Learning in quantum control: High-dimensional global optimization for noisy quantum dynamics

Pantita Palittapongarnpim [a,*], Peter Wittek [b,c], Ehsan Zahedinejad [a], Shakib Vedaie [a], Barry C. Sanders [a,d,e,f]

[a] *Institute for Quantum Science and Technology, University of Calgary, Calgary, Alberta T2N 1N4 Canada*
[b] *ICFO – The Institute of Photonic Sciences, Mediterranean Technology Park Av. Carl Friedrich Gauss, 3 08860 Castelldefels, Barcelona, Spain*
[c] *University of Borås SE-501 90 Borås, Sweden*
[d] *Program in Quantum Information Science, Canadian Institute for Advanced Research, Toronto, Ontario M5G 1Z8 Canada*
[e] *Hefei National Laboratory for Physical Sciences at Microscale, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China*
[f] *CAS Center for Excellence and Synergetic Innovation Center in Quantum Information and Quantum Physics, University of Science and Technology of China, Shanghai Branch, Shanghai 201315, People's Republic of China*

ARTICLE INFO

ABSTRACT

Quantum control is valuable for various quantum technologies such as high-fidelity gates for universal quantum computing, adaptive quantum-enhanced metrology, and ultra-cold atom manipulation. Although supervised machine learning and reinforcement learning are widely used for optimizing control parameters in classical systems, quantum control for parameter optimization is mainly pursued via gradient-based greedy algorithms. Although the quantum fitness landscape is often compatible with greedy algorithms, sometimes greedy algorithms yield poor results, especially for large-dimensional quantum systems. We employ differential evolution algorithms to circumvent the stagnation problem of non-convex optimization. We improve quantum control fidelity for noisy system by averaging over the objective function. To reduce computational cost, we introduce heuristics for early termination of runs and for adaptive selection of search subspaces. Our implementation is massively parallel and vectorized to reduce run time even further. We demonstrate our methods with two examples, namely quantum phase estimation and quantum-gate design, for which we achieve superior fidelity and scalability than obtained using greedy algorithms.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Quantum mechanics has been recognized as a superior foundation for performing computation [1–3], secure communication [4,5] and metrology [6,7], also leading to technological advancements such as nuclear magnetic resonance and other resonators [8–12], femtosecond lasers [13,14] and laser-driven molecular reactions [15,16]. Central to these applications is the ability to steer quantum dynamics towards closely realizing specific quantum states or operations; i.e., the ability to control the system [17].

Control theory of classical systems has a long history and is extremely well developed. Control theory most often relies on a mathematical model of a physical system. The primary goal of control is to make the system's dynamics follow a reference trajectory or optimize the dynamics according to an objective function if a reference trajectory is not available. A mathematical model, however, can be difficult to specify exactly or solve analytically. Reinforcement learning is an alternative approach to control that does not necessarily have an explicit mathematical model of the underlying physical system, but rather it optimizes system's behavior by studying responses given a set of inputs [18,19]. If the control provided by the reinforcement learning algorithm is discrete in time, we can view the algorithm as a form of machine learning where the typical assumption of sampling from independent and identical distributions is replaced by that of a Markov decision process.

The purpose of quantum control is identical to the classical case: generate a feasible *policy* for the given control problem. A policy is a set of instructions that determine the control parame-

ters, and hence the effectiveness of the control scheme. This task is complicated by the quantum mechanical nature of the system, which allows non-classical correlations and noncontinuous jumps of the system's state [17]. For control tasks that involve continuous control over quantum states, usually through applying control pulses, algorithms such as GRadient-Ascent Pulse Engineering (GRAPE) have been applied to generate policies. These tasks are found in spectroscopy [9], ultracold-atom research [20,21], and implementation of quantum computation [22]. When feedback is included, such as for adaptive parameter estimation [23–25] and for stabilization of a quantum state [26,27], the dynamics of the state becomes nonlinear and noncontinuous. In this case, the optimization methods account for all allowed trajectories the quantum state takes over the course of the procedure [28–31].

