Adaptive Optimal Control of Entangled Qubits

David L. Goodwin,[1] Pranav Singh,[2*] Mohammadali Foroozandeh[1*]

[1]Chemistry Research Laboratory, University of Oxford,
Mansfield Road, Oxford OX1 3TA, UK
[2]Department of Mathematical Sciences, University of Bath,
Bath BA2 7AY, UK

*To whom correspondence should be addressed;
E-mail: ps2106@bath.ac.uk, mohammadali.foroozandeh@chem.ox.ac.uk.

Developing fast, robust, and accurate methods for optimal control of quantum systems comprising interacting particles is one of the most active areas of current science. Although a valuable repository of algorithms is available for numerical applications in quantum control, the high computational cost is somewhat overlooked. Here we present a fast and accurate optimal control algorithm for systems of interacting qubits, QOALA (Quantum Optimal control by Adaptive Low-cost Algorithm), that is predicted to offer $\mathcal{O}(M^2)$ speedup for an M-qubit system, compared to the state-of-the-art exact methods, without compromising overall accuracy of the optimal solution. The method is general and compatible with diverse Hamiltonian structures. The proposed approach uses inexpensive low-accuracy approximations of propagators far from the optimum, adaptively switching to higher accuracy, higher-cost propagators when approaching the optimum. Additionally, the utilisation of analytical Lie algebraic derivatives that do not require computationally expensive matrix exponential brings even better performance.

1 Introduction

Driving quantum spins systems to a desired target state via optimal control theory (*1*) has been widely applied to a range of areas form high-resolution magnetic resonance spectroscopy and imaging (*2–5*), and electron paramagnetic resonance (*6–8*), quantum error-correction and quantum information registers (*9–11*), cold atoms (*12–14*), terahertz technologies (*15,16*), control of

trapped ions (*17*), and NV-centres in diamond (*18, 19*). Along with applications in measurement science, algorithmic and numerical developments of optimal control methods remain active and challenging, with examples including geometric optimal control (*20, 21*), adiabatic optimal control (*22, 23*), GRAPE (GRadient Ascent Pulse Engineering) (*24, 25*) and Krotov (*26, 27*) algorithms, tensor product approach for large quantum systems (*28*), and optimal control over approximate control landscapes (*29*).

A crucial ingredient in these optimal control applications is a numerical method for computing the dynamics of spins systems, which is utilised for computing the objective function at each iteration of an optimisation algorithm. A common feature in most numerical methods is that a uniformly and highly-accurate method is utilised throughout the optimisation process when, for the significant part, the optimisation can not benefit from the provided accuracy, and hence suffers from the computational burden without much gain. This problem becomes more evident and cumbersome, especially, in the context of large, multi-particle systems and therefore more important to address.

In this manuscript we present a general and highly flexible approach for solving optimal control problems for systems of entangled qubits in a computationally efficient manner without sacrificing the desired accuracy. In an optimisation process, computationally inexpensive methods for computing spin dynamics can be used far from the optimum, where loss of accuracy is less crucial and where the most iterations in a numerical optimisation routine are often spent. Conversely, high accuracy methods should only be used for computing the dynamics of quantum systems close to the optimum when high fidelity is achievable and desired. This *adaptive optimal control* method, QOALA (Quantum Optimal control by Adaptive Low-cost Algorithm), uses a set of approximant propagators that are designed to allow variable degrees of trade-off between accuracy and computational expense, and achieves significant speedup in practice. We elucidate the potential speedup of QOALA with the concrete example of a class of numerical methods called *propagator splittings*. However, the overall framework developed in this manuscript is flexible enough to incorporate any combination of numerical methods with different cost and accuracy trade-offs.

This paper is structured as follows: In Sections 2.1 to 2.3 a brief and general theory of optimal control is presented, and the concept of adaptive approach is introduced. In Section 2.4 the ingredients of the QOALA method including the Hamiltonian structure, propagator splitting, and computation of derivatives is presented, along with potential benchmarking strategies. Finally in Section 2.5 demonstrations of the the proposed method are presented in the context of nuclear magnetic resonance (NMR) using a range of examples including state transfers and swap gates on 2-, 3-, and 4-spin systems. In each case, the convergence and wall-clock time of the method are compared to methods (*30–32*) available via the versatile software package *Spinach* (*33*). It is demonstrated that the proposed adaptive framework consistently outperforms available methods. Relative speedup of QOALA compared to an exact method is shown in Figure 1.

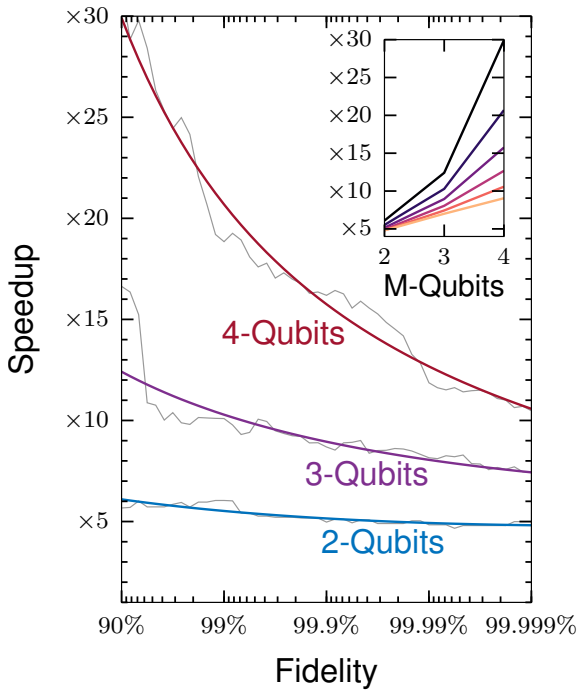

Figure 1: **Speedup of QOALA state-transfer optimal control problems.** Speedup in wall-clock time achieved by utilising the QOALA method when compared to an exact method for state-transfer optimal control problems. Smooth, thick lines are a polynomial fit to the thin lines. The inset plot shows the expected speedup, per qubit, for different fidelities from 90% (top line) to 99.9999% (bottom line).

2 Results

2.1 Formulation of optimal control problem

The state of a system with M spins at time t is described by a density matrix $\rho(t)$, and its dynamics are governed by the Liouville–von Neumann equation,

$$\frac{\partial \rho(t)}{\partial t} = -i\mathcal{H}(t)\rho(t), \qquad \rho(t_0) = \rho_0, \tag{1}$$

where ρ_0 is the initial state of the system at time t_0, and $\mathcal{H} = \mathbb{1} \otimes H - H^\top \otimes \mathbb{1}$ is the time-dependent Hamiltonian super-operator of the system. The solution of eq. (1) can be expressed as

$$\rho(t) = \mathbf{U}(t)\rho(0). \tag{2}$$

Formally, the propagator $\mathbf{U}(t)$ is a time-ordered exponential of $-i\mathcal{H}$, denoted

$$\mathbf{U}(t) = \mathcal{T}\mathrm{e}^{-i\int_0^t \mathcal{H}(t')\mathrm{d}t'}. \tag{3}$$

In optimal control applications the Hamiltonian, $\mathcal{H}(t)$, depends on some control parameters, θ, and optimal choices of these parameters are sought which can either (i) drive the state of the system $\rho(T)$ at a specified time T to a desirable target ϱ, or (ii) ensure that the propagator $\mathbf{U}(T)$ implements a desirable propagator $\mathbf{U}_{\text{target}}$. To highlight the dependence on θ, we often write $\mathcal{H}(t;\theta)$, $\mathbf{U}(t;\theta)$ and $\rho(t;\theta)$ instead of $\mathcal{H}(t)$, $\mathbf{U}(t)$ and $\rho(t)$, respectively, where normally $t \in [0, T]$. Further, we assume that $\mathcal{H}(t;\theta)$ and $\mathbf{U}(t;\theta)$ are represented by $N \times N$ matrices.

The optimal control problem is expressed in terms of maximising an objective function

$$\mathcal{F}(\theta) = f(\mathbf{U}(T;\theta)) \tag{4}$$

where f is a continuously differentiable real-valued functional of propagator matrices, i.e.

$$f \in \mathrm{C}(\mathbb{C}^{N\times N};\mathbb{R}). \tag{5}$$

In this manuscript we are particularly concerned with cases where the value of the objective at the optimal parameters,

$$\theta^* = \underset{\theta}{\operatorname{argmax}}\, \mathcal{F}(\theta), \tag{6}$$

is known a-priori. Without loss of generality, we assume that after suitable scaling of f, the objective $\mathcal{F}$ is able to achieve the maximum possible value of 1,

$$\mathcal{F}(\theta^*) = 1, \tag{7}$$

A large class of objective functions $\mathcal{F}$ of this form that appear in the context of optimal control of spins and gate design are *fidelity* functions. For instance,

$$f(X) = \mathrm{Re}\left[\mathrm{Tr}\left(\varrho^\dagger X \rho_0\right)\right] \tag{8}$$

leads to the fidelity functional

$$\mathcal{F}(\theta) = \mathrm{Re}\left[\mathrm{Tr}\left(\varrho^\dagger \mathbf{U}(T;\theta)\rho_0\right)\right] = \mathrm{Re}\left[\mathrm{Tr}\left(\varrho^\dagger \rho(T;\theta)\right)\right] \quad \in [0,1], \tag{9}$$

which measures the overlap of the quantum state $\rho(T;\theta)$ at a specified time $t = T$, with a target state ϱ. Here, both ρ_0 and ϱ are assumed to be *improper* (i.e. trace-free, $\mathrm{Tr}(\rho) = 0$), normalised ($\|\rho\|_2 = 1$), Hermitian ($\rho^\dagger = \rho$) density matrices. Functionals such as eq. (8) where an initial state ρ_0 appears, lead to fidelity functions that are utilised in the context of state-to-state transfers, i.e. case (i).

Here normalisation is only assumed so that value at the optima is 1, i.e. to satisfy eq. (7). Note that the requirement of trace-free density matrices only applies to the fidelity functional described in eq. (9), and the overall approach applies equally to fidelity functions of proper density matrices (i.e. non trace-free density matrices).

The functional

$$f(X) = \mathrm{Re}\left[\mathrm{Tr}\left(\mathbf{U}_{\text{target}}^\dagger X\right)\right], \tag{10}$$

leads to the fidelity

$$\mathcal{F}(\theta) = \mathrm{Re}[\mathrm{Tr}(\mathbf{U}_{\text{target}}^{\dagger}\mathbf{U}(T;\theta))], \tag{11}$$

which does not depend on initial state ρ_0, and can be utilised where an effective design of a specified propagator $\mathbf{U}_{\text{target}}$ (a desired unitary gate, for instance) is required, i.e. case (ii). Note that, instead of working with the real-value of the trace in eqs. (8) and (10), i.e. applying the function $\mathrm{Re}(\cdot)$, we can also apply any other continuously differentiable function, e.g. absolute value squared, $|\cdot|^2$.

