Hsieh, and C. M. Rosenthal, Science **303**, 1998 (2004), https://www.science.org/doi/pdf/10.1126/science.1093649.
[9] A. N. Pechen and D. J. Tannor, Phys. Rev. Lett. **106**, 120402 (2011).
[10] P. De Fouquieres and S. G. Schirmer, Infinite Dimensional Analysis, Quantum Probability and Related Topics **16**, 1350021 (2013), https://doi.org/10.1142/S0219025713500215.
[11] D. V. Zhdanov, Journal of Physics A: Mathematical and Theoretical **51**, 508001 (2018).
[12] M. Larocca, P. M. Poggi, and D. A. Wisniacki, Journal of Physics A: Mathematical and Theoretical **51**, 385305 (2018).
[13] Z. Shen, M. Hsieh, and H. Rabitz, The Journal of chemical physics **124** (2006).
[14] M. Hsieh, R. Wu, and H. Rabitz, The Journal of chemical physics **130** (2009).
[15] A. Nanduri, A. Donovan, T.-S. Ho, and H. Rabitz, Physical Review A—Atomic, Molecular, and Optical Physics **88**, 033425 (2013).
[16] B. Koczor and S. C. Benjamin, Phys. Rev. Res. **4**, 023017 (2022).
[17] J. Landman, S. Thabet, C. Dalyac, H. Mhiri, and E. Kashefi, Classically approximating variational quantum machine learning with random fourier features (2022), arXiv:2210.13200 [quant-ph].
[18] F. J. Schreiber, J. Eisert, and J. J. Meyer, Phys. Rev. Lett. **131**, 100803 (2023).
[19] E. Fontana, I. Rungger, R. Duncan, and C. Cîrstoiu, Spectral analysis for noise diagnostics and filter-based digital error mitigation (2022), arXiv:2206.08811 [quant-ph].
[20] M. S. Rudolph, E. Fontana, Z. Holmes, and L. Cincio, Classical surrogate simulation of quantum systems with lowesa (2023), arXiv:2308.09109 [quant-ph].
[21] E. Fontana, M. S. Rudolph, R. Duncan, I. Rungger, and C. Cîrstoiu, Classical simulations of noisy variational quantum circuits (2023), arXiv:2306.05400 [quant-ph].
[22] C. Kokail, C. Maier, R. van Bijnen, T. Brydges, M. K. Joshi, P. Jurcevic, C. A. Muschik, P. Silvi, R. Blatt, C. F. Roos, and P. Zoller, Nature **569**, 355 (2019).
[23] F. Sauvage and F. Mintert, PRX Quantum **1**, 020322 (2020).
[24] F. Motzoi, J. M. Gambetta, S. T. Merkel, and F. K. Wilhelm, Physical Review A—Atomic, Molecular, and Optical Physics **84**, 022307 (2011).
[25] M. Schuld, R. Sweke, and J. J. Meyer, Phys. Rev. A **103**, 032430 (2021).
[26] M. Wiedmann, M. Periyasamy, and D. D. Scherer, Fourier analysis of variational quantum circuits for supervised learning (2024), arXiv:2411.03450 [cs.LG].
[27] C. M. Bishop, *Pattern Recognition and Machine Learning*, 1st ed., Information Science and Statistics (Springer, New York, NY, 2006).
[28] M. Reed and B. Simon, *Methods of modern mathematical physics: Functional analysis v. 1* (Academic Press, San Diego, CA, 1972).
[29] H. F. Trotter, Proceedings of the American Mathematical Society **10**, 545–551 (1959).

[30] S. Lloyd, Science **273**, 1073 (1996), https://www.science.org/doi/pdf/10.1126/science.273.5278.1073.

[31] M. Suzuki, Journal of Mathematical Physics **32**, 400 (1991), https://pubs.aip.org/aip/jmp/article-pdf/32/2/400/19166143/400_1_online.pdf.

[32] M. Schilling, F. Preti, M. M. Müller, T. Calarco, and F. Motzoi, Phys. Rev. Res. **6**, 043278 (2024).

[33] S. Kazi, M. Larocca, M. Farinati, P. J. Coles, M. Cerezo, and R. Zeier, Analyzing the quantum approximate optimization algorithm: ansätze, symmetries, and lie algebras (2024), arXiv:2410.05187 [quant-ph].

[34] M. Kormos, M. Collura, G. Takács, and P. Calabrese, Nature Physics **13**, 246 (2017).

[35] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm (2014), arXiv:1411.4028 [quant-ph].

[36] D. D'Alessandro and J. T. Hartwig, Journal of Dynamical and Control Systems **27**, 1 (2021).

[37] The random states $|r\rangle, |r'\rangle$ are obtained by sampling independently real and imaginary part of each coefficient in the computational basis, using a uniform distribution. Multiplication by an overall factor ensures correct normalization.

[38] Here we are thinking about $\mathcal{S}_n^{\Delta}$ as a vector of frequencies in ascending order, numbered from 0 to $n_\Delta - 1$.

[39] T. Caneva, T. Calarco, and S. Montangero, Physical Review A—Atomic, Molecular, and Optical Physics **84**, 022326 (2011).

[40] J. J. W. H. Sørensen, M. O. Aranburu, T. Heinzel, and J. F. Sherson, Phys. Rev. A **98**, 022119 (2018).

[41] Typically the Taylor expansion is written using a multiindex that avoids multiple counting, but as Lemma 7 shows, the two are equivalent.

[42] V. Smirnov, *Corso di matematica superiore* (Editori Riuniti, 2011).

[43] D. L. Goodwin and I. Kuprov, The Journal of Chemical Physics **144**, 204107 (2016), https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.4949534/15510918/204107_1_online.pdf.

[44] E. T. Whittaker, Proc. Roy. Soc. Edinb. **35**, 181 (1915).