As in the case of classical control, if the mathematical model of the quantum physical system is overly complex or elusive altogether, we can turn to reinforcement learning. For instance, in quantum control problems that involve measurement and feedback, reinforcement learning is gaining attention. Examples include an agent-based model in measurement-based quantum computation [32], mapping quantum gates on a spin system [33], suppressing errors in quantum memory [34], optimization in ultra-cold-atom experiments [35], and earlier work on the adaptive quantum phase estimation problem using heuristic optimization [36,37]. Whereas optimization methods such as Bayesian and Markovian feedback require the knowledge of the system dynamics [30], the machine learning approach [38] enables us to treat the quantum system as a black box. The policy is generated in response to the outcome such that the quantum dynamics closely approximates the target channel of the procedure irrespective of the dynamics involved. This approach has been used successfully to find policies for quantum control problems, such as the classification of qubits and trajectories [39,40].

Greedy algorithms are used for finding successful policies because the algorithms are fast in converging on successful solutions when performing local searches. Greedy algorithms are not guaranteed to succeed (i) when the search domain is non-convex or (ii) when the computational resource or the time for performing the control task is constrained [41]. Early quantum-control schemes employ standard heuristics, such as genetic algorithms, to find successful policies [42,43]. These algorithms fail when the number of control parameters increases or when decoherence and loss are included.

As greedy algorithms fail to provide successful policies for the quantum control problems at hand, and simple heuristics for global optimizations also fail under realistic conditions of quantum systems, we develop new variants of optimization algorithms. We consider in particular differential evolution (DE) as a basis. Our decision is due to the algorithm's superior performance to particle swarm optimization (PSO) and other evolutionary algorithms for high-dimensional optimization problems [44–46].

We demonstrate learning-based quantum-control schemes to find successful policies for two topics relevant to quantum control: phase estimation via adaptive quantum-enhanced metrology (AQEM) and designing fast quantum logic gates. Adaptive phase estimation aims to estimate an unknown phase shift on a light field such that the precision is enhanced by the use of a quantum state of light [29,31,47–49]. Quantum metrology has many applications, such as in atomic clocks and gravitational wave detection. Quantum-gate design tackles the problem of performing a specific gate operation on a quantum bit by steering its evolution [33,50,51]. Fast quantum logic gates are required for designing fast quantum processing units as the timescales are limited by the decoherence time of the qubits. Both problems require optimization of the procedure over the limited resource and time.

Building on the DE-based reinforcement learning and machine learning algorithm, we address critical issues for applying quantum control in realistic physical systems in the following way.

1. We develop a scheme that can operate when practical imperfections such as noise and loss are included. The primary means to this is the way in which the objective function is evaluated.
2. We improve scalability to a higher dimensional search space. For the problem of adaptive phase estimation, scalability is achieved by the accept–reject criterion that allows an early or a late termination of calculations. For the problem of gate design, we devise a subspace-selective self-adaptive DE (SuSSADE) that alternates between a search in the subspace and the overall space while adaptively updates the algorithmic constants of the standard DE algorithm during the search process.
3. Furthermore, we vectorize the time-critical operations to use the parallel resources available efficiently in contemporary CPUs and GPUs.

This paper is structured as follows. In Section 2, we introduce the relevant concepts in quantum mechanics and quantum control. We also explain the control procedures in our two examples: adaptive phase estimation and quantum-gate design. In Section 3, we describe the connection between machine learning, reinforcement learning, and control. In Section 4 we formulate adaptive phase estimation and quantum-gate design as learning problems and show the methods for creating noise-resistant DE and increasing the scalability of our learning algorithms. The results for both control problems are in Section 5.

2. Quantum control framework

Quantum control concerns the application of control procedures to systems whose dynamics are governed by quantum mechanics [17]. In this regime, behaviors that are associated with classical dynamics are violated, leading to challenges in applying classical control theory directly to the system [52,53]. In this section, we explain the key concepts in quantum mechanics that are necessary to understand quantum control, especially to adaptive phase estimation and quantum-gate design. Readers interested in the complete formalism of quantum mechanics are referred to the many publications on this subject [54,55].