Fidelity functions measure the similarity between mixed quantum states. A very wide range of fidelity functions exist, among them notably the Uhlmann–Jozsa fidelity (*34*), which satisfy the *fidelity axioms* to different extents. We refer the reader to (*35*) for a more detailed discussion of the appropriateness of various fidelity measures. For the purposes of this manuscript, however, all objective functions $\mathcal{F}$ satisfying eqs. (4) to (7), will be considered valid *fidelity functions* that can be maximised using the proposed approach.

The fidelity functions can be maximised using gradient-based optimisation schemes, where one needs the gradient of the fidelity function $\mathcal{F}$. We compute the gradient using the chain rule,

$$\frac{\partial \mathcal{F}(\theta)}{\partial \theta} = \mathbf{D}f(\mathbf{U}(T;\theta))\frac{\partial \mathbf{U}(T;\theta)}{\partial \theta}, \tag{12}$$

which further requires computation of the gradient of the propagator $\mathbf{U}(T;\theta)$. Here $\mathbf{D}f(X)$ is the Fréchet derivative of f at X, which acts linearly on Y, and $\mathbf{D}f(X)Y$ quantifies the rate of change in the *direction* Y. Equation (12) can be generalised further to use Gateuax derivatives (or directional derivatives) (*36, 37*), where instead of $\mathbf{D}f(X)Y$ we write $\mathbf{D}f(X;Y)$ or $\mathbf{D}_Y f(X)$.

For instance, the action of the Fréchet derivative of the functional eq. (8) is

$$\mathbf{D}f(X)Y = \mathrm{Re}\left[\mathrm{Tr}\left(\varrho^{\dagger}Y\rho_0\right)\right],$$

so that the gradient of the fidelity eq. (9) is

$$\frac{\partial \mathcal{F}(\theta)}{\partial \theta} = \mathrm{Re}\left[\mathrm{Tr}\left(\varrho^{\dagger}\frac{\partial \mathbf{U}(T;\theta)}{\partial \theta}\rho_0\right)\right] = f\left(\frac{\partial \mathbf{U}(T;\theta)}{\partial \theta}\right). \tag{13}$$

The exact propagator $\mathbf{U}(T;\theta)$ (and hence its derivatives) are not available since an analytical solution of eq. (3) is not available except in very specialised and restrictive circumstances. Instead, in practice, we rely on numerical methods for solving eq. (3).

We assume that a family of numerical solvers $\mathcal{S}_{(1)}, \mathcal{S}_{(2)}, \ldots, \mathcal{S}_{(L)}$ to approximate the propagator $\mathbf{U}(T;\theta)$ is available and arranged in increasing accuracy and cost, i.e. we assume that the solver $\mathcal{S}_{(\ell+1)}$ is more accurate and computationally more expensive than $\mathcal{S}_{(\ell)}$. Here $\mathcal{S}_{(\ell)}$ is a (θ-dependent) linear map from the initial state ρ_0 to the state at time T,

$$\rho^{(\ell)}(T;\theta) = \mathcal{S}_{(\ell)}(\theta)\rho_0$$

and it produces an approximation $\rho^{(\ell)}(T;\theta)$ to $\rho(T;\theta)$, the true solution of eq. (1).

The use of a particular propagator $\mathcal{S}_{(\ell)}$ naturally leads to the computation of an approximate value of the fidelity function

$$\mathcal{F}_{(\ell)}(\theta) := f(\mathcal{S}_{(\ell)}(\theta)) \;\; \approx \;\; f(\mathbf{U}(T;\theta)) = \mathcal{F}(\theta).$$

When a range of numerical solvers with different cost and accuracy trade-offs are available, a natural question to ask is: What is the best choice of solver? On the one hand, the most inexpensive solver $\mathcal{S}_{(1)}$ makes the optimisation fast. However, it also leads to a less accurate approximation to the fidelity, $\mathcal{F}_{(1)}(\theta)$. This becomes particularly important in applications where high fidelities (approaching 1) are feasible and desired, and using low accuracy approximations limits the fidelity achievable. In such cases, the most accurate solver $\mathcal{S}_{(L)}$ seems appealing. However, such a solver typically comes with a high computational cost. The use of such an accurate but costly solver seems less justified in the initial stages of optimisation where the objective is very far from the optimal value.

In this paper we propose an adaptive procedure where inexpensive solvers with low accuracy are used far from the optimum, and high accuracy solvers with large computational costs are utilised only closer to the optimum θ^*. The framework requires the following ingredients. **(i)** A set of numerical solvers $\mathcal{S}_{(1)}, \mathcal{S}_{(2)}, \ldots, \mathcal{S}_{(L)}$ with different cost and accuracy trade-offs, **(ii)** Method for computing gradients of these solvers with respect to θ for approximation of the fidelity gradient eq. (12),and **(iii)** a low-cost, adaptive strategy to switch from a solver $\mathcal{S}_{(\ell)}$ to a more accurate solver $\mathcal{S}_{(\ell+1)}$, that includes an error estimation of the computed fidelity $\mathcal{F}_{(\ell)}$ and a measure of proximity to the optimum θ^*.

2.2 Termination criteria

In theory, a nonlinear optimisation process is terminated when the distance of the j^{th} iterate $\theta^{(j)}$ from the optimum θ^* becomes sufficiently small. In practice, however, since θ^* is not available a-priori, we estimate $\|\theta^* - \theta^{(j-1)}\|$ by $\|\theta^{(j)} - \theta^{(j-1)}\|$ and terminate iterations when

$$\|\theta^{(j)} - \theta^{(j-1)}\| \quad \leq \quad \mathrm{tol}_\theta, \tag{14}$$

for some user defined tolerance tol_θ.

Another criterion is motivated by the fact that at critical points (including, but not limited to, the optimum), the gradient vanishes, e.g. $\nabla_\theta \mathcal{F}(\theta^*) = 0$. Again, some user defined tolerance tol_g is used to asses this criterion as

$$\|\nabla_\theta \mathcal{F}(\theta^{(j)})\| \quad \leq \quad \mathrm{tol}_g. \tag{15}$$

The last criterion to consider is the difference of the value of the objective $\mathcal{F}$ at the previous iterate $\theta^{(j-1)}$ from the optimal value $\mathcal{F}(\theta^*)$,

$$|\mathcal{F}(\theta^*) - \mathcal{F}(\theta^{(j)})|. \tag{16}$$

However, we know that when a system is fully controllable, given a set of parameters, a fidelity of 1 can be achieved, i.e. $\mathcal{F}(\theta^*) = 1$, and therefore, for some user-defined tolerance $\mathrm{tol}_{\mathcal{F}}$,

$$|1 - \mathcal{F}(\theta^{(j)})| \quad \leq \quad \mathrm{tol}_{\mathcal{F}} \tag{17}$$

can be used alongside eq. (14) and eq. (15) as a termination criterion. Note that even if we do not know whether a fidelity of 1 is achievable in a particular application, in quantum optimal control we are guaranteed that $\mathcal{F} \leq 1$ (*38*) for suitably normalised initial and target states. Thus, eq. (17) does not lead to unnecessarily early termination and should be used alongside eq. (14) and eq. (15).

2.3 The adaptive procedure

An additional and important aspect in an adaptive process is the numerical error in the approximation of the objective $\mathcal{F}$. We can, at best, rely on $\mathcal{F}_{(L)}$, generated using the most accurate numerical solver $\mathcal{S}_{(L)}$ available to us, but we would like to use the least expensive solvers $\mathcal{S}_{(1)}$, wherever possible. Here we consider the *true fidelity* $\mathcal{F}(\theta^{(j)})$, a fidelity measure that would results from exact calculation methods, and $\mathcal{F}_{(\ell)}(\theta^{(j)})$, a fidelity when the numerical solver $\mathcal{S}_{(\ell)}$ is being used in the j^{th} iteration. For ease of notation, we have suppressed the dependence on the parameters θ in this section. Using eqs. (16) and (17) we can construct a general system of inequalities for the adaptive procedure.

$$|1 - \mathcal{F}| \quad \leq \quad |1 - \mathcal{F}_{(\ell)}| \; + \; |\mathcal{F}_{(\ell)} - \mathcal{F}| \quad \leq \quad (1 + \kappa_{\mathcal{F}})|1 - \mathcal{F}_{(\ell)}| \quad \leq \quad \mathrm{tol}_{\mathcal{F}}, \tag{18}$$

where $\kappa_{\mathcal{F}} \in (0, 1)$ is a user-defined parameter. This system of inequalities enforces the termination criteria such that

$$|1 - \mathcal{F}_{(\ell)}| \quad \leq \quad \frac{\mathrm{tol}_{\mathcal{F}}}{1 + \kappa_{\mathcal{F}}}, \tag{19}$$

and ensures that

$$|\mathcal{F}_{(\ell)} - \mathcal{F}| \quad \leq \quad \kappa_{\mathcal{F}}|1 - \mathcal{F}_{(\ell)}|. \tag{20}$$

The *true fidelity* $\mathcal{F}$ also satisfies the termination criterion eq. (17). Note that $\kappa_{\mathcal{F}}$ in eq. (20) guarantees that at all times during the optimisation procedure we keep the numerical error several times smaller than $|1 - \mathcal{F}_{(\ell)}|$, which measures the distance from optimal value (assuming $\mathcal{F}(\theta^*) = 1$).

As we approach the optimal value, we need to use more accurate solvers. When eq. (20) is violated, the optimiser switches to a more accurate solver $\mathcal{S}_{(\ell+1)}$. From the left side of eq. (20) the error of the fidelity approximation $\mathcal{F}_{(\ell)}$ using the solver $\mathcal{S}_{(\ell)}$ should be computed

$$\varepsilon^{(\ell)} := |\mathcal{F}_{(\ell)} - \mathcal{F}|,$$

As the true fidelity $\mathcal{F}$ is not available, we use a more accurate scheme for approximating $\mathcal{F}$,

$$\varepsilon^{(\ell_1,\ell_2)} := |\mathcal{F}_{(\ell_1)} - \mathcal{F}_{(\ell_2)}| = 1 - \mathrm{Tr}\left((\rho^{(\ell_1)})^\dagger \rho^{(\ell_2)}\right), \qquad \ell_2 > \ell_1.$$

where $\rho^{(\ell)} = \rho^{(\ell)}(t_f;\theta) = \mathcal{S}_{(\ell)}(\rho_0;\theta)$. The most inexpensive estimate is given by

$$\varepsilon^{(\ell,\ell+1)} := |\mathcal{F}_{(\ell)} - \mathcal{F}_{(\ell+1)}|. \tag{21}$$

The computation of $\varepsilon^{(\ell,\ell+1)}$ as in eq. (21) involves both $\mathcal{S}_{(\ell)}$ and $\mathcal{S}_{(\ell+1)}$, and therefore is more than twice as expensive as using $\mathcal{S}_{(\ell)}$ alone. A simple way to mitigate this problem is to only make the decision to switch from $\mathcal{S}_{(\ell)}$ to $\mathcal{S}_{(\ell+1)}$ once every 5 to 10 iterations, which reduces the overall burden of computing $\varepsilon^{(\ell,\ell+1)}$. However, the gap between tests should not be too large, to avoid the fidelity deviating from the *true fidelity*. Although other methods like defect-based error estimators (*39*) can provide a highly accurate estimate of the error in $\mathcal{S}_{(\ell_1)}$ without the use of a more accurate scheme $\mathcal{S}_{(\ell_2)}$ ($\ell_2 > \ell_1$), and can be easily incorporated in our framework, as our current application does not require a very high accuracy of error estimation, we restrict the present implementation to the estimator presented in eq. (21).