[45] Hypersphere Packing – from Wolfram MathWorld — mathworld.wolfram.com, `https://mathworld.wolfram.com/HyperspherePacking.html`.

[46] One can check that by carrying on this infinitesimal quantity in the calculations the result is the same.

[47] In order to get around this difficulty, we could instead bound the number of approximate isolated critical points (without the need of specifying if they are minima or their depth as we defined it) by making use of the Taylor representation of the landscape, which we discussed in Sec. V. This problem is equivalent to finding the isolated zeros of a system of multivariate polynomials, whose number is bounded by Bezout's theorem and generalizations thereof [62]. This would potentially be informative to bound the complexity of the landscape in situations when a large number of suboptimal solutions are expected, which typically happens for critically constrained systems [2, 63].

[48] N. Rach, M. M. Müller, T. Calarco, and S. Montangero, Phys. Rev. A **92**, 062343 (2015).

[49] F. Preti, T. Calarco, and F. Motzoi, PRX Quantum **3**, 040311 (2022).

[50] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, Journal of Magnetic Resonance **172**, 296 (2005).

[51] M. Dalgaard, F. Motzoi, J. H. M. Jensen, and J. Sherson, Physical Review A **102**, 042612 (2020).

[52] M. Dalgaard, F. Motzoi, J. J. Sørensen, and J. Sherson, NPJ quantum information **6**, 6 (2020).

[53] D. R. Jones, C. D. Perttunen, and B. E. Stuckman, Journal of Optimization Theory and Applications **79**, 157 (1993).

[54] B. O. Shubert, SIAM Journal on Numerical Analysis **9**, 379 (1972), https://doi.org/10.1137/0709036.

[55] P. J. Nicholas (2015).

[56] H. Liu, S. Xu, X. Wang, J. Wu, and Y. Song, Engineering Optimization **47**, 1441 (2015), https://doi.org/10.1080/0305215X.2014.971777.

[57] In the original paper the quantity h is called δ.

[58] F. Motzoi, M. P. Kaicher, and F. K. Wilhelm, Physical review letters **119**, 160503 (2017).

[59] M. Goerz, D. Basilewitsch, F. Gago-Encinas, M. G. Krauss, K. P. Horn, D. M. Reich, and C. Koch, SciPost physics **7**, 080 (2019).

[60] A. Zygmund and R. Fefferman, *Trigonometric Series* (Cambridge University Press, 2003).

[61] M. A. Pinsky, *Introduction to Fourier analysis and wavelets*, Graduate studies in mathematics (American Mathematical Society, Providence, RI, 2009).

[62] D. N. Bernshtein, Functional Analysis and Its Applications **9**, 183 (1975).

[63] M. Bukov, A. G. R. Day, D. Sels, P. Weinberg, A. Polkovnikov, and P. Mehta, Phys. Rev. X **8**, 031086 (2018).

Chapter 7

Discussion

As anticipated in Part I, in this thesis we have dealt with several aspects concerning the development of a quantum firmware for cold atom platforms. In Chapter 4 we have focused on the transfer function mapping hardware inputs of a DMD into optical dipole potentials for the manipulation of ultracold gasses. We developed methods that allow for precise shaping of this potential while keeping in with the constraint of the rather low repetition rate of the experimental cycle. We envisage these methods to be useful in the shaping of time-dependent control pulses, especially if they involve the complex structures required to implement engineered systems such as quantum field thermal machines [103].

We then moved our focus in Chapter 5 from the transfer functions to the quantum system itself, considering the technically important control task of transporting an atom between different sites of an array of optical tweezers. We derived solutions for the transport problem that speed up this operation considerably compared to state-of-the-art experiments, while keeping residual vibrational excitations under control and by only using a single pair of moving tweezers. The possibility of rearranging the atoms without individual control of all the static sites reduces considerably the hardware requirements, simplifying platform design and potentially increasing the amount of atoms that can be dealt with using finite resources. In combination with two-qubit local collisional gates [104, 105], this could in perspective deliver a fermionic platform with a high degree of connectivity and a rather simple hardware design.

Finally, in Chapter 6 we still concerned ourselves with quantum control, but now from the perspective of characterizing the quantum cost landscape associated to this problem. We find that limited time and energy budgets translate in limited bandwidth and Lipschitz constant, which flatten the landscape as time discretization is refined. From the point of view of regression, this makes bandwidth-limited kernels particularly suited feature maps. In the context of closed-loop optimal control schemes, we envisage them as efficient local landscape estimators with

good noise-filtering properties, which can be used for robust gradient estimation or to devise analytic descent strategies [106]. Moreover, our analysis of landscape metrics is informative for the tuning of gradient-free optimizers, and highlights the relevance within QOC of Lipschitz-aware algorithms such as DIRECT [107, 108] in order to deal with landscape flattening.

These contributions therefore provide a collection of methods and insights that tackle important challenges in the development of quantum firmware for cold atom platforms, concerning a variety of experimental setups, different building blocks constituting these systems and all the three fundamental tasks that we identified in Chapter 3 as characterization, calibration and control. On a conclusive note, while the road to realize a quantum platform delivering an advantage in real-world simulational or computational tasks might still be very long, and the contributions presented in this thesis barely stir the surface of a deep ocean of complexity, it is the hope of the Author that what has been covered here might help those who thread this route after him save their own time and efforts, at least by avoiding the mistakes he made along the way.

Bibliography

[1] J. P. Dowling and G. J. Milburn. Quantum technology: the second quantum revolution, 2002. arXiv: quant-ph/0206091 [quant-ph]. URL: https://arxiv.org/abs/quant-ph/0206091.

[2] W. Gerlach and O. Stern. Der experimentelle nachweis der richtungsquantelung im magnetfeld. *Zeitschrift für Physik*, 9(1):349–352, Dec. 1922. ISSN: 0044-3328. DOI: 10.1007/BF01326983. URL: https://doi.org/10.1007/BF01326983.