2.1. Elements of quantum mechanics

In the regime of quantum mechanics, the state of an isolated particle A is given by a vector $|\psi\rangle_A$ of norm one in a Hilbert space $\mathscr{H}_A$. We restrict our attention to the finite dimensional case where the Hilbert space is $\mathbb{C}^n$. More generally, the state of a particle is a self-adjoint trace class operator of trace one [56], given in the case of the adjoint vector $|\psi\rangle_A$ as $\hat{\rho} = |\psi\rangle_A\langle\psi|$, called a density operator. The density operator can be represented by a matrix given for a chosen basis of the Hilbert space.

Whereas a classical particle has definite values for its characteristics such as position and momentum, a characteristic of a quantum particle can only be described in terms of its probability distribution. Given a chosen basis $\{|i\rangle_A\}$, the state can be represented by $|\psi\rangle_A = \sum_i c_i |i\rangle_A$, where $c_i \in \mathbb{C}$. The absolute square of c_i determines the probability distribution on the chosen basis and, hence, must satisfy the condition $\sum_i |c_i|^2 = 1$. In the matrix representation of the density matrix, the distribution is determined by the diagonal elements, which leads to the trace of one.

When two or more particles exist in the system, correlations can exist that cannot be expressed by any local hidden variable model. For instance, a qubit is a state in $\mathbb{C}^2$, and a two-qubit state

of particles A and B is in $\mathbb{C}_A^2 \otimes \mathbb{C}_B^2$. A quantum correlation between the qubits leads to the state $|\psi\rangle_{AB}$ that cannot be expressed as $|\psi\rangle_A \otimes |\psi\rangle_B$ and a phenomenon known as entanglement where a local operation on the subsystem A affects the state of the subsystem B regardless of the space separation between A and B.

The connection between state and probability distribution becomes apparent when we consider the measurement of the state. Measurements are described by a positive operator-valued measure (POVM) $\{\hat{E}_x\}$ [57] acting on system A, which delivers outcome x with probability $\text{tr}(\hat{E}_x|\psi\rangle_A\langle\psi|)$. POVMs satisfies the condition $\sum_x \hat{E}_x^\dagger \hat{E}_x = \mathbb{1}$, and $\hat{E}_x$ is positive for all x. The state after the measurement is

$$|\psi, x\rangle_A = \frac{\hat{E}_x|\psi\rangle_A}{\sqrt{\left|\hat{E}_x|\psi\rangle_A\right|^2}}. \tag{1}$$

The outcome x is random. Therefore, the evolution of the quantum state due to a measurement is a random jump unless the state is an eigenstate of $\hat{E}_x$.

Manipulation of the state is accomplished by providing outside interaction to the system. This interaction is described by a completely positive trace-preserving (CPTP) map [55], ensuring that the vector remains a quantum state. We refer to these maps as quantum channels $\{\hat{C}_j\}$ that satisfy $\sum_j \hat{C}_j \hat{C}_j^\dagger = \mathbb{1}$. If these linear operators are unitary, the system does not interact with its environment, and its dynamics remains reversible.

As particles, in reality, are not perfectly isolated, the system can be considered as having a constant weak interaction with a bath, which is another quantum system in a larger Hilbert space. The state of the bath is not accessible through measurement, and this loss of information leads to the decoherence of the system's state. Quantum correlations suffer from this kind of interaction because the state turns into a convex combination of the resulting states of all possible interactions between bath and system, $\hat{\rho}_A = \sum_i P_i|\psi^{(i)}\rangle_A\langle\psi^{(i)}|$, where $\{P_i\}$ is a classical probability distribution. The state is no longer pure, but forms a mixed state. In the case of strong system-bath interaction, the system loses the quantum-mechanical characteristics altogether [58].

The essence of quantum control is to steer a quantum channel towards the desired operator. One way to test the channel resulting from the control is to monitor the channel through process tomography [59]. A known quantum state is injected as the input, and the output state is measured. This random measurement outcome is then used to infer the operation performed by the channel. In the process of obtaining a successful policy, the outcome is monitored, and the policy that is used to adjust the control parameters is updated accordingly. The difference between the channel and the target is determined using an objective function. However, if the goal of the control does not explicitly involve the channel, other methods and objective functions can be selected.