2.4 QOALA

We start with presenting the total Hamiltonian in terms of interaction ($\mathcal{H}_{\text{in}}$) and single spin ($\mathcal{H}_{\text{ss}}$) sub-Hamiltonians,

$$\mathcal{H} = \mathcal{H}_{\text{ss}} + \mathcal{H}_{\text{in}}. \tag{22}$$

Therefore $-i\delta t\mathcal{H} = \mathcal{A} + \mathcal{B}$ such that

$$\mathcal{A} = -i\delta t\mathcal{H}_{\text{in}}, \qquad \mathcal{B} = -i\delta t\mathcal{H}_{\text{ss}}. \tag{23}$$

where δt is a small change in t, derived from the differential in eq. (3).

The Hamiltonian for an isolated (non-interacting) spin can be written as

$$h = a\sigma_{\text{x}} + b\sigma_{\text{y}} + c\sigma_{\text{z}}, \tag{24}$$

where σ_α, $\alpha \in \{\text{x},\text{y},\text{z}\}$ are Pauli matrices for spin$-\frac{1}{2}$, (or Pauli matrices equivalent for spins $> \frac{1}{2}$). For M spins a general form of $\mathcal{H}_{\text{ss}}$ can expressed as

$$\begin{aligned}\mathcal{H}_{\text{ss}} &= \sum_{k=1}^{M} \underbrace{\mathbb{1}\otimes\cdots\otimes\mathbb{1}}_{M-k \text{ times}} \otimes h_k \otimes \underbrace{\mathbb{1}\otimes\cdots\otimes\mathbb{1}}_{k-1 \text{ times}} \\ &= \left[\bigoplus_{k=1}^{M} h_k\right].\end{aligned} \tag{25}$$

Any other parts of the Hamiltonian, e.g interactions, can also be included in $\mathcal{H}_{\text{in}}$. Thanks to utilisation of super-operator formalism (Liouville space) in eq. (1), $\mathcal{H}_{\text{in}}$ can contain additional terms such as decoherence via Lindblad or Redfield equations. While $\exp(\mathcal{A})$ can be computed once prior to the optimisation, due to properties of the Kronecker sum, $\exp(\mathcal{B})$ can be expressed as

$$\exp(\mathcal{B}) = \exp\left(-i\delta t h_1\right) \otimes \exp\left(-i\delta t h_2\right) \otimes \cdots \otimes \exp\left(-i\delta t h_M\right). \tag{26}$$

As it will become clear in the next section these decomposition and separation play a key role in the mechanism of the proposed method. Although the theoretical basis of this work can be extended to any quantum systems, it is formulated and demonstrated in the context of nuclear magnetic resonance (NMR) in Section 4.

In principle, one can employ any set of solvers for the approximation of the propagator, ordered in ascending accuracy and cost. In this work we restrict our attention to a range of splitting methods, which are numerical solvers that approximate $\mathrm{e}^{\mathcal{A}+\mathcal{B}}$ in terms of products of $\mathrm{e}^{\mathcal{A}}$ and $\mathrm{e}^{\mathcal{B}}$. The idea of approximating a propagator using splitting techniques is well-explored in the literature (*40–48*). Propagator splittings of arbitrarily high order accuracy can be derived using techniques such as Baker–Campbell–Hausdorff formula (*49*), Zassenhaus splitting (*50*), Magnus expansion (*51*), and their combinations. In what follows, a specified solver using a splitting method of order p, having an accuracy $\mathcal{O}(\delta t^p)$, will be denoted with a subscript $\mathcal{S}_p$.

Moreover, any splitting method $\mathcal{S}_p$ can be combined with with Trotterisation (*40*), which divides a small time slice of duration δt into q equal, smaller, time slices of durations $\frac{\delta t}{q}$ for improved numerical accuracy, to generate a larger group of solvers.

$$\mathcal{S}_{p,q}(\delta t) = \left[S_p\left(\frac{\delta t}{q}\right)\right]^q + \mathcal{O}\left(\frac{\delta t^{p+1}}{q^{p+1}}\right). \tag{27}$$

where the Trotter number q is now also included as a subscript in the solver, $\mathcal{S}_{p,q}$. Splitting methods utilised as solvers in the present work are summarised in Table 1, with a schematic diagram outlining 2nd order splitting, $p = 2$, with Trotter number $q = 3$ shown in Figure 2.

Table 1: Summary of Trotterised splitting methods as solvers to approximate $\mathrm{e}^{\mathcal{A}+\mathcal{B}}$. A pictorial representation of $\mathcal{S}_{2,3}$ is shown in Figure 2. The computational cost, $\mathcal{O}_{\text{cost}}$, is of a forward time-propagation to obtain $\mathbf{U}(T;\theta)$ for a fidelity calculation in eq. (4).

$\mathcal{S}$	**Formula**	$\mathcal{O}_{\text{acc.}}$	$\mathcal{O}_{\text{cost}}$	**Ref.**
$S_{1,q}$	$\left(\mathrm{e}^{\frac{1}{q}\mathcal{A}}\mathrm{e}^{\frac{1}{q}\mathcal{B}}\right)^q$	$\frac{\delta t}{q}$	$2q$	(*40*)
$S_{2,q}$	$\left(\mathrm{e}^{\frac{1}{2q}\mathcal{A}}\mathrm{e}^{\frac{1}{q}\mathcal{B}}\mathrm{e}^{\frac{1}{2q}\mathcal{A}}\right)^q$	$\frac{\delta t^2}{q^2}$	$2q+1$	(*41, 42*)
$S_{3,q}$	$\left(\mathrm{e}^{\frac{\bar{a}_1}{q}\mathcal{A}}\mathrm{e}^{\frac{b_1}{q}\mathcal{B}}\mathrm{e}^{\frac{\bar{a}_2}{q}\mathcal{A}}\mathrm{e}^{\frac{b_2}{q}\mathcal{B}}\mathrm{e}^{\frac{a_2}{q}\mathcal{A}}\mathrm{e}^{\frac{b_1}{q}\mathcal{B}}\mathrm{e}^{\frac{a_1}{q}\mathcal{A}}\right)^q$	$\frac{\delta t^3}{q^3}$	$6q+1$	(*48*)[a]
$S_{4,q}$	$\left(\mathrm{e}^{\frac{a_1}{q}\mathcal{A}}\mathrm{e}^{\frac{b_1}{q}\mathcal{B}}\mathrm{e}^{\frac{a_2}{q}\mathcal{A}}\mathrm{e}^{\frac{b_2}{q}\mathcal{B}}\mathrm{e}^{\frac{a_3}{q}\mathcal{A}}\mathrm{e}^{\frac{b_3}{q}\mathcal{B}}\mathrm{e}^{\frac{a_4}{q}\mathcal{A}}\mathrm{e}^{\frac{b_3}{q}\mathcal{B}}\mathrm{e}^{\frac{a_3}{q}\mathcal{A}}\mathrm{e}^{\frac{b_2}{q}\mathcal{B}}\mathrm{e}^{\frac{a_2}{q}\mathcal{A}}\mathrm{e}^{\frac{b_1}{q}\mathcal{B}}\mathrm{e}^{\frac{a_1}{q}\mathcal{A}}\right)^q$	$\frac{\delta t^4}{q^4}$	$12q+1$	(*43, 47*)[b]

[a] $a_1 = \frac{13}{126} - i\frac{\sqrt{59/2}}{63}$, $a_2 = \frac{25}{63} + i\frac{5\sqrt{59/2}}{126}$, $b_1 = \frac{3}{10}$, $b_2 = \frac{2}{5}$

[b] $a_1 = 0.0792036964311957$, $b_1 = 0.209515106613362$
$a_2 = 0.353172906049774$, $b_2 = -0.143851773179818$
$a_3 = -0.0420650803577195$, $b_3 = 1/2 - b_1 - b_2$
$a_4 = 1 - 2\left(a_1 + a_2 + a_3\right)$,

Let us recall the definition of fidelity in eq. (4)

$$\mathcal{F}(\theta) = f(\mathbf{U}(T;\theta)) \quad \in [0,1],$$

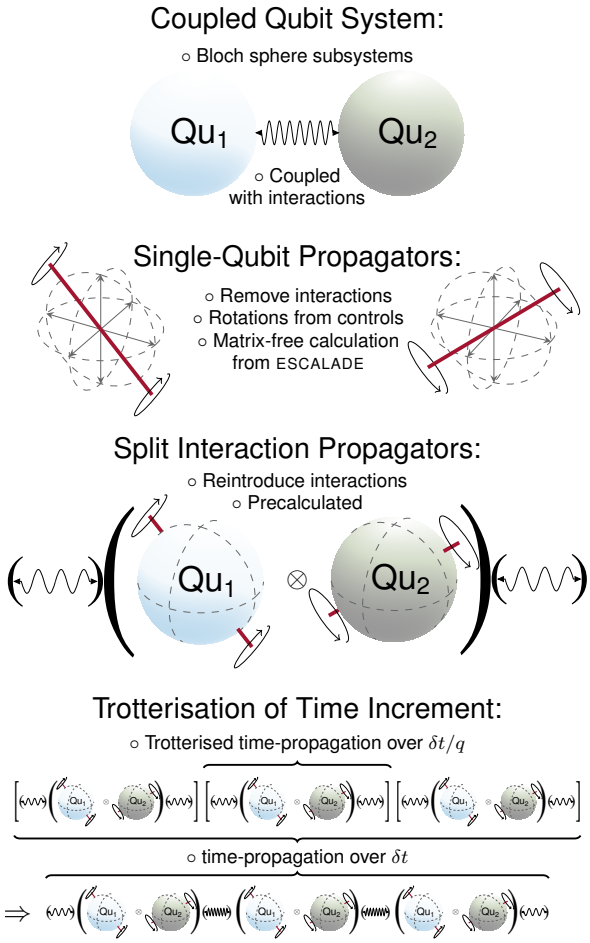

Figure 2: **Pictorial schematics of time-propagator splitting and Trotterisation.** A schematic diagram of a coupled 2-qubit system, an angle-axis representation of ESCALADE (*52*), a pictorial representation of 2nd order operator splitting, and Trotterisation of time-slices with $q = 3$. This corresponds to $\mathcal{S}_{2,3}$ in Table 1. The final image shows a Trotterised, split operator, time-propagation with the left-most and right-most split coupling of the operator splitting method, allowable for $\mathcal{S}_{2,q}$, $\mathcal{S}_{3,q}$, and $\mathcal{S}_{4,q}$ in Table 1, combined into one exponential.

which is the objective function to be minimised in the optimisation.