[3] T. H. Maiman. Stimulated optical radiation in ruby. *Nature*, 187(4736):493–494, Aug. 1960. ISSN: 1476-4687. DOI: 10.1038/187493a0. URL: https://doi.org/10.1038/187493a0.

[4] C. N. Cohen-Tannoudji. Nobel lecture: manipulating atoms with photons. *Rev. Mod. Phys.*, 70:707–719, 3, July 1998. DOI: 10.1103/RevModPhys.70.707. URL: https://link.aps.org/doi/10.1103/RevModPhys.70.707.

[5] M. H. Anderson, J. R. Ensher, M. R. Matthews, C. E. Wieman, and E. A. Cornell. Observation of bose-einstein condensation in a dilute atomic vapor. *Science*, 269(5221):198–201, 1995. DOI: 10.1126/science.269.5221.198. URL: https://www.science.org/doi/abs/10.1126/science.269.5221.198.

[6] K. B. Davis, M. O. Mewes, M. R. Andrews, N. J. van Druten, D. S. Durfee, D. M. Kurn, and W. Ketterle. Bose-einstein condensation in a gas of sodium atoms. *Phys. Rev. Lett.*, 75:3969–3973, 22, Nov. 1995. DOI: 10.1103/PhysRevLett.75.3969. URL: https://link.aps.org/doi/10.1103/PhysRevLett.75.3969.

[7] W. S. Bakr, J. I. Gillen, A. Peng, S. Fölling, and M. Greiner. A quantum gas microscope for detecting single atoms in a hubbard-regime optical lattice. *Nature*, 462(7269):74–77, Nov. 2009. ISSN: 1476-4687. DOI: 10.1038/nature08482.

[8] L. W. Cheuk, M. A. Nichols, M. Okan, T. Gersdorf, V. V. Ramasesh, W. S. Bakr, T. Lompe, and M. W. Zwierlein. Quantum-gas microscope for fermionic atoms. *Phys. Rev. Lett.*, 114:193001, 19, May 2015. DOI: 10.1103/PhysRevLett.114.193001. URL: https://link.aps.org/doi/10.1103/PhysRevLett.114.193001.

[9] D. Barredo, S. de Léséleuc, V. Lienhard, T. Lahaye, and A. Browaeys. An atom-by-atom assembler of defect-free arbitrary two-dimensional atomic arrays. *Science*, 354(6315):1021–1023, 2016. DOI: 10.1126/science.aah3778. URL: https://www.science.org/doi/abs/10.1126/science.aah3778.

[10] 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 (New York, N.Y.)*, 354(6315):1024–1027, Nov. 2016. ISSN: 0036-8075. DOI: 10.1126/science.aah3752. URL: https://doi.org/10.1126/science.aah3752.

[11] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, J. P. Bonilla Ataides, N. Maskara, I. Cong, X. Gao, P. Sales Rodriguez, T. Karolyshyn, G. Semeghini, M. J. Gullans, M. Greiner, V. Vuletić, and M. D. Lukin. Logical quantum processor based on reconfigurable atom arrays. *Nature*, 626(7997):58–65, Feb. 2024. ISSN: 1476-4687. DOI: 10.1038/s41586-023-06927-3. URL: https://doi.org/10.1038/s41586-023-06927-3.

[12] F. Gyger, M. Ammenwerth, R. Tao, H. Timme, S. Snigirev, I. Bloch, and J. Zeiher. Continuous operation of large-scale atom arrays in optical lattices. *Phys. Rev. Res.*, 6:033104, 3, July 2024. DOI: 10.1103/PhysRevResearch.6.033104. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.6.033104.

[13] H. J. Manetsch, G. Nomura, E. Bataille, X. Lv, K. H. Leung, and M. Endres. A tweezer array with 6100 highly coherent atomic qubits. *Nature*, Sept. 2025. ISSN: 1476-4687. DOI: 10.1038/s41586-025-09641-4. URL: https://doi.org/10.1038/s41586-025-09641-4.

[14] M. Tajik, B. Rauer, T. Schweigler, F. Cataldini, J. Sabino, F. S. Møller, S.-C. Ji, I. E. Mazets, and J. Schmiedmayer. Designing arbitrary one-dimensional potentials on an atom chip. *Opt. Express*, 27(23):33474–33487, Nov. 2019. DOI: 10.1364/OE.27.033474. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-27-23-33474.

[15] T. Schweigler. *Correlations and dynamics of tunnel-coupled one-dimensional Bose gases*. PhD thesis, Technische Universität Wien, 2019. DOI: https://doi.org/10.34726/hss.2019.63111.

[16] G. Gauthier, I. Lenton, N. M. Parry, M. Baker, M. J. Davis, H. Rubinsztein-Dunlop, and T. W. Neely. Direct imaging of a digital-micromirror device for configurable microscopic optical potentials. *Optica*, 3(10):1136–1143, Oct. 2016. DOI: 10.1364/OPTICA.3.001136. URL: https://opg.optica.org/optica/abstract.cfm?URI=optica-3-10-1136.

[17] R. Folman, P. Krüger, J. Schmiedmayer, J. Denschlag, and C. Henkel. Microscopic atom optics: from wires to an atom chip. In B. Bederson and H. Walther, editors. Volume 48, Advances In Atomic, Molecular, and Optical Physics, pages 263–356. Academic Press, 2002. DOI: https://doi.org/10.1016/S1049-250X(02)80011-8. URL: https://www.sciencedirect.com/science/article/pii/S1049250X02800118.

[18] A. Browaeys and T. Lahaye. Many-body physics with individually controlled rydberg atoms. *Nature Physics*, 16(2):132–142, Feb. 2020. ISSN: 1745-2481. DOI: 10.1038/s41567-019-0733-z. URL: https://doi.org/10.1038/s41567-019-0733-z.