In this work, we consider two examples of quantum-control procedures. The first example is the adaptive quantum-enhanced metrology, which utilizes a multiparticle entangled state to attain quantum-enhanced estimation of an unknown parameter and consolidates previous material [60]. The second case study is quantum-gate design, which uses the control to apply logic gates on three quantum bits (qubits) [41].

2.2. Adaptive quantum-enhanced metrology

The task of a quantum-enhanced metrology (QEM) scheme is to infer an unknown parameter ϕ using entangled states of N particles such that the scaling in uncertainty surpasses $\Delta\phi \propto 1/\sqrt{N}$ obtained using a classical strategy [7]. This scaling is known as the standard quantum limit (SQL). The use of quantum resources enables a QEM scheme to approach the Heisenberg limit (HL) corre-
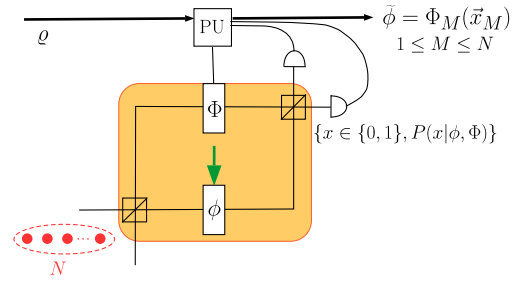

Fig. 1. An adaptive phase estimation scheme. An N-photon entangled state is separated into single-particle bundles. A particle is injected into a Mach–Zehnder interferometer, representing a two-mode interferometer. Contained in this interferometer is an unknown phase shifter ϕ and a controllable phase shifter Φ, whose value after the mth measurement is Φ_m. The path x_m in which the photon exits is detected by single-photon detectors connected to a processing unit (PU). The PU uses the information to adjust $\Phi_{m\geq1} \to \Phi_m$ according to policy ϱ. After all the photons are measured, the estimate $\tilde{\phi}$ is inferred from Φ_M, allowing for loss in photons which leads to $1 \leq M \leq N$. Hence, the estimate is a function of the history of measurement outcomes $\boldsymbol{x}_M = (x_1 x_2 \ldots x_M)$.

sponding to the scaling of $\Delta\phi \propto 1/N$ [61]. This quadratic improvement in precision is valuable for applications where the measurement is operating at the limit of N that can be safely produced or detected.

AQEM is one strategy for performing QEM that involves splitting the input state into a sequence of single-particle bundles [62]. A bundle is injected into the channel and measured at the output. The measurement outcome is used to update the control parameter in preparation for the next bundle. The value of the control parameter after the Nth measurement is then taken to be the estimate.

The evolution of the quantum state over the course of the measurement process is noncontinuous, and so the policy that can achieve quantum-enhanced precision is non-trivial to find. For this reason, machine learning has been introduced [29,36,37]. In our work, we focus on the channel that includes noise and loss, which are imperfections present in every practical measurement setup.

We consider the problem of optical interferometric-phase estimation (Fig. 1), which is well-studied due to its connection to the detection of gravitational wave [63,64] and atomic clocks [65]. The interferometer has two input modes and two output modes. The input state containing N entangled photons is injected into the interferometer one photon at a time.

Neglecting loss, the mth photon comes out from either of the output modes with a probability that depends on $\phi - \Phi_{m-1}$. Our interferometer model includes Gaussian noise on the phase shift with standard deviation σ. We label the outcome by $x_m \in \{0, 1\}$, where 0 refers to the photon exiting the first port and 1 to the photon exiting the second port. The sequence of outcomes from the first to the mth photon is given by $\boldsymbol{x}_m = (x_1 x_2 \ldots x_m)$.