Although the scope of the QOALA method is not limited to the GRAPE algorithm (*24*), in this paper it is presented in the context of the GRAPE method, where the time-dependent control pulses are generally presented as piecewise-constant over a small time interval δt

$$\theta_{a,k}(t) \rightarrow \begin{bmatrix} \theta_{a,k,1} & \theta_{a,k,2} & \cdots & \theta_{a,k,N} \end{bmatrix},$$

with $a \in \{\mathrm{x}, \mathrm{y}\}$ controlling the k^{th} spin. For notational convenience, the explicit time interval dependence of $\theta_{a,k,n}(\delta t)$ has been dropped. In this context, the effective propagator over the

total pulsing time duration, $T = N\delta t$ is given by

$$\mathbf{U}(T;\theta) = \mathbf{U}_{\text{tot}} := \mathbf{U}_N \mathbf{U}_{N-1} \cdots \mathbf{U}_2 \mathbf{U}_1, \qquad \text{with} \quad \mathbf{U}_n = \mathrm{e}^{-i(\delta t)\mathcal{H}(\theta_n)}, \tag{28}$$

where θ_n is a general control parameter solely affecting the n^{th} propagator, $\mathbf{U}_n$, and the shorthand $\theta_n = \theta_{a,k,n}$ is valid for any $a \in \{\mathrm{x}, \mathrm{y}\}$ and $k \in \{1, \ldots, M\}$. Note that for pulses that are not piecewise constant, the decomposition eq. (28) does not hold exactly. However, similar decompositions can be obtained for arbitrary pulses using the Magnus expansion (*51*), for instance.

The gradient of the true fidelity, i.e. the derivative of the fidelity at every time point n with respect to a discrete control, θ_n, can be written as

$$\frac{\partial \mathcal{F}(\theta)}{\partial \theta_n} = \mathbf{D}f(\mathbf{U}_{\text{tot}}) \frac{\partial \mathbf{U}_{\text{tot}}}{\partial \theta_n}, \tag{29}$$

with

$$\frac{\partial \mathbf{U}_{\text{tot}}}{\partial \theta_n} = \mathbf{U}_N \cdots \mathbf{U}_{n+1} \frac{\partial \mathbf{U}_n}{\partial \theta_n} \mathbf{U}_{n-1} \cdots \mathbf{U}_1. \tag{30}$$

However, due to computational costs, we never compute the true fidelity eq. (4) or its gradient eq. (29). The central idea presented here relies on the approximation of $\mathbf{U}_n = \mathrm{e}^{-i(\delta t)\mathcal{H}(\theta_n)}$ using a solver $\mathcal{S}_{(\ell),n}$, and of the propagator $\mathbf{U}(T;\theta) = \mathbf{U}_{\text{tot}}$ by $\mathcal{S}_{(\ell)}$,

$$\mathcal{S}_{(\ell)} := \mathcal{S}_{(\ell),N}\, \mathcal{S}_{(\ell),N-1}\, \cdots\, \mathcal{S}_{(\ell),2}\, \mathcal{S}_{(\ell),1}, \qquad \text{with} \quad \mathcal{S}_{(\ell),n} \approx \mathbf{U}_n, \tag{31}$$

Thus, instead of eq. (4), we compute the approximate fidelity,

$$\mathcal{F}_{(\ell)}(\theta) = f(\mathcal{S}_{(\ell)}(\theta)), \tag{32}$$

and the exact gradient of the approximate fidelity eq. (32),

$$\frac{\partial \mathcal{F}_{(\ell)}(\theta)}{\partial \theta_n} = \mathbf{D}f(\mathcal{S}_{(\ell)}(\theta)) \frac{\partial \mathcal{S}_{(\ell)}(\theta)}{\partial \theta_n}, \tag{33}$$

where

$$\frac{\partial \mathcal{S}_{(\ell)}(\theta)}{\partial \theta_n} = \mathcal{S}_{(\ell),N}\, \cdots\, \mathcal{S}_{(\ell),n+1}\, \frac{\partial \mathcal{S}_{(\ell),n}}{\partial \theta_n}\, \mathcal{S}_{(\ell),n-1}\, \cdots\, \mathcal{S}_{(\ell),1}. \tag{34}$$

As a concrete example of approximate solvers, $\mathcal{S}_{(\ell)}$, we consider the use of the splitting propagators introduced in Table 1. The n^{th} propagator $\mathbf{U}_n$ is approximated by a split propagator, $\mathcal{S}_{p,q}(\theta_n) = (\mathcal{S}_p(\theta_n))^q$. For ease of notation, we drop the dependency on θ_n. Assuming an odd number of multiplications, $2P+1$, in a general solver, $\mathcal{S}_p = \mathbf{S}_{2P+1}\mathbf{S}_{2P} \cdots \mathbf{S}_2\mathbf{S}_1$ (making a distinction between the solver, $\mathcal{S}_p$, and one of its constituent matrix exponentials, $\mathbf{S}$), and by considering that $\mathcal{A}$ and $\mathcal{B}$ appear in odd and even $\mathbf{S}$ terms respectively (without loss of generality, as presented in Table 1), i.e.

$$\mathbf{S}_{2j} = \mathrm{e}^{b_{2j}\mathcal{B}}, \quad \mathbf{S}_{2j+1} = \mathrm{e}^{a_{2j+1}\mathcal{A}}, \tag{35}$$

and given the fact that $\frac{\partial \mathcal{A}}{\partial \theta_n} = \mathbf{0}$, $\frac{\partial \mathcal{B}}{\partial \theta_n} \neq \mathbf{0}$ and $\left[\frac{\partial \mathcal{B}}{\partial \theta_n}, \mathcal{B}\right] \neq \mathbf{0}$, for a general solver, $\mathcal{S}_{p,q}$, with Trotter number q we have

$$\frac{\partial \mathcal{S}_{p,q}}{\partial \theta_n} = \sum_{i=1}^{q} (\mathcal{S}_p)^{i-1} \frac{\partial \mathcal{S}_p}{\partial \theta_n} (\mathcal{S}_p)^{q-i}, \tag{36}$$

where, using $\mathcal{S}_p = \mathbf{S}_{2P+1}\mathbf{S}_{2P}\cdots\mathbf{S}_2\mathbf{S}_1$,

$$\frac{\partial \mathcal{S}_p}{\partial \theta_n} = \sum_{j=1}^{P} \left[\left(\prod_{i=2P+1}^{2j+1} \mathbf{S}_i \right) \frac{\partial \mathbf{S}_{2j}}{\partial \theta_n} \left(\prod_{i=2j-1}^{1} \mathbf{S}_i \right) \right]. \tag{37}$$

$\frac{\partial \mathbf{S}_{2j}}{\partial \theta_n}$ can be computed using finite difference (*53*) or exact (*32*) methods. While finite difference approximations of the gradient are easy to programme, they can also be inaccurate, expensive and, most importantly, in the present context, unstable when the underlying numerical propagator has low accuracy (*52, 54*). Here, taking advantage of the decomposition presented in eq. (26) we use analytical Lie algebraic derivatives using the recent ESCALADE (Efficient Spin Control using Analytical Lie Algebraic Derivatives) method (*52*), which, in contrast to the auxiliary matrix method (*31, 32, 55–58*), does not require the computation of an expensive matrix exponential, resulting in an additional significant speedup.

An alternative approach to error estimation presented in Section 2.3 is benchmarking of the solvers prior to the optimisation. For any solver in the set of solvers, the achievable accuracy is known, and since the dimensions of the Hamiltonian do not change in the course of the optimisation, we can consider the number of matrix-vector multiplications as a measure for cost. Pre-computed benchmarking consists of identifying switch points on a cost vs. accuracy plot that minimises the overall cost while maximising the accuracy.

If we want to apply this benchmarking strategy to the set of solvers in Table 1 it can be seen that with efficient matrix caching, the minimum cost of the computation of $\mathcal{S}_p = \mathbf{S}_{2P+1}\mathbf{S}_{2P}\cdots\mathbf{S}_2\mathbf{S}_1$ and its derivative in the presence of q^{th} order Trotterisation is $(4+K)Pq+2$, where K is the number of controls. Using these known attributed accuracy and cost, we can obtain a generic benchmarking graph for our set of solvers (Figure 3). This plot, calculated for $\delta t = 0.1$ ms and 4 controls, shows that there is no benefit in using 2^{nd} and 3^{rd} order solvers with Trotter numbers larger than 2 and 1 respectively.

2.5 Numerical examples

The state-transfer problem is investigated for two different spin systems: a 2-spin system, labelled system-1 in table 2 and inset in Figure 4(**D**); a 3-spin system, labelled system-2 in table 2 and inset in Figure 4(**E**). The 2-spin and 3-spin systems are both heteronuclear with $2M$ controls (x and y controls on each spin) and pulse B_1 amplitude limited to 1 kHz.

Heteronuclear 2-spin and 3-spin systems are tasked with transferring an initial z-magnetisation from the first spin, to a desired z-magnetisation at the final spin. The 2-spin system is optimised

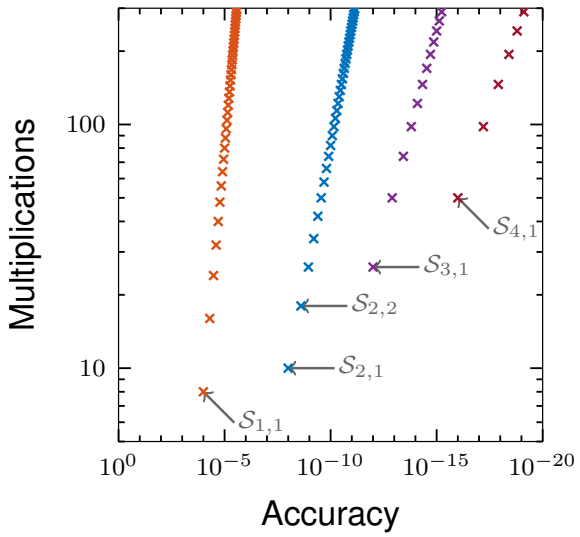

Figure 3: **Benchmarking graph showing the computational cost of different Trotterised time-propagator splitting methods.** The plot shows the number of multiplications per time-slice required of a fidelity and gradient calculation for a desired accuracy, for $\delta t = 0.1$ ms and 4 controls, as a measure of the computational cost of the set of solvers presented in Table 1; arrows and $\mathcal{S}_{p,q}$ terms represent switching points between members of the solver set.

over a pulse duration of $T = 10$ ms and the 3-spin system is optimised over $T = 22$ ms, both systems are allowed a time-slice of $\delta t = 0.2$ ms.