[19] C. Chin, R. Grimm, P. Julienne, and E. Tiesinga. Feshbach resonances in ultracold gases. *Rev. Mod. Phys.*, 82:1225–1286, 2, Apr. 2010. DOI: 10.1103/RevModPhys.82.1225. URL: https://link.aps.org/doi/10.1103/RevModPhys.82.1225.

[20] R. P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6):467–488, June 1982. ISSN: 1572-9575. DOI: 10.1007/BF02650179. URL: https://doi.org/10.1007/BF02650179.

[21] M. Greiner, O. Mandel, T. Esslinger, T. W. Hänsch, and I. Bloch. Quantum phase transition from a superfluid to a mott insulator in a gas of ultracold atoms. *Nature*, 415(6867):39–44, Jan. 2002. ISSN: 1476-4687. DOI: 10.1038/415039a. URL: https://doi.org/10.1038/415039a.

[22] T. Stöferle, H. Moritz, C. Schori, M. Köhl, and T. Esslinger. Transition from a strongly interacting 1d superfluid to a mott insulator. *Phys. Rev. Lett.*, 92:130403, 13, Mar. 2004. DOI: 10.1103/PhysRevLett.92.130403. URL: https://link.aps.org/doi/10.1103/PhysRevLett.92.130403.

[23] I. Bloch, J. Dalibard, and W. Zwerger. Many-body physics with ultracold gases. *Rev. Mod. Phys.*, 80:885–964, 3, July 2008. DOI: 10.1103/RevModPhys.80.885. URL: https://link.aps.org/doi/10.1103/RevModPhys.80.885.

[24] C. Gross and I. Bloch. Quantum simulations with ultracold atoms in optical lattices. *Science*, 357(6355):995–1001, 2017. DOI: 10.1126/science.aal3837. URL: https://www.science.org/doi/abs/10.1126/science.aal3837.

[25] J. J. García-Ripoll and J. I. Cirac. Spin dynamics for bosons in an optical lattice. *New Journal of Physics*, 5(1):76, June 2003. DOI: 10.1088/1367-2630/5/1/376. URL: https://dx.doi.org/10.1088/1367-2630/5/1/376.

[26] J. Simon, W. S. Bakr, R. Ma, M. E. Tai, P. M. Preiss, and M. Greiner. Quantum simulation of antiferromagnetic spin chains in an optical lattice. *Nature*, 472(7343):307–312, Apr. 2011. ISSN: 1476-4687. DOI: 10.1038/nature09994. URL: https://doi.org/10.1038/nature09994.

[27] H. Labuhn, D. Barredo, S. Ravets, S. de Léséleuc, T. Macrì, T. Lahaye, and A. Browaeys. Tunable two-dimensional arrays of single rydberg atoms for realizing quantum ising models. *Nature*, 534(7609):667–670, June 2016. ISSN: 1476-4687. DOI: 10.1038/nature18274. URL: https://doi.org/10.1038/nature18274.

[28] D.-W. Zhang, Y.-Q. Zhu, Y. X. Zhao, H. Yan, and S.-L. Zhu. Topological quantum matter with cold atoms. *Advances in Physics*, 67(4):253–402, 2018. DOI: 10.1080/00018732.2019.1594094. URL: https://doi.org/10.1080/00018732.2019.1594094.

[29] A. Celi, P. Massignan, J. Ruseckas, N. Goldman, I. B. Spielman, G. Juzeli ūnas, and M. Lewenstein. Synthetic gauge fields in synthetic dimensions. *Phys. Rev. Lett.*, 112:043001, 4, Jan. 2014. DOI: 10.1103/PhysRevLett.112.043001. URL: https://link.aps.org/doi/10.1103/PhysRevLett.112.043001.

[30] G. Jotzu, M. Messer, R. Desbuquois, M. Lebrat, T. Uehlinger, D. Greif, and T. Esslinger. Experimental realization of the topological haldane model with ultracold fermions. *Nature*, 515(7526):237–240, Nov. 2014. ISSN: 1476-4687. DOI: 10.1038/nature13915. URL: https://doi.org/10.1038/nature13915.

[31] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.

[32] S. Lloyd. Universal quantum simulators. *Science*, 273(5278):1073–1078, 1996. DOI: 10.1126/science.273.5278.1073. URL: https://www.science.org/doi/abs/10.1126/science.273.5278.1073.

[33] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, 1997. DOI: 10.1137/S0097539795293172. URL: https://doi.org/10.1137/S0097539795293172.

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

[35] A. W. Harrow, A. Hassidim, and S. Lloyd. Quantum algorithm for linear systems of equations. *Phys. Rev. Lett.*, 103:150502, 15, Oct. 2009. DOI: 10.1103/PhysRevLett.103.150502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.103.150502.

[36] P. W. Shor. Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A*, 52:R2493–R2496, 4, Oct. 1995. DOI: 10.1103/PhysRevA.52.R2493. URL: https://link.aps.org/doi/10.1103/PhysRevA.52.R2493.

[37] S. Ebadi, T. T. Wang, H. Levine, A. Keesling, G. Semeghini, A. Omran, D. Bluvstein, R. Samajdar, H. Pichler, W. W. Ho, S. Choi, S. Sachdev, M. Greiner, V. Vuletić, and M. D. Lukin. Quantum phases of matter on a 256-atom programmable quantum simulator. *Nature*, 595(7866):227–232, July 2021. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03582-4. URL: https://doi.org/10.1038/s41586-021-03582-4.