The exit path of the photon is used to determine the value of Φ_m for the next round of measurement. Once all photons are put to use in the Mth measurement, allowing for the loss of photons such that $1 \leq M \leq N$, the estimate $\tilde{\phi}$ of ϕ is inferred from Φ_M to be $\tilde{\phi} \equiv \Phi_M$. As the measurement outcome $\boldsymbol{x}_M$ is a string of discrete random variables, the estimate of phase from this scheme, which is a function of $\boldsymbol{x}_M$, is also discrete.

Because the distribution of the estimate $\tilde{\phi}$ is periodic, the standard deviation is not an appropriate choice to quantify the imprecision $\Delta\tilde{\phi}$. Unless the domain is bounded, the standard deviation is skewed by the existing peaks appearing in the distribution outside of the domain $[0, 2\pi)$ [66]. The imprecision is instead quantified by the Holevo variance [29],
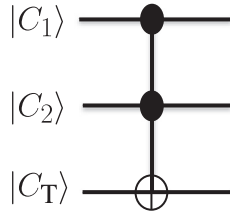
$$V_H = S^{-2} - 1, \tag{2}$$

Fig. 2. The quantum circuit representation of the Toffoli (CCNOT) gate. The horizontal solid black lines are circuit wires with each wire representing a qubit. • shows the control qubits and $\oplus$ denotes the NOT operator acting on the target qubit. $|C_1\rangle$, $|C_1\rangle$, $|C_T\rangle$ refer to the states of the first, second and target qubits. The gate accepts the input from the left side and output a new state on the right side.

Table 1
The truth table representation of Toffoli gate. C_1 and C_2 denote the control qubits, and C_T represents the target qubit. The input and output columns show the states of the three qubits before and after applying Toffoli.

Input			Output		
C_1	C_2	C_T	C_1	C_2	C_T
$|0\rangle$	$|0\rangle$	$|0\rangle$	$|0\rangle$	$|0\rangle$	$|0\rangle$
$|0\rangle$	$|0\rangle$	$|1\rangle$	$|0\rangle$	$|0\rangle$	$|1\rangle$
$|0\rangle$	$|1\rangle$	$|0\rangle$	$|0\rangle$	$|1\rangle$	$|0\rangle$
$|0\rangle$	$|1\rangle$	$|1\rangle$	$|0\rangle$	$|1\rangle$	$|1\rangle$
$|1\rangle$	$|0\rangle$	$|0\rangle$	$|1\rangle$	$|0\rangle$	$|0\rangle$
$|1\rangle$	$|0\rangle$	$|1\rangle$	$|1\rangle$	$|0\rangle$	$|1\rangle$
$|1\rangle$	$|1\rangle$	$|0\rangle$	$|1\rangle$	$|1\rangle$	$|1\rangle$
$|1\rangle$	$|1\rangle$	$|1\rangle$	$|1\rangle$	$|1\rangle$	$|0\rangle$

$$S = \left| \sum_{\widetilde{\phi}} P(\widetilde{\phi}|\phi) e^{i(\phi - \widetilde{\phi})} \right|, \tag{3}$$

which is one possible choice for a periodic distribution [67]. Our goal is to generate a feedback policy such that V_H is minimized and the power-law scaling with N exceeds $1/\sqrt{N}$.

2.3. Quantum-gate design

Quantum computing employs quantum mechanics to perform computation and promises speed-up in computational time for algorithms such as factorization [2] and database search [3]. Implementing quantum computing has been challenging due to the interaction between the quantum system and the environment, which introduces errors and may even nullify the advantage of using quantum resources altogether [68]. If the error rate can be reduced to a value that is less than a specified threshold, error correction can be introduced, and quantum information is thereby protected [69].

Information is encoded on qubits. Each qubit has the state spanned by the basis {$|0\rangle$, $|1\rangle$}. Therefore, unlike a classical bit, a qubit can exist in any superposition $a|0\rangle + b|1\rangle$, where $a, b \in \mathbb{C}$ and $|a|^2 + |b|^2 = 1$. Quantum algorithms use these qubits as resources to perform computations. Just as in classical computation, the quantum algorithm operating on a large number of qubits can be decomposed into gates acting on a few qubits at a time. An ideal quantum gate is reversible and is represented by a unitary transformation U. For each of these gates, error threshold can be assigned such that fault-tolerant quantum computing is attained.