The optimal control task of a 4-spin system is purposefully made more difficult: the initial state is set as the entangled multi-spin state defined in eq. (45); a relaxation term is included in the uncontrollable part of the Hamiltonian. The relaxation term is calculated using *Spinach* (*31, 33, 59, 60*) Bloch-Redfield-Wangsness relaxation theory with isotropic rotational diffusion (correlation time $\tau_c = 50$ ps). The system, labelled system-3 in table 2 and inset in Figure 4(**F**), is a mixture of heteronuclear and homonuclear systems and is controlled with 4 controls (x and y controls on each type of spin), limited to 10 kHz pulses, optimised over a pulse duration of $T = 100$ ms with a time-slice width of $\delta t = 0.02$ ms.

Essentially, there is the requirement that there are enough controls (enough time-slices) to control the system to its desired target. This is a requirement for both an exact method and the solver based methods, $\mathcal{S}_{p,q}$, presented in this work. Accordingly, the control problems are designed to have enough time-slices to attain good convergence, but no more.

The convergence of the 2-spin system follows a typical quadratic convergence, which can be seen from the average convergence of the exact method (dashed black lines) and in Figure 4 (**A**). The interquartile range is also shown (solid grey area bounded by thin dashed lines), with small deviations from the average indicating a predictable/robust optimisation. This can be considered an *easy* optimal control problem. However, at the time-slice width of $\delta t = 0.2$ ms,

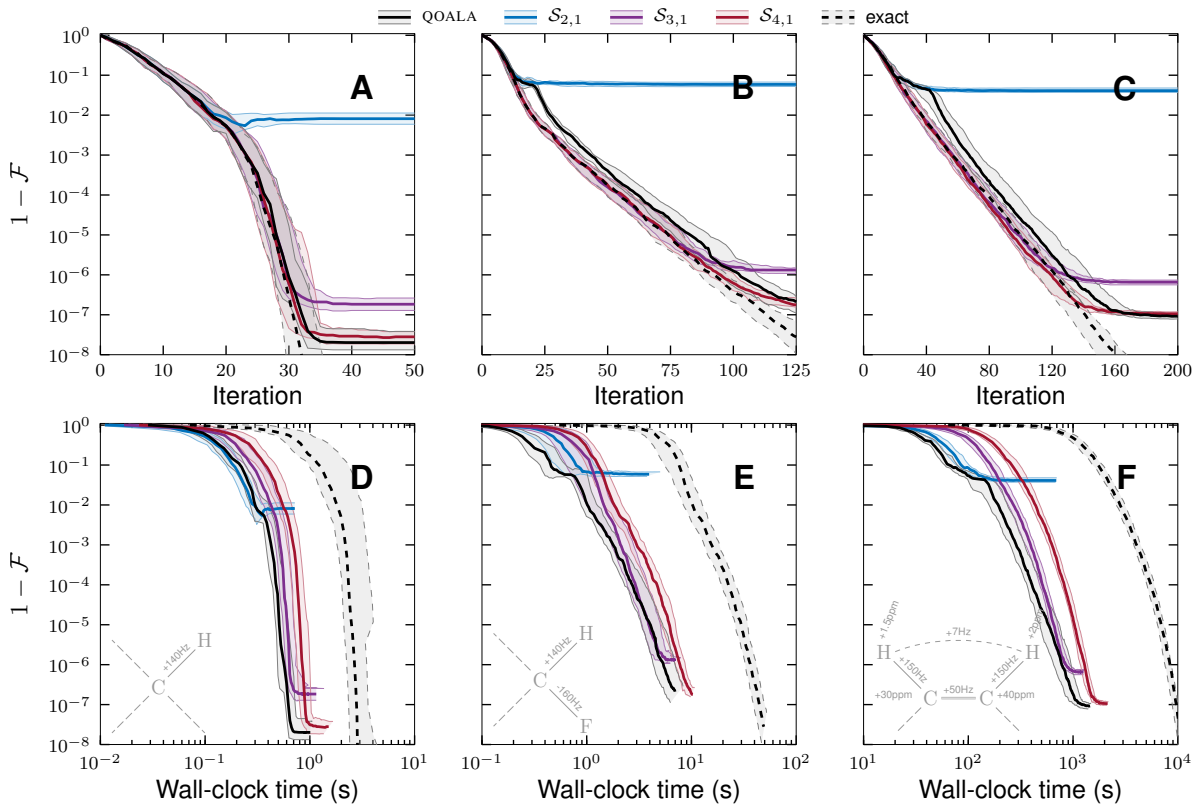

Figure 4: **Computation time of state-transfer optimal control problems.** Comparison of the convergence of the 2-spin, (**A**), 3-spin, (**B**), and 4-spin, (**C**), state transfers with solvers $\mathcal{S}_{2,1}$, $\mathcal{S}_{3,1}$, $\mathcal{S}_{4,1}$ (solid coloured) and QOALA with solver $\mathcal{S}_{\{2,3,4\},1}$ (solid black) to an exact method (dashed black). The lower plots, (**D**-**F**), are the corresponding wall-clock time of computation. The 2-spin system in (**A** and **D**) is given a pulse duration of $T = 10$ ms and a time-slice width of $\delta t = 0.2$ ms. The 3-spin system in (**B** and **E**) is given a pulse duration of $T = 22$ ms and a time-slice width of $\delta t = 0.2$ ms. The 4-spin system in (**C** and **F**) is given a pulse duration of $T = 100$ ms and a time-slice width of $\delta t = 0.02$ ms. In all plots a central thick line is an average over 28 optimisations, and the surrounding shaded area is bound by the first and third quartiles. Adaptive splitting order selection (solid black) allows the optimisation to follow a computationally inexpensive route through a convergence trajectory, while achieving an eventual high fidelity.

the second order splitting solver $\mathcal{S}_{2,1}$ fails to reach a fidelity above 99%. A similar behaviour is observed for the 3-spin and 4-spin problems, when using the $\mathcal{S}_{2,1}$ solver, but with a fidelity limited to approximately 90%. The difficulty of these two optimal control problems is evident from the less-than quadratic convergence of the exact method which, after the initial damped stage of convergence, has a lower acceleration of convergence when compared to Figure 4 (**A**).

Higher order solvers, $\mathcal{S}_{3,1}$ and $\mathcal{S}_{4,1}$, increase the limit on achievable fidelity for all spin systems in Figure 4 with an additional increase in computation time. Taking advantage of the short computation time of low accuracy solvers when allowable, the QOALA algorithm switches to higher accuracy solvers only when needed. Given a switching threshold tolerance of $\kappa_{\mathcal{F}} = 0.5$ in eq. (20) and a switching decision every 10 iterations, the convergence of QOALA in Figure 4 does indeed converge to the high fidelities of the high accuracy solver $\mathcal{S}_{4,1}$ while benefiting from the time savings of the low accuracy solvers at the beginning of the optimisation. Furthermore, the average convergence trajectory of the 2-spin system in Figure 4(**A**) follows the average convergence trajectory of the exact method very closely. The first and third quartiles are also closely aligned, showing both the exact and adaptive methods converge in a predictably quadratic fashion from different, random, initial guesses. From Figure 4(**B** and **C**), the switch to $\mathcal{S}_{3,1}$ looks to have happened a little too late, following $\mathcal{S}_{2,1}$ for too many iterations, and losing a small amount of convergence acceleration in the process. Even considering this small loss in potential time saving, all spin systems achieve an appreciable time saving compared to the exact method. Emphasising this point further, for a 4-spin system, depicted in Figure 4(**F**), the exact method takes over 1.5 hours to achieve a fidelity of 99.99% whereas QOALA takes less than 7 minutes to achieve the same fidelity.

It should be noted that the speedup achieved with the QOALA approach compared to the exact method depends on the type of optimal control problems and the system under study, but we predict that an $\sim \mathcal{O}(M^2)$ speedup is expected for a system consisting of M qubits, as summarised in Figure 1. High order splitting methods are required at the highest fidelities and the computation time of these high order splitting methods dominate the overall computation time. The speedup tends to $\mathcal{O}(M)$ at these high fidelities because of this domination and because there is a splitting order/Trotter number pair that gives the same accuracy as the exact method with a matrix exponential.

The universal swap-gate problem, with the fidelity defined in eq. (46), is investigated for two different heteronuclear spin systems: a 2-spin system, labelled system-1 in table 2 and inset in Figure 5(**C**); a 3-spin system swapping the first and third spin, labelled system-2 in table 2 and inset in Figure 5(**D**). The potential time-saving should be more for this type of optimisation because the efficient Krylov propagation of the exact method in state-transfer problems is not appropriate here, and an explicit matrix exponential must be made in gradient calculations. In this case the exponential of a matrix is evaluated using a Taylor series, with scaling and squaring (*37*), and truncated with a predefined tolerance of 1×10^{-12}. Both problems have $2M$ controls (x and y controls on each spin) with pulse B_1 amplitudes limited to 1 kHz. The adaptivity of the QOALA is set with a switching threshold tolerance of $\kappa_{\mathcal{F}} = 0.5$ (eq. (20)) and a switching decision every 10 iterations. In the context of our set of examples, we observed that universal swap-gate problems require slightly longer pulses than the state-transfer problems. The 2- and 3-spin systems are optimised with $T = 12$ ms and $T = 26.4$ ms, respectively. An example of the effect of a 3-spin swap-gate is shown in Figure 6, showing the first and thirds spins swap states, and the state of the second spin is left unchanged.