[38] H. Zhou, C. Duckering, C. Zhao, D. Bluvstein, M. Cain, A. Kubica, S.-T. Wang, and M. D. Lukin. Resource analysis of low-overhead transversal architectures for reconfigurable atom arrays. In *Proceedings of the 52nd Annual International Symposium on Computer Architecture*, ISCA '25, pages 1432–1448. Association for Computing Machinery, 2025. ISBN: 9798400712616. DOI: 10.1145/3695053.3731039. URL: https://doi.org/10.1145/3695053.3731039.

[39] H. Ball, M. J. Biercuk, and M. R. Hush. Quantum firmware and the quantum computing stack. *Physics Today*, 74(3):28–34, Mar. 2021. ISSN: 0031-9228. DOI: 10.1063/PT.3.4698. URL: https://doi.org/10.1063/PT.3.4698.

[40] N. Wittler, F. Roy, K. Pack, M. Werninghaus, A. S. Roy, D. J. Egger, S. Filipp, F. K. Wilhelm, and S. Machnes. Integrated tool set for control, calibration, and characterization of quantum devices applied to su-

perconducting qubits. *Phys. Rev. Appl.*, 15:034080, 3, Mar. 2021. DOI: 10.1103/PhysRevApplied.15.034080. URL: https://link.aps.org/doi/10.1103/PhysRevApplied.15.034080.

[41] A. S. Roy, K. Pack, N. Wittler, and S. Machnes. Software tool-set for automated quantum system identification and device bring up. In *2025 17th International Conference on COMmunication Systems and NETworks (COMSNETS)*, pages 1062–1067, 2025. DOI: 10.1109/COMSNETS63942.2025.10885595.

[42] J. Singh, R. Zeier, T. Calarco, and F. Motzoi. Compensating for nonlinear distortions in controlled quantum systems. *Phys. Rev. Appl.*, 19:064067, 6, June 2023. DOI: 10.1103/PhysRevApplied.19.064067. URL: https://link.aps.org/doi/10.1103/PhysRevApplied.19.064067.

[43] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, June 2021. ISSN: 2522-5820. DOI: 10.1038/s42254-021-00314-5. URL: https://doi.org/10.1038/s42254-021-00314-5.

[44] C. Meng, S. Griesemer, D. Cao, S. Seo, and Y. Liu. When physics meets machine learning: a survey of physics-informed machine learning. *Machine Learning for Computational Science and Engineering*, 1(1):20, May 2025. ISSN: 3005-1436. DOI: 10.1007/s44379-025-00016-0. URL: https://doi.org/10.1007/s44379-025-00016-0.

[45] C. Cohen-Tannoudji, B. Diu, and F. Laloë. *Band 1 Quantenmechanik*. De Gruyter, Berlin, Boston, 2019. ISBN: 9783110638738. DOI: doi:10.1515/9783110638738. URL: https://doi.org/10.1515/9783110638738.

[46] C. Cohen-Tannoudji, B. Diu, and F. Laloë. *Band 2 Quantenmechanik*. De Gruyter, Berlin, Boston, 2019. ISBN: 9783110638769. DOI: doi:10.1515/9783110638769. URL: https://doi.org/10.1515/9783110638769.

[47] C. Cohen-Tannoudji, J. Dupont-Roc, and G. Grynberg. *Atom-photon interactions*. en. Blackwell Verlag, Berlin, Germany, July 2024.

[48] D. A. Steck. Quantum and atom optics. 2025. URL: http://steck.us/teaching (visited on 09/25/2025).

[49] W. Ketterle. Atomic and optical physics i. 2014. URL: https://ocw.mit.edu/courses/8-421-atomic-and-optical-physics-i-spring-2014/ (visited on 09/25/2025).

[50] W. Ketterle. Atomic and optical physics ii. 2013. URL: https://ocw.mit.edu/courses/8-422-atomic-and-optical-physics-ii-spring-2013/ (visited on 09/25/2025).

[51] P. van der Straten and H. Metcalf. *The alkali-metal atoms*. In *Atoms and Molecules Interacting with Light: Atomic Physics for the Laser Era*. Cambridge University Press, 2016, pages 164–180.

[52] D. A. Steck. Rubidium 87 d line data. 2025. URL: https://steck.us/alkalidata/ (visited on 09/25/2025).

[53] E. Arimondo, M. Inguscio, and P. Violino. Experimental determinations of the hyperfine structure in the alkali atoms. *Rev. Mod. Phys.*, 49:31–75, 1, Jan. 1977. DOI: 10.1103/RevModPhys.49.31. URL: https://link.aps.org/doi/10.1103/RevModPhys.49.31.

[54] R. Grimm, M. Weidemüller, and Y. B. Ovchinnikov. Optical dipole traps for neutral atoms. In B. Bederson and H. Walther, editors. Volume 42, Advances In Atomic, Molecular, and Optical Physics, pages 95–170. Academic Press, 2000. DOI: https://doi.org/10.1016/S1049-250X(08)60186-X. URL: https://www.sciencedirect.com/science/article/pii/S1049250X0860186X.

[55] J. D. Jackson. *Classical Electrodynamics*. en. John Wiley & Sons, Nashville, TN, 3rd edition, July 1998.

[56] W. Ketterle and N. V. Druten. Evaporative cooling of trapped atoms. In B. Bederson and H. Walther, editors. Volume 37, Advances In Atomic, Molecular, and Optical Physics, pages 181–236. Academic Press, 1996. DOI: https://doi.org/10.1016/S1049-250X(08)60101-9. URL: https://www.sciencedirect.com/science/article/pii/S1049250X08601019.

[57] M. Tajik. *Arbitrary one-dimensional optical dipole potentials on an atom chip*. Master's thesis, Technische Universität Wien, 2017. DOI: https://doi.org/10.34726/hss.2017.43483.

[58] T. Wilk, A. Gaëtan, C. Evellin, J. Wolters, Y. Miroshnychenko, P. Grangier, and A. Browaeys. Entanglement of two individual neutral atoms using rydberg blockade. *Phys. Rev. Lett.*, 104:010502, 1, Jan. 2010. DOI: 10.1103/PhysRevLett.104.010502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.104.010502.