Quantum computing operations can be decomposed efficiently down to a set of one- and two-qubit gates [70,71]. However, this decomposition results in an increased processing time and lower overall fidelity. One way to increase the efficiency is to convert these operations to gates that act on more than two qubits. The Toffoli gate is one such gate. This gate is a controlled-controlled-not gate acting on three qubits (Fig. 2), whose action is summarized in Table 1.

The current experimental schemes to design fast Toffoli gate are limited to decomposition approaches, with the fidelity limited to 68.5% in a three-qubit circuit QED system [72], 71% in an ion-trap system [73], 78% in a four-qubit circuit QED system [74] and 81% in a post-selected photonic circuit [75]. Here, our goal is to devise a machine learning algorithm to design a single-shot threshold-fidelity Toffoli gate without any need to resort to decomposition approach. Our approach to creating a high-fidelity Toffoli gate is not restricted to a specific physical model for quantum computation. We choose to design a Toffoli gate for an architecture of three nearest-neighbour capacitively coupled superconducting transmons [76].

In a transmon system, the computational basis {$|0\rangle$, $|1\rangle$} is assigned to the two lowest energy levels, although the transmon system consists of multiple discrete levels. The evolution of the quantum states for the three transmons are controlled through electrical pulses changing the frequency of each of the transmons. Because the three transmons are coupled, these changes tune their interactions and allow us to perform gate evolution on the computational bases of the three qubits.

The unitary transformation of the Toffoli gate is approximated using a sequence of constant control pulses lasting $\delta t = 1$ ns applied over time τ. At time t, the control pulses $\boldsymbol{\varepsilon}(t) = (\varepsilon_1(t), \varepsilon_2(t), \varepsilon_3(t))$ is applied to the three transmons, effectively subjecting the system to a unitary transformation $U(\boldsymbol{\varepsilon}(t), \delta t)$. The approximate unitary for the gate is, therefore,

$$\tilde{U}(\boldsymbol{\varepsilon}_\tau, \tau) = U(\boldsymbol{\varepsilon}(\tau - \delta t), \delta t)U(\boldsymbol{\varepsilon}(\tau - 2\delta t), \delta t)) \ldots U(\boldsymbol{\varepsilon}(0), \delta t), \tag{4}$$

where $\boldsymbol{\varepsilon}_\tau$ denotes the entire sequence of the three control pulses.

The gate's performance is quantified using the intrinsic fidelity [51],

$$\mathcal{F} = \frac{1}{8} \left| \text{tr}\left(U_T^\dagger \tilde{U}(\varepsilon(\tau), \tau) \right) \right|, \tag{5}$$

where U_T is the unitary transformation of the Toffoli gate. The fidelity $\mathcal{F}$ has a value between 0 and 1, where $\mathcal{F} = 1$ is attained when $\tilde{U} = U_T$. We aim for an intrinsic fidelity beyond 0.9999 for the design of the Toffoli gate as required by fault-tolerant quantum computing.

In the experimental realization, non ideal behavior arises due to electronic imperfections. One is the distortion of the control pulses, which we treat as sequences of piecewise constant functions. A more realistic pulse shape takes into account the response time of the electronics, which acts as a Gaussian filter. Another source of imperfection is the disturbance of the pulses associated with the thermal noise in the electronics. Although this noise is not included in the optimization, we test the robustness of our control pulses by adding random noise $\delta\varepsilon \cdot \text{rand}(-1, 1)$, where $\text{rand}(-1, 1)$ uniformly generates a random number in $(-1, 1)$ to the control parameters at each time bin. The value of $\delta\varepsilon$ is varied from 0 to 300 kHz. We then use the distorted pulse to calculate the intrinsic fidelity for each value of $\delta\varepsilon$.

Note that we have already devised a quantum control scheme to design a high-fidelity quantum gate [50,51]. Our main goal in this article is to present the problem in a different framework, namely machine learning. In particular, we formulate the problem for supervised learning. Expressing the problem in the machine learning context can provide a new perspective on finding control pulses for when the transformations, $U(\varepsilon(t), \delta t)$, are no longer unitary.