In comparison to the state-transfer problems in Figure 4, the inclusion of Trotterisation ($\mathcal{S}_{2,2}$)

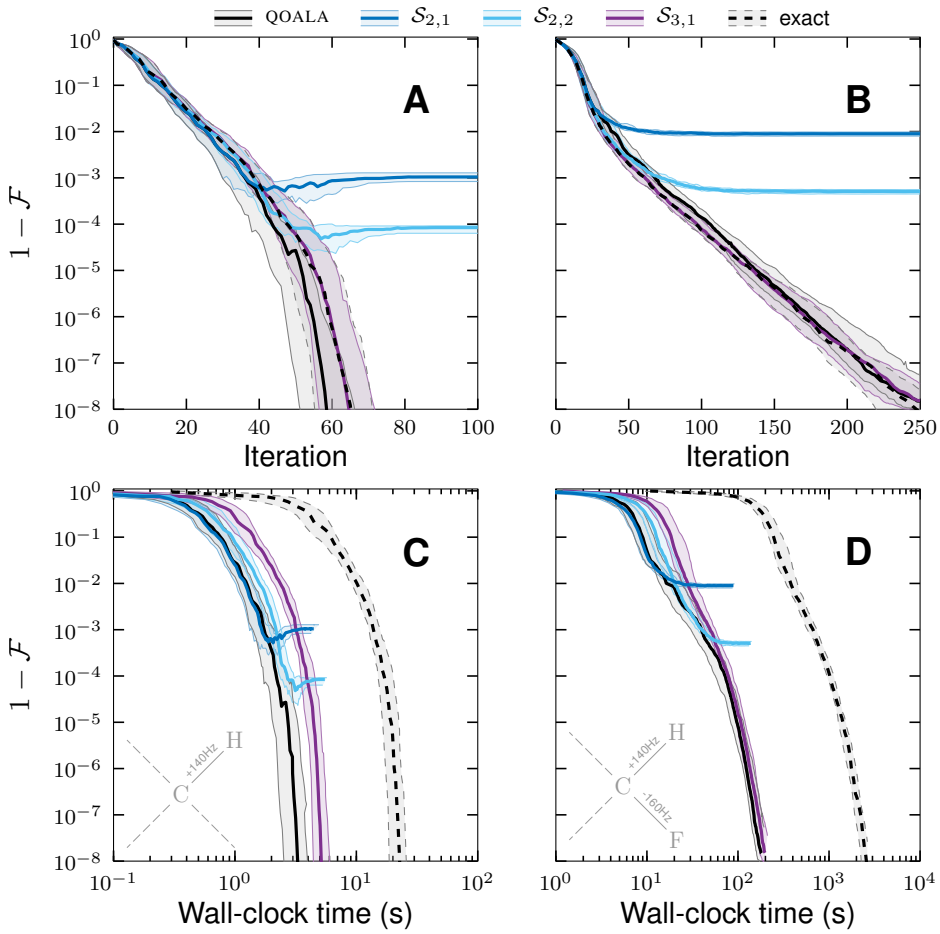

Figure 5: **Computation time of universal swap-gate optimal control problems.** Comparison of the convergence of the 2-spin and 3-spin swap-gate with solvers $\mathcal{S}_{2,1}$, $\mathcal{S}_{2,2}$, $\mathcal{S}_{3,1}$ (solid coloured) and QOALA using those same solvers adaptively (solid black) to an exact method (black dashed) (**A** and **B**) and the corresponding wall-clock time of computation (**C** and **D**). The 2-spin system in (**A** and **C**) is given a pulse duration of $T = 12$ ms and the 3-spin system, swapping the first and third spin, in (**B** and **D**) is given a pulse duration of $T = 26.4$ ms. Both have a time-slice width of $\delta t = 0.1$ ms. In all plots a central thick line is an average over 28 optimisations, and the surrounding shaded area is bound by the first and third quartiles. Adaptive splitting order and Trotter number selection (solid black) allows the optimisation to follow an inexpensive route through a convergence trajectory: Using approximate methods far from the optimum, and accurate methods close to the optimum.

in Figure 5 makes the convergence trajectory more stable by avoiding utilising the solver $\mathcal{S}_{2,1}$ for too many iterations, following the average convergence trajectory of the exact method closely in the case of the 3-spin system in Figure 5(**B**), and even finding a short-cut to convergence for

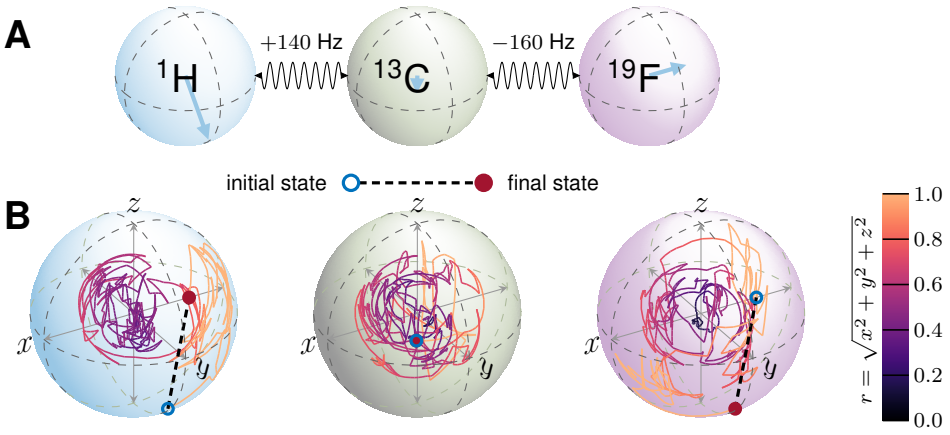

Figure 6: **Bloch sphere trajectories for an optimal universal 3-spin swap-gate.** Analysis of a 3-spin swap-gate from Figure 5(**B**) and (**D**). (**A**) shows the spin system, (**B**) shows the optimised trajectories obtained after 250 iterations of QOALA. The initial and final states indicated in (**B**) show the random states on the first and third spins are swapped, and the random initial state of the second spin is the same as the final state.

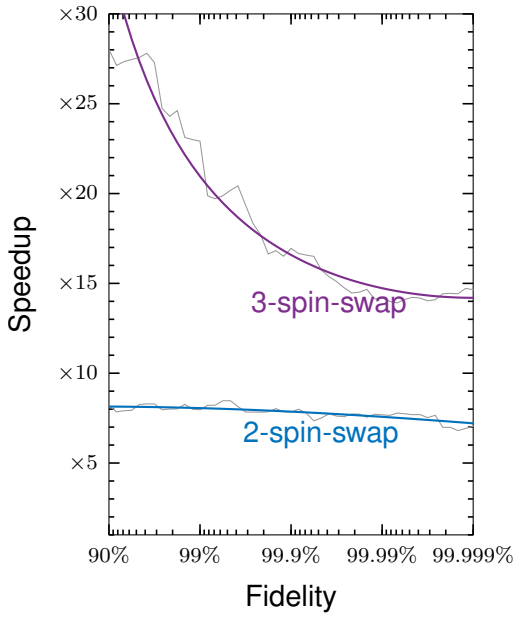

Figure 7: **Speedup of QOALA universal gate optimal control problems.** Speedup in wall-clock time achieved by utilising the QOALA method when compared to an exact method for universal swap-gate optimisations. Smooth, thick lines are a polynomial fit to the thin lines.

the 2-spin system in Figure 5(**A**). Again, both methods converge in a predictable fashion from initial random guesses, shown by the first and third quartiles.

Figure 7 shows the speedup achieved with the QOALA method for these two universal swap-gate optimisations. It shows an even greater (an additional twofold) speedup than for state-transfer problems in Figure 1. This is because the efficiency of Krylov propagation cannot be used for propagator derivative calculations in the exact method. The slowdown from using matrix-matrix products, as opposed to efficient matrix-vector products of the state-transfer problems, is inherited by both QOALA and the exact method.

3 Discussion

We presented a fast, accurate, general, and highly flexible optimal control algorithm, QOALA (Quantum Optimal control by Adaptive Low-cost Algorithm), for solving optimal control problems involving systems of entangled qubits in a computationally efficient manner without sacrificing the desired accuracy of the optimal solution. The QOALA method starts with low-cost, low-accuracy approximations of propagators far from the optimum and adaptively switches to high-cost, high-accuracy approximations as it approaches the optimum. With a set of approximant propagators, the QOALA method offers a versatile trade-off mechanism between accuracy and computational cost. Using a GRAPE-based implementation of the QOALA algorithm and through a diverse set of examples, we demonstrated that it consistently outperforms the state-of-the-art exact methods and we predict $\mathcal{O}(M^2)$ speedup for an M-qubit system, while maintaining the accuracy of the optimal solution. The results of this manuscript can be generalised in a straightforward manner to functionals that involve the values of the density matrix ρ at multiple points, as well as functionals that involve regularisation and penalty terms.

4 Materials and Methods

4.1 Liouville space

In this work, we use the spherical-tensor basis (*61*) in a Liouville space, allowing a general implementation to include effects such as relaxation or decoherence. Using the Liouville-Hilbert dual relationship,

$$\rho(T;\theta) = \mathbf{U}_{\text{tot}}\rho_0\mathbf{U}_{\text{tot}}^{\dagger} \Longleftrightarrow \text{vec}\left(\rho(T;\theta)\right) = \mathbf{U}_{\text{tot}}^{*} \otimes \mathbf{U}_{\text{tot}}\,\text{vec}\left(\rho_0\right), \tag{38}$$

The method is formulated in super-operator formalism

$$\mathcal{H} \mapsto \mathbb{1} \otimes \mathcal{H} - \mathcal{H}^{\top} \otimes \mathbb{1}.$$

4.2 Hamiltonians

Here we present the formulation of our proposed method in the context on NMR for a spin system consisting of M spin-$\frac{1}{2}$ particles interacting with each other via scalar coupling J and under a radio-frequency pulse. The non-interacting, single-spin part of the Hamiltonian can be written as

$$\mathcal{H}_{\mathrm{ss}} = \sum_{k=1}^{M} \theta_{\mathrm{x},k}\mathbf{I}_{\mathrm{x},k} + \sum_{k=1}^{M} \theta_{\mathrm{y},k}\mathbf{I}_{\mathrm{y},k} + \sum_{k=1}^{M} \Omega_k \mathbf{I}_{\mathrm{z},k}, \tag{39}$$

where

$$\mathbf{I}_{a,k} = \underbrace{\mathbb{1} \otimes \cdots \otimes \mathbb{1}}_{M-k \text{ times}} \otimes \sigma_a \otimes \underbrace{\mathbb{1} \otimes \cdots \otimes \mathbb{1}}_{k-1 \text{ times}}, \qquad a \in \{\mathrm{x}, \mathrm{y}, \mathrm{z}\}. \tag{40}$$

Therefore, we can prepare this Hamiltonian for our adaptive approach by

$$\mathcal{H}_{\mathrm{ss}} = \left[\bigoplus_{k=1}^{M} \left(\theta_{\mathrm{x},k}\sigma_x + \theta_{\mathrm{y},k}\sigma_y + \Omega_k \sigma_z \right) \right], \tag{41}$$

where θ_{x} and θ_{y} are the x and y components of the time-dependent pulse amplitudes. In the special case of M homonuclei there are only two controls affecting those M spins, i.e. in eq. (39) $\sum_k \theta_{a,k}\mathbf{I}_{a,k} \to \theta_a \sum_k \mathbf{I}_{a,k}$. The interaction Hamiltonian can be written as

$$\mathcal{H}_{\mathrm{in}} = \sum_{j>k}^{M} 2\pi J_{jk} \Big(\mathbf{I}_{\mathrm{x},j}\mathbf{I}_{\mathrm{x},k} + \mathbf{I}_{\mathrm{y},j}\mathbf{I}_{\mathrm{y},k} + \mathbf{I}_{\mathrm{z},j}\mathbf{I}_{\mathrm{z},k} \Big). \tag{42}$$

In the case of heteronuclear coupling, the spin system can be considered weakly coupled, reducing the interaction Hamiltonian to

$$\mathcal{H}_{\mathrm{in}} = \sum_{j>k}^{M} 2\pi J_{jk} \mathbf{I}_{\mathrm{z},j}\mathbf{I}_{\mathrm{z},k}. \tag{43}$$

4.3 Auxiliary matrix method

One method to find the time-propagators and analytical propagator derivatives is to use an auxiliary matrix method (*31, 32, 55–58*) for computing the matrix exponential of a block triangular matrix:

$$\exp\left\{ \begin{bmatrix} \mathcal{A} + \mathcal{B}_n & \mathcal{C}_{a,k} \\ \mathbf{0} & \mathcal{A} + \mathcal{B}_n \end{bmatrix} \right\} = \begin{bmatrix} \mathbf{U}_n & \dfrac{\partial \mathbf{U}_n}{\partial \theta_{a,k,n}} \\ \mathbf{0} & \mathbf{U}_n \end{bmatrix}, \tag{44}$$

where $\mathcal{C}_{a,k} = -i\delta t \mathbf{I}_{a,k}$ for a control pulse amplitude $\theta_{a,k,n}$ of duration δt. The propagators can be recovered from one of the diagonal blocks, and the propagator derivatives with respect to $\theta_{a,k,n}$, in the direction of the control operator $\mathbf{I}_{a,k}$ (eq. (40)) can be extracted from the upper right block. For state-transfer problems, this method should also use efficient Krylov propagation, avoiding the need to explicitly compute the matrix exponential (*62, 63*).