[59] D. A. Smith, S. Aigner, S. Hofferberth, M. Gring, M. Andersson, S. Wildermuth, P. Krüger, S. Schneider, T. Schumm, and J. Schmiedmayer. Absorption imaging of ultracold atoms on atom chips. *Opt. Express*, 19(9):8471–8485, Apr. 2011. DOI: 10.1364/OE.19.008471. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-19-9-8471.

[60] R. Folman, P. Krüger, D. Cassettari, B. Hessmo, T. Maier, and J. Schmiedmayer. Controlling cold atoms using nanofabricated surfaces: atom chips. *Phys. Rev. Lett.*, 84:4749–4752, 20, May 2000. DOI: 10.1103/PhysRevLett.84.4749. URL: https://link.aps.org/doi/10.1103/PhysRevLett.84.4749.

[61] F. Nogrette, H. Labuhn, S. Ravets, D. Barredo, L. Béguin, A. Vernier, T. Lahaye, and A. Browaeys. Single-atom trapping in holographic 2d arrays of microtraps with arbitrary geometries. *Phys. Rev. X*, 4:021034, 2, May 2014. DOI: 10.1103/PhysRevX.4.021034. URL: https://link.aps.org/doi/10.1103/PhysRevX.4.021034.

[62] Y.-Q. Zou, É. Le Cerf, B. Bakkali-Hassani, C. Maury, G. Chauveau, P. C. M. Castilho, R. Saint-Jalm, S. Nascimbene, J. Dalibard, and J. Beugnon. Optical control of the density and spin spatial profiles of a planar bose gas. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 54(8):08LT01, Apr. 2021. DOI: 10.1088/1361-6455/abf298. URL: https://doi.org/10.1088/1361-6455/abf298.

[63] A. Mazurenko, C. S. Chiu, G. Ji, M. F. Parsons, M. Kanász-Nagy, R. Schmidt, F. Grusdt, E. Demler, D. Greif, and M. Greiner. A cold-atom fermi–hubbard antiferromagnet. *Nature*, 545(7655):462–466, May 2017. ISSN: 1476-4687. DOI: 10.1038/nature22362. URL: https://doi.org/10.1038/nature22362.

[64] L.-M. Duan, E. Demler, and M. D. Lukin. Controlling spin exchange interactions of ultracold atoms in optical lattices. *Phys. Rev. Lett.*, 91:090402, 9, Aug. 2003. DOI: 10.1103/PhysRevLett.91.090402. URL: https://link.aps.org/doi/10.1103/PhysRevLett.91.090402.

[65] R. Liao, J. Sun, P. Zhao, S. Yang, H. Li, X. Huang, W. Xiong, X. Zhou, D. Li, X. Liu, and X. Chen. Simulation of exact quantum ising models with a mott insulator of paired atoms. *Phys. Rev. A*, 106:053308, 5, Nov. 2022. DOI: 10.1103/PhysRevA.106.053308. URL: https://link.aps.org/doi/10.1103/PhysRevA.106.053308.

[66] S. Sachdev, K. Sengupta, and S. M. Girvin. Mott insulators in strong electric fields. *Phys. Rev. B*, 66:075128, 7, Aug. 2002. DOI: 10.1103/PhysRevB.66.075128. URL: https://link.aps.org/doi/10.1103/PhysRevB.66.075128.

[67] L. Magni and R. Scattolini. *Advanced and multivariable control*. en. Societa Editrice Esculapio, June 2023.

[68] 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(2):296–305, 2005. ISSN: 1090-7807. DOI: https://doi.org/10.1016/j.jmr.2004.11.004.

[69] P. Doria, T. Calarco, and S. Montangero. Optimal control technique for many-body quantum dynamics. *Phys. Rev. Lett.*, 106:190501, 19, May 2011. DOI: 10.1103/PhysRevLett.106.190501. URL: https://link.aps.org/doi/10.1103/PhysRevLett.106.190501.

[70] C. M. Bishop. *Pattern Recognition and Machine Learning*. en. Information Science and Statistics. Springer, New York, NY, 1st edition, Aug. 2006.

[71] D. C. Liu and J. Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical Programming*, 45(1):503–528, Aug. 1989. ISSN: 1436-4646. DOI: 10.1007/BF01589116.

[72] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab. A high-bias, low-variance introduction to machine learning for physicists. *Physics Reports*, 810:1–124, 2019. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2019.03.001. URL: https://www.sciencedirect.com/science/article/pii/S0370157319300766. A high-bias, low-variance introduction to Machine Learning for physicists.

[73] Jax: high performance array computing. 2024. URL: https://docs.jax.dev/en/latest/ (visited on 09/26/2025).

[74] S. J. Glaser, U. Boscain, T. Calarco, C. P. Koch, W. Köckenberger, R. Kosloff, I. Kuprov, B. Luy, S. Schirmer, T. Schulte-Herbrüggen, D. Sugny, and F. K. Wilhelm. Training schrödinger's cat: quantum optimal control. *The European Physical Journal D*, 69(12):279, Dec. 2015. ISSN: 1434-6079. DOI: 10.1140/epjd/e2015-60464-1. URL: https://doi.org/10.1140/epjd/e2015-60464-1.

[75] C. P. Koch, U. Boscain, T. Calarco, G. Dirr, S. Filipp, S. J. Glaser, R. Kosloff, S. Montangero, T. Schulte-Herbrüggen, D. Sugny, and F. K. Wilhelm. Quantum optimal control in quantum technologies. strategic report on current status, visions and goals for research in europe. *EPJ Quantum Technology*, 9(1):19, July 2022. ISSN: 2196-0763. DOI: 10.1140/epjqt/s40507-022-00138-x. URL: https://doi.org/10.1140/epjqt/s40507-022-00138-x.