3. Machine learning and evolutionary algorithms for control

In this section, we explain how machine learning and reinforcement learning can be used as tools for control. A control problem involves optimization of a control policy such that a target performance is met [77]. Optimizing based on a model is not useful if the model is incorrect, for example by not properly incorporating environmental interaction. As a satisfactory noise model might not exist, the alternative is to implement optimization procedures that are independent of the underlying interaction, with evolutionary algorithms being one such example.

Control theory concerns finding a way to steer a system to achieve a target [77]. The method of control may involve monitoring of a system, in which case the control may involve a feedback loop adjusting the control signal following a policy. Such a control procedure is known as closed-loop control [78]. For a system that cannot be monitored, or for a system whose monitoring is deemed unnecessary, an open-loop strategy is applied, and the control signal is predetermined using a model of the evolution of the system [77,79].

For the cases where the system models are not known, not accurate, or too complicated to be used for generating feasible control policies, machine learning and reinforcement learning can be used to generate the policy through trial and error. Machine learning typically assumes sampling from random variables that are independent and identically distributed (i.i.d.). Concentration theorems based on this assumption give guarantees of estimation accuracy or generalization performance [80,81]. A variant of machine learning takes training instances one by one and updates its model accordingly. This scheme is known as online learning. Some results from statistical learning theory extend to this case if the samples remain i.i.d. [82].

In reinforcement learning, an agent responds to a random outcome from its interaction with an environment [83]. The agent is equipped with a set of possible actions and a scoring function with which it can evaluate its performance. The agent can either greedily optimize the scoring function or aim to maximize long-term performance quantified after the task is completed. In this case, due to the interaction between the learning agent and the environment, the sampling loses the i.i.d. assumption. The most we can assume is that the interaction is modeled by a Markov decision process.

In both machine learning and reinforcement learning, we often boil down the learning problem to a constrained optimization problem with an objective function that can be regularized. The function is seldom convex, making it hard to find the global optimum of the problem. We can use a relaxation and a convex approximation or substitution. For instance, support vector machines replace the optimal 0–1 loss with the hinge-loss to have a convex optimization problem to solve [84]. Another option is to use a non-convex optimization algorithm. This approach works if a convex relaxation is not easy to derive or when we are not satisfied with a local optimum. Sometimes even a greedy algorithm works well for the non-convex case, depending on the topology of the space.

Derivative-free heuristics are common in global optimization. Two widely used algorithms are PSO [85], which is a form of swarm intelligence, and DE [44], which is an evolutionary algorithm. In comparative studies, DE has been shown to be the most powerful in that it converges on a near-optimal solution quickly and can find solutions without stagnation in higher dimensional search spaces than other algorithms [45].

Quantum control tasks also employ closed-loop and open-loop controls to accomplish the desired goals. A key difference between quantum and classical control lies in the quantum system's response to the measurement. Whereas classical systems are un-affected by the measurement procedure, the quantum states response to a measurement with a random jump depending on the measurement outcome. For applications where the quantum states are important to the goal, such as gate design, measurements are to be avoided, and the control procedure for the task is designed based on open-loop control. On the other hand, when a quantum state is used as a resource for accomplishing the desired goal, or if measurements are possible, closed-loop control can be applied.

To implement quantum control in the experimental settings, the control procedures must be designed to be resilient to imperfections, including the nonideal evolution of the quantum state due to its interaction with the environment. Although noise models can be incorporated into computer simulations of the control procedures, the models might not match the noises in the real-world system. One approach is to design a procedure that can learn the control policies directly from experiments, including as few assumptions about the noise as possible.

Reinforcement learning is a valuable tool for applying control as it provides a method to generate a policy that does not rely on the knowledge of the system [18,19,86]. Learning is performed through trial and error, and the system's dynamics is treated as a black box. In fact, this black box treatment of a process is not restricted to a control scenario: if we only want to approximate the output of the process and we have a certain number of disposal of the black box, we can employ a supervised learning algorithm with the assumption of the sampling being i.i.d. For quantum control, this feature is valuable because the knowledge of the dynamics of a real quantum system is never exact. Furthermore, in systems involving measurements, the back-action leads to an exponentially-growing number of state trajectories with the number of possible measurement outcomes $|\{x\}|$, making the problem difficult to solve analytically.