4.4 Examples

Three types of optimal control problem are compared in this work:

z to z transfer: The goal is to transfer an initial z-magnetisation on a spin j, to z-magnetisation on a spin k: $\mathbf{I}_{z,j} \rightarrow \mathbf{I}_{z,k}$. Magnetisation on each spin can be depicted in a simple way through Cartesian coordinates on each spin's Bloch sphere, as in Figure 2. Here we use the state-to-state transfer fidelity eq. (9), i.e. $\mathcal{F} = \mathrm{Re}[\mathrm{Tr}(\varrho^{\dagger}\mathbf{U}_{\mathrm{tot}}\rho_0)]$.

Entangled to z state transfer: An initial 2-spin entangled state between two spins j and k asks the optimisation to collapse the entanglement to a desired localised z-magnetisation on a further spin l: $\mathbf{E}^{(j,k)} \rightarrow \mathbf{I}_{z,l}$. The normalised, entangled, 2-spin state between spins j and k is defined as

$$\mathbf{E}^{(j,k)} = \frac{1}{2}\mathbb{1} - \Big(2\mathbf{I}_{z,j}\mathbf{I}_{z,k} + \mathbf{I}_{+,j}\mathbf{I}_{-,k} + \mathbf{I}_{-,j}\mathbf{I}_{+,k}\Big), \tag{45}$$

where $\mathbb{1}$ is a unit state, a trivial but complete addition to the basis set, and $\mathbf{I}_{\pm} = \mathbf{I}_{x} \pm \mathrm{i}\mathbf{I}_{y}$. Once again, the fidelity used is eq. (9), $\mathcal{F} = \mathrm{Re}[\mathrm{Tr}(\varrho^{\dagger}\mathbf{U}_{\mathrm{tot}}\rho_0)]$.

Universal swap-gate: A universal multi-spin operation designed to swap the arbitrary initial magnetisation between two spins, j and k, is termed a *swap-gate*: $\mathbf{I}_{(j)} \rightleftharpoons \mathbf{I}_{(k)}$. The swap-gate can be found by optimising to a desired effective propagator, with a fidelity defined through only the effective propagator of the target:

$$\mathcal{F} = \mathrm{Re}[\mathrm{Tr}(\mathbf{U}_{\mathrm{swap}}^{\dagger}\mathbf{U}_{\mathrm{tot}})], \tag{46}$$

where $\mathbf{U}_{\mathrm{swap}}$ is the effective propagator of the swap operation (*64*). This fidelity function is of the form eq. (11), and the overlap is now defined in a more general way by the Hilbert-Schmidt inner product (*65*). The real part disregards a trivial global phase which does not affect the swap-gate operation.

A number of different spin systems are set out in table 2 which are used to investigate convergence and benchmark wall-clock time of QOALA in comparison to exact calculations of propagators and propagator derivatives using the auxiliary matrix method (*31, 58*) of eq. (44). As outlined in the main text, state-to-state simulations for the exact method were performed using the efficient *Spinach* implementation (*33*) of Krylov propagation within MATLAB.

4.5 Optimisation strategy

All results were obtained using a workstation running CentOS 7 with a 14-core Intel Xeon W-2175 CPU @ 2.50GHz and 64GB of RAM. All optimisations were run 28 times, from 28 different random initial pulse guesses, with their convergence trajectories and time of computation (the wall-clock time), averaged to show typical convergence and computation behaviour. The ℓ-BFGS gradient-following optimisation method is used in all cases, with the 25 most recent gradients forming a Hessian approximation (*30*). A bracketing and sectioning line-search method (with cubic interpolation) calculates an appropriate step-length using strong Wolfe conditions (*54*). When an adaptive method switches between solver, a step-length of 1 is

System	k	Isotope	Offset, Ω_k	Coupling, J_{jk}	B_0	B_1	Controls
0	1	^{1}H	$\Omega_1 = +2000$Hz	$J_{12} = +20$Hz	9.4T	1kHz	2
	2	^{1}H	$\Omega_2 = -1200$Hz				
1	1	^{1}H	$\Omega_1 = 0$	$J_{12} = +140$Hz	9.4T	1kHz	4
	2	^{13}C	$\Omega_2 = 0$				
2	1	^{1}H	$\Omega_1 = 0$	$J_{12} = +140$Hz	9.4T	1kHz	6
	2	^{13}C	$\Omega_2 = 0$	$J_{23} = -160$Hz			
	3	^{19}F	$\Omega_3 = 0$				
3	1	^{1}H	$\Omega_1 = -900$Hz	$J_{12} = 150$Hz	14.1T	10kHz	4
	2	^{13}C	$\Omega_2 = -4530$Hz	$J_{23} = 50$Hz			
	3	^{13}C	$\Omega_3 = -1200$Hz	$J_{34} = 150$Hz			
	4	^{1}H	$\Omega_4 = -6040$Hz	$J_{14} = 7$Hz			

Table 2: Spin system parameters for 4 different spin systems, labelled system-0 to system-3 in the first column. The second column refers to a numeric label, k, for each spin in the system, with the isotope of that spin in the third column. The resonance offset (chemical shift), Ω_k, is shown for each spin in Hz in the fourth column. Coupling, J_{jk}, between different spins, labelled j and k, is shown as a list in the fifth column in Hz. The static magnetic field, B_0, is shown in the sixth column, and the nominal power-level of the control pulses, B_1, is shown in kHz in the seventh column. The final column shows the number of control channels. Simulations use a penalty function to ensure control pulses do not exceed to nominal power-level, $\pm B_1$, of the control pulses.

selected, with no further line-search at that iteration. Termination conditions of optimisation are $\mathrm{tol}_\theta = 1 \times 10^{-12}$ in eq. (14) and $\mathrm{tol}_g = 1 \times 10^{-12}$ in eq. (15). All optimal control problems are given sufficient pulse duration to achieve high fidelity (*20*), enforce a pulse amplitude spill-out penalty beyond 1 kHz (or 10 kHz for 4-spin examples) (*32, 58*), and fidelities are normalised so the globally optimal fidelity is unity.

References

1. 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, F. K. Wilhelm, Training schrödinger's cat: quantum optimal control. *Eur. Phys. J. D* **69**, 1–24 (2015).

2. T. E. Skinner, T. O. Reiss, B. Luy, N. Khaneja, S. J. Glaser, Application of optimal control theory to the design of broadband excitation pulses for high-resolution NMR. *J. Magn. Reson.* **163**, 8–15 (2003).

3. Z. Tošner, S. J. Glaser, N. Khaneja, N. C. Nielsen, Effective hamiltonians by optimal control: Solid-state NMR double-quantum planar and isotropic dipolar recoupling. *J. Chem. Phys.* **125**, 184502 (2006).

4. M. S. Vinding, D. Brenner, D. H. Y. Tse, S. Vellmer, T. Vosegaard, D. Suter, T. Stöcker, I. I. Maximov, Application of the limited-memory quasi-newton algorithm for multi-dimensional, large flip-angle RF pulses at 7T. *Magn. Reson. Mater. Phy.* **30**, 29–39 (2017).

5. E. V. Reeth, H. Ratiney, K. Tse Ve Koon, M. Tesch, D. Grenier, O. Beuf, S. J. Glaser, D. Sugny, A simplified framework to optimize MRI contrast preparation. *Magn. Reson. Med.* **81**, 424–438 (2019).

6. P. E. Spindler, Y. Zhang, B. Endeward, N. Gershernzon, T. E. Skinner, S. J. Glaser, T. F. Prisner, Shaped optimal control pulses for increased excitation bandwidth in EPR. *J. Magn. Reson.* **218**, 49–58 (2012).

7. A. Doll, S. Pribitzer, R. Tschaggelar, G. Jeschke, Adiabatic and fast passage ultra-wideband inversion in pulsed EPR. *J. Magn. Reson.* **230**, 27–39 (2013).

8. T. Kaufmann, T. J. Keller, J. M. Franck, R. P. Barnes, S. J. Glaser, J. M. Martinis, S. Han, DAC-board based X-band EPR spectrometer with arbitrary waveform control. *J. Magn. Reson.* **235**, 95–108 (2013).

9. J. Zhang, R. Laflamme, D. Suter, Experimental implementation of encoded logical qubit operations in a perfect quantum error correcting code. *Phys. Rev. Lett.* **109**, 100503 (2012).

10. G. Waldherr, Y. Wang, S. Zaiser, M. Jamali, T. Schulte-Herbrüggen, H. Abe, T. Ohshima, J. Isoya, J. F. Du, P. Neumann, J. Wrachtrup, Quantum error correction in a solid-state hybrid spin register. *Nature* **506**, 204–207 (2014).

11. F. Dolde, V. Bergholm, Y. Wang, I. Jakobi, B. Naydenov, S. Pezzagna, J. Meijer, F. Jelezko, P. Neumann, T. Schulte-Herbrüggen, J. Biamonte, J. Wrachtrup, High-fidelity spin entanglement using optimal control. *Nat. Commun.* **5**, 1–9 (2014).

12. J. C. Saywell, I. Kuprov, D. Goodwin, M. Carey, T. Freegarde, Optimal control of mirror pulses for cold-atom interferometry. *Phys. Rev. A* **98**, 023625 (2018).