[76] J. J. Sakurai and J. Napolitano. *Modern Quantum Mechanics*. Cambridge University Press, 3rd edition, 2020.

[77] M. A. Nielsen. A simple formula for the average gate fidelity of a quantum dynamical operation. *Physics Letters A*, 303(4):249–252, Oct. 2002. ISSN: 0375-9601. DOI: 10.1016/s0375-9601(02)01272-0. URL: http://dx.doi.org/10.1016/S0375-9601(02)01272-0.

[78] M. Goerz, D. Basilewitsch, F. Gago-Encinas, M. G. Krauss, K. P. Horn, D. M. Reich, and C. Koch. Krotov: a python implementation of krotov's method for quantum optimal control. *SciPost physics*, 7(6):080, 2019.

[79] S. Machnes, E. Assémat, D. Tannor, and F. K. Wilhelm. Tunable, flexible, and efficient optimization of control pulses for practical qubits. *Phys. Rev. Lett.*, 120:150401, 15, Apr. 2018. DOI: 10.1103/PhysRevLett.120.150401. URL: https://link.aps.org/doi/10.1103/PhysRevLett.120.150401.

[80] J. J. W. H. Sørensen, M. O. Aranburu, T. Heinzel, and J. F. Sherson. Quantum optimal control in a chopped basis: applications in control of bose-einstein condensates. *Phys. Rev. A*, 98:022119, 2, Aug. 2018. DOI: 10.1103/PhysRevA.98.022119. URL: https://link.aps.org/doi/10.1103/PhysRevA.98.022119.

[81] D. L. Goodwin and I. Kuprov. Auxiliary matrix formalism for interaction representation transformations, optimal control, and spin relaxation theories. *The Journal of Chemical Physics*, 143(8):084113, Aug. 2015. ISSN: 0021-9606. DOI: 10.1063/1.4928978. URL: https://doi.org/10.1063/1.4928978.

[82] D. L. Goodwin and I. Kuprov. Modified newton-raphson grape methods for optimal control of spin systems. *The Journal of Chemical Physics*, 144(20):204107, May 2016. ISSN: 0021-9606. DOI: 10.1063/1.4949534. URL: https://doi.org/10.1063/1.4949534.

[83] N. Rach, M. M. Müller, T. Calarco, and S. Montangero. Dressing the chopped-random-basis optimization: a bandwidth-limited access to the trap-free landscape. *Phys. Rev. A*, 92:062343, 6, Dec. 2015. DOI: 10.1103/PhysRevA.92.062343. URL: https://link.aps.org/doi/10.1103/PhysRevA.92.062343.

[84] E. Torrontegui, S. Ibáñez, S. Martínez-Garaot, M. Modugno, A. del Campo, D. Guéry-Odelin, A. Ruschhaupt, X. Chen, and J. G. Muga. Chapter 2 - shortcuts to adiabaticity. In E. Arimondo, P. R. Berman, and C. C. Lin, editors, *Advances in Atomic, Molecular, and Optical Physics*. Volume 62, Advances In Atomic, Molecular, and Optical Physics, pages 117–169. Aca-

demic Press, 2013. DOI: https://doi.org/10.1016/B978-0-12-408090-4.00002-5.

[85] D. Guéry-Odelin, A. Ruschhaupt, A. Kiely, E. Torrontegui, S. Martínez-Garaot, and J. G. Muga. Shortcuts to adiabaticity: concepts, methods, and applications. *Rev. Mod. Phys.*, 91:045001, 4, Oct. 2019. DOI: 10.1103/RevModPhys.91.045001. URL: https://link.aps.org/doi/10.1103/RevModPhys.91.045001.

[86] M. V. Berry. Transitionless quantum driving. *Journal of Physics A: Mathematical and Theoretical*, 42(36):365303, Aug. 2009. DOI: 10.1088/1751-8113/42/36/365303. URL: https://doi.org/10.1088/1751-8113/42/36/365303.

[87] A. del Campo. Shortcuts to adiabaticity by counterdiabatic driving. *Phys. Rev. Lett.*, 111:100502, 10, Sept. 2013. DOI: 10.1103/PhysRevLett.111.100502. URL: https://link.aps.org/doi/10.1103/PhysRevLett.111.100502.

[88] H. R. Lewis and W. B. Riesenfeld. An exact quantum theory of the time-dependent harmonic oscillator and of a charged particle in a time-dependent electromagnetic field. *J. Math. Phys.*, 10(8):1458–1473, 1969. DOI: 10.1063/1.1664991.

[89] X. Chen, A. Ruschhaupt, S. Schmidt, A. del Campo, D. Guéry-Odelin, and J. G. Muga. Fast optimal frictionless atom cooling in harmonic traps: shortcut to adiabaticity. *Phys. Rev. Lett.*, 104:063002, 6, Feb. 2010. DOI: 10.1103/PhysRevLett.104.063002. URL: https://link.aps.org/doi/10.1103/PhysRevLett.104.063002.

[90] E. Torrontegui, S. Ibáñez, X. Chen, A. Ruschhaupt, D. Guéry-Odelin, and J. G. Muga. Fast atomic transport without vibrational heating. *Phys. Rev. A*, 83:013415, 1, Jan. 2011. DOI: 10.1103/PhysRevA.83.013415. URL: https://link.aps.org/doi/10.1103/PhysRevA.83.013415.

[91] A. K. Dhara and S. V. Lawande. Feynman propagator for time-dependent lagrangians possessing an invariant quadratic in momentum. *J. Phys. A*, 17(12):2423, Aug. 1984. DOI: 10.1088/0305-4470/17/12/014. URL: https://dx.doi.org/10.1088/0305-4470/17/12/014.

[92] M. Calzavara, Y. Kuriatnikov, A. Deutschmann-Olek, F. Motzoi, S. Erne, A. Kugi, T. Calarco, J. Schmiedmayer, and M. Prüfer. Optimizing optical potentials with physics-inspired learning algorithms. *Phys. Rev. Appl.*, 19:044090, 4, Apr. 2023. DOI: 10.1103/PhysRevApplied.19.044090. URL: https://link.aps.org/doi/10.1103/PhysRevApplied.19.044090.

[93] P. Zupancic, P. M. Preiss, R. Ma, A. Lukin, M. E. Tai, M. Rispoli, R. Islam, and M. Greiner. Ultra-precise holographic beam shaping for microscopic quantum control. *Opt. Express*, 24(13):13881–13893, June 2016. DOI: 10.1364/OE.24.013881. URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-24-13-13881.

[94] N. Navon, R. P. Smith, and Z. Hadzibabic. Quantum gases in optical boxes. *Nature Physics*, 17(12):1334–1341, Dec. 2021. ISSN: 1745-2481. DOI: 10.1038/s41567-021-01403-z. URL: https://doi.org/10.1038/s41567-021-01403-z.

[95] A. del Campo and M. G. Boshier. Shortcuts to adiabaticity in a time-dependent box. *Scientific Reports*, 2(1):648, Sept. 2012. ISSN: 2045-2322. DOI: 10.1038/srep00648. URL: https://doi.org/10.1038/srep00648.

[96] A. Deutschmann-Olek, M. Tajik, M. Calzavara, J. Schmiedmayer, T. Calarco, and A. Kugi. Iterative shaping of optical potentials for one-dimensional bose-einstein condensates. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pages 5801–5806, 2022. DOI: 10.1109/CDC51059.2022.9993271.

[97] C. Cicali, M. Calzavara, E. Cuestas, T. Calarco, R. Zeier, and F. Motzoi. Fast neutral-atom transport and transfer between optical tweezers. *Phys. Rev. Appl.*, 24:024070, 2, Aug. 2025. DOI: 10.1103/7r3w-8m61. URL: https://link.aps.org/doi/10.1103/7r3w-8m61.

[98] M. Calzavara, T. Calarco, and F. Motzoi. Classical feature map surrogates and metrics for quantum control landscapes, 2025. arXiv: 2509.25930 [quant-ph]. URL: https://arxiv.org/abs/2509.25930.

[99] M. Schuld, R. Sweke, and J. J. Meyer. Effect of data encoding on the expressive power of variational quantum-machine-learning models. *Phys. Rev. A*, 103:032430, 3, Mar. 2021. DOI: 10.1103/PhysRevA.103.032430. URL: https://link.aps.org/doi/10.1103/PhysRevA.103.032430.

[100] N. Beato, P. Patil, and M. Bukov. Towards a theory of phase transitions in quantum control landscapes, 2024. arXiv: 2408.11110 [quant-ph]. URL: https://arxiv.org/abs/2408.11110.

[101] M. Dalgaard, F. Motzoi, and J. Sherson. Predicting quantum dynamical cost landscapes with deep learning. *Phys. Rev. A*, 105:012402, 1, Jan. 2022. DOI: 10.1103/PhysRevA.105.012402. URL: https://link.aps.org/doi/10.1103/PhysRevA.105.012402.

[102] M. Larocca, P. Czarnik, K. Sharma, G. Muraleedharan, P. J. Coles, and M. Cerezo. Diagnosing Barren Plateaus with Tools from Quantum Optimal Control. *Quantum*, 6:824, Sept. 2022. ISSN: 2521-327X. DOI: 10.22331/q-2022-09-29-824. URL: https://doi.org/10.22331/q-2022-09-29-824.

[103] M. Gluza, J. Sabino, N. H. Ng, G. Vitagliano, M. Pezzutto, Y. Omar, I. Mazets, M. Huber, J. Schmiedmayer, and J. Eisert. Quantum field thermal machines. *PRX Quantum*, 2:030310, 3, July 2021. DOI: 10.1103/PRXQuantum.2.030310. URL: https://link.aps.org/doi/10.1103/PRXQuantum.2.030310.

[104] J. Singh, J. A. Reuter, T. Calarco, F. Motzoi, and R. Zeier. Optimizing two-qubit gates for ultracold atoms using fermi-hubbard models. *Phys. Rev. Appl.*, 24:034007, 3, Sept. 2025. DOI: 10.1103/xqzw-m27l. URL: https://link.aps.org/doi/10.1103/xqzw-m27l.

[105] P. Bojović, T. Hilker, S. Wang, J. Obermeyer, M. Barendregt, D. Tell, T. Chalopin, P. M. Preiss, I. Bloch, and T. Franz. High-fidelity collisional quantum gates with fermionic atoms, 2025. arXiv: 2506.14711 [cond-mat.quant-gas]. URL: https://arxiv.org/abs/2506.14711.

[106] B. Koczor and S. C. Benjamin. Quantum analytic descent. *Phys. Rev. Res.*, 4:023017, 2, Apr. 2022. DOI: 10.1103/PhysRevResearch.4.023017. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.4.023017.

[107] D. R. Jones, C. D. Perttunen, and B. E. Stuckman. Lipschitzian optimization without the lipschitz constant. *Journal of Optimization Theory and Applications*, 79:157–181, 1993. URL: https://api.semanticscholar.org/CorpusID:123674634.

[108] C. Kokail, C. Maier, R. van Bijnen, T. Brydges, M. K. Joshi, P. Jurcevic, C. A. Muschik, P. Silvi, R. Blatt, C. F. Roos, and P. Zoller. Self-verifying variational quantum simulation of lattice models. *Nature*, 569(7756):355–360, May 2019. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1177-4. URL: https://doi.org/10.1038/s41586-019-1177-4.