13. J. Saywell, M. Carey, I. Kuprov, T. Freegarde, Biselective pulses for large-area atom interferometry. *Phys. Rev. A* **101**, 063625 (2020).

14. J. C. Saywell, M. Carey, M. Belal, I. Kuprov, T. Freegarde, Optimal control of raman pulse sequences for atom interferometry. *J. Phys. B* **53**, 085006 (2020).

15. L. H. Coudert, Optimal control of the orientation and alignment of an asymmetric-top molecule with terahertz and laser pulses. *J. Chem. Phys.* **148**, 094306 (2018).

16. E. Räsänen, A. Castro, J. Werschnik, A. Rubio, E. K. Gross, Optimal control of quantum rings by terahertz laser pulses. *Phys. Rev. Lett.* **98**, 157404 (2007).

17. M. Zhao, D. Babikov, Coherent and optimal control of adiabatic motion of ions in a trap. *Phys. Rev. A* **77**, 012338 (2008).

18. Y. Chou, S. Huang, H. Goan, Optimal control of fast and high-fidelity quantum gates with electron and nuclear spins of a nitrogen-vacancy center in diamond. *Phys. Rev. A* **91**, 052315 (2015).

19. J. Tian, T. Du, Y. Liu, H. Liu, F. Jin, R. S. Said, J. Cai, Optimal quantum optical control of spin in diamond. *Phys. Rev. A* **100**, 012110 (2019).

20. N. Khaneja, S. J. Glaser, R. Brockett, Sub-riemannian geometry and time optimal control of three spin systems: Quantum gates and coherence transfer. *Phys. Rev. A* **65**, 1–11 (2002).

21. B. Bonnard, S. J. Glaser, D. Sugny, A review of geometric optimal control for quantum systems in nuclear magnetic resonance. *Adv. Math. Phys.* **2012**, 1–29 (2012).

22. N. Augier, U. Boscain, M. Sigalotti, Adiabatic ensemble control of a continuum of quantum systems. *SIAM J. Control Optim.* **56**, 4045–4068 (2018).

23. C. Brif, M. D. Grace, M. Sarovar, K. C. Young, Exploring adiabatic quantum trajectories via optimal control. *New J. Phys.* **16**, 065013 (2014).

24. N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, S. J. Glaser, Optimal control of coupled spin dynamics: design of NMR pulse sequences by gradient ascent algorithms. *J. Magn. Reson.* **172**, 296–305 (2005).

25. D. Lu, K. Li, J. Li, H. Katiyar, A. J. Park, G. Feng, T. Xin, H. Li, G. Long, A. Brodutch, J. Baugh, B. Zeng, R. Laflamme, Enhancing quantum control by bootstrapping a quantum processor of 12 qubits. *npj Quantum Inf.* **3**, 1–7 (2017).

26. S. G. Schirmer, P. de Fouquieres, Efficient algorithms for optimal control of quantum dynamics: the krotov method unencumbered. *New J. Phys.* **13**, 073029 (2011).

27. D. M. Reich, M. Ndong, C. P. Koch, Monotonically convergent optimization in quantum control using krotov's method. *J. Chem. Phys.* **136**, 104103 (2012).

28. D. Quiñones Valles, S. Dologov, D. Savostyanov, Tensor product approach to quantum control. *Integral Methods in Science and Engineering*, C. Constanda, P. Harris, eds. (Birkhäuser, Cham, 2019), pp. 367–379.

29. J. H. M. Jensen, F. S. Møller, J. J. Sørensen, J. F. Sherson, Approximate dynamics leading to more optimal control: Efficient exact derivatives. *Phys. Rev. A* **103**, 062612 (2021).

30. P. de Fouquieres, S. G. Schirmer, S. J. Glaser, I. Kuprov, Second order gradient ascent pulse engineering. *J. Magn. Reson.* **212**, 412–7 (2011).

31. D. L. Goodwin, I. Kuprov, Auxiliary matrix formalism for interaction representation transformations, optimal control, and spin relaxation theories. *J. Chem. Phys.* **143**, 084113 (2015).

32. D. L. Goodwin, I. Kuprov, Modified Newton-Raphson GRAPE methods for optimal control of spin systems. *J. Chem. Phys.* **144**, 204107 (2016).

33. H. Hogben, M. Krzystyniak, G. Charnock, P. Hore, I. Kuprov, Spinach – a software library for simulation of spin dynamics in large spin systems. *J. Magn. Reson.* **208**, 179–194 (2011).

34. R. Jozsa, Fidelity for mixed quantum states. *J. Mod. Opt.* **41**, 2315-2323 (1994).

35. Y.-C. Liang, Y.-H. Yeh, P. E. M. F. Mendonça, R. Y. Teh, M. D. Reid, P. D. Drummond, Quantum fidelity measures for mixed states. *Rep. Prog. Phys.* **82**, 076001 (2019).

36. J. Munkres, *Analysis on Manifolds* (CRC Press, 1991).

37. N. J. Higham, *Functions of Matrices: Theory and Computation* (Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008).

38. S. J. Glaser, T. Schulte-Herbrüggen, M. Sieveking, O. Schedletzky, N. C. Nielsen, O. W. Sørensen, C. Griesinger, Unitary control in quantum ensembles: Maximizing signal intensity in coherent spectroscopy. *Science* **280**, 421–424 (1998).

39. W. Auzinger, H. Hofstätter, O. Koch, M. Thalhammer, Defect-based local error estimators for splitting methods, with application to schrödinger equations, part iii: The nonlinear case. *J. Comput. Appl. Math.* **273**, 182–204 (2015).

40. H. F. Trotter, On the product of semi-group operators. *Proc. Amer. Math. Soc.* **10**, 545–551 (1959).

41. G. Strang, On the construction and comparison of different schemes. *SIAM J. Numer. Anal.* **5**, 506–517 (1968).

42. G. I. Marchuk, Some application of splitting-up methods to the solution of mathematical physics problems. *Appl. Math.-Czech.* **13**, 103–132 (1968).

43. S. Blanes, P. C. Moan, Practical symplectic partitioned runge–kutta and runge–kutta–nyström methods. *J. Comput. Appl. Math.* **142**, 313–330 (2002).

44. I. P. Omelyan, I. M. Mryglod, R. Folk, Symplectic analytically integrable decomposition algorithms: classification, derivation, and application to molecular dynamics, quantum and celestial mechanics simulations. *Comput. Phys. Commun.* **151**, 272–314 (2003).

45. R. I. McLachlan, G. R. W. Quispel, Splitting methods. *Acta Numer.* **11**, 341–434 (2002).

46. S. Blanes, F. Casas, A. Murua, Splitting and composition methods in the numerical integration of differential equations. *Bol. Soc. Esp. Mat. Apl.* **45**, 89–145 (2008).

47. A. Iserles, K. Kropielnicka, P. Singh, Compact schemes for laser–matter interaction in schrödinger equation based on effective splittings of magnus expansion. *Comput. Phys. Commun.* **234**, 195–201 (2019).

48. S. Blanes, F. Casas, A. Escorihuela-Tomàs, Applying splitting methods with complex coefficients to the numerical integration of unitary problems. *J. Comput. Dyn.* **9**, 85–101 (2022).

49. J. A. Oteo, The Baker–Campbell–Hausdorff formula and nested commutator identities. *J. Math. Phys.* **32**, 419–424 (1991).

50. P. Bader, A. Iserles, K. Kropielnicka, P. Singh, Effective approximation for the semiclassical Schrödinger equation. *Found. Comput. Math.* **14**, 689–720 (2014).

51. A. Iserles, H. Z. Munthe-Kaas, S. P. Nørsett, A. Zanna, Lie-group methods. *Acta Numerica* **9**, 215–365 (2000).

52. M. Foroozandeh, P. Singh, Optimal control of spins by analytical lie algebraic derivatives. *Automatica J. IFAC* **129**, 109611 (2021).

53. G. Forsythe, W. Wasow, *Finite-difference methods for partial differential equations* (John Wiley and Sons, New York ; London, 1960).

54. J. Nocedal, S. J. Wright, *Numerical Optimization* (Springer, New York, NY, USA, 2006), second edn.

55. C. F. Van Loan, Computing integrals involving the matrix exponential. *IEEE Trans. Automat. Contr.* **23**, 395–404 (1978).

56. I. Najfeld, T. F. Havel, Derivatives of the matrix exponential and their computation. *Adv. Appl. Math.* **16**, 321–375 (1995).

57. F. F. Floether, P. de Fouquieres, S. G. Schirmer, Robust quantum gates for open systems via optimal control: Markovian versus non-markovian dynamics. *New J. Phys.* **14**, 073023 (2012).

58. D. L. Goodwin, Advanced optimal control methods for spin systems, Ph.D. thesis, University of Southampton, UK (2017).

59. I. Kuprov, N. Wagner-Rundell, P. J. Hore, Polynomially scaling spin dynamics simulation algorithm based on adaptive state-space restriction. *J. Magn. Reson.* **189**, 241–250 (2007).

60. I. Kuprov, Diagonalization-free implementation of spin relaxation theory for large spin systems. *J. Magn. Reson.* **209**, 31–38 (2011).

61. D. J. Siminovitch, Rotations in NMR: Part I. Euler-Rodrigues parameters and quaternions. *Concepts Magn. Reson.* **9**, 149–171 (1997).

62. T. J. Park, J. C. Light, Unitary quantum time evolution by iterative lanczos reduction. *J. Chem. Phys.* **85**, 5870-5876 (1986).

63. M. Hochbruck, C. Lubich, On krylov subspace approximations to the matrix exponential operator. *SIAM J. Numer. Anal.* **34**, 1911-1925 (1997).

64. N. Linden, H. Barjat, Ē. Kupče, R. Freeman, How to exchange information between two coupled nuclear spins: the universal SWAP operation. *Chem. Phys. Lett.* **307**, 198–204 (1999).

65. D. L. Goodwin, M. R. M. Koos, B. Luy, Second order phase dispersion by optimised rotation pulses. *Phys. Rev. Research* **2**, 033157 (2020).

Acknowledgements

MF is grateful to the Royal Society for a University Research Fellowship and a University Research Fellow Enhancement Award (URF\R1\180233 and RGF\EA\181018) that have funded this project. DLG would like the thank Burkhard Luy for his continued support.

The idea of QOALA was conceived by MF and PS; the theory evolved and improved in collaboration with DLG. A stable numerical implementation of the method, made available as the QOALA package, was done by DLG. All numerical results and all figures reported in the paper were produced by DLG. The article was written by MF, DLG, and PS.

All authors declare that they have no competing interests.

All data needed to evaluate the conclusions in the paper are present in the paper and are also publicly hosted on `GitHub` (`https://github.com/superego101/qoala/releases/tag/v1.0.0`) and `Zenodo` (`https://doi.org/10.5281/zenodo.6942692`) in the form of a MATLAB toolbox QOALA, which includes functions to produce the results presented in this paper and an additional set of examples.