4. Method

We now explain how to construct quantum-control procedures for AQEM and quantum-gate design as learning problems and the challenges they pose to learning algorithms. We then explain how we create noise-resistant differential evolution and methods for generating policies in high-dimensional problems. Here, we assign a fixed mutation rule, although a new DE variant created by random selection of the mutation is possible [87].

4.1. Quantum control procedures as learning problems

In this section, we discuss how the control procedures are performed in an adaptive phase estimation scheme and on the superconducting circuit to create a Toffoli gate. The feedback control applied in the adaptive measurement fits into the framework of reinforcement learning, for which we discuss the policy and the training set. Quantum-gate design, on the other hand, fits with open-loop control as direct monitoring of quantum states interfere with their evolution. To move beyond treating the problem as optimization, in which the noise models are required, we formulate the gate design as a learning problem fit for supervised learning.

4.1.1. Generating feedback policy for AQEM

The goal of adaptive phase estimation is to infer the value of an unknown parameter ϕ following a sequence of measurements performed on a string of N single-photon pulses. In order to deliver quantum-enhanced precision, not only must the particles be entangled, but a policy must be generated that can attain the desired precision. In principle, the policy must be optimized for all sequences $\boldsymbol{x}_M$ and all value of $\phi \in [0, 2\pi)$, which is impossible as ϕ is in a continuous domain. For this reason, we employ the reinforcement learning approach to generate a feasible policy using a training set selected from the domain of ϕ.

Reinforcement learning is the suitable framework because the approach is designed to optimize a decision-making process given random inputs. The adaptive phase measurement involves decision making by the processing unit in response to the random measurement outcomes, which fits in this framework. In the learning algorithm, the adaptive procedure is simulated many times with a fixed input quantum state and treated by the learning algorithm as a black box. The fitness values of the policies are computed over a training set of randomly generated ϕ and are used by the algorithm for the optimization.

In AQEM, the feedback policy is a set of rules that determines how the controllable phase shifter Φ is adjusted. In the mth round of measurement, the policy is a function of the sequence of previous outcomes $\boldsymbol{x}_{m-1} \in \{0, 1\}^{m-1}$. The process is better understood by representing a policy as a binary decision tree, where each branch from the root to the leaf corresponds to a sequence of $\boldsymbol{x}_N$. An advantage of using this representation is that the size of a policy is readily determined from the number of branches, and its size scales as $2^N - 1$.

The exponential scaling of the policy size makes generating the policy increasingly expensive in computational time, which limits the number of particles N in which the application of reinforcement learning is practical. To reduce the policy's size and make searching for the policy tractable, we restrict to Markov feedback [88], in which only the current outcome x_m is used to determined the value of Φ_m. In particular, we impose the update rule

$$\Phi_m = \Phi_{m-1} + (-1)^{x_m}\Delta_m; \tag{6}$$

thus the decision tree is parameterized by a vector $\boldsymbol{\Delta} = (\Delta_1, \ldots, \Delta_N)$. Hence, the size of the policy $\varrho = \boldsymbol{\Delta}$ is reduced to scale linearly with N. The space in which the policy can be searched is therefore restricted to $[0, 2\pi)^N$. Through the implementation of this rule in previous work [36,37,89], we found that the rule leads to feasible policies for adaptive phase estimation.

The update rule (6) restricts the estimate $\widetilde{\phi} \equiv \Phi_M(\boldsymbol{x}_M)$ to a discrete value even though ϕ is continuous in $[0, 2\pi)$. In other words, the AQEM scheme delivers a discrete approximation of ϕ based on the discrete outcomes $\boldsymbol{x}_M$. As N increases, the approximation becomes refined, and an increasingly precise estimate may be obtained.

To determine the performance of a policy, the imprecision of all possible value of $\phi \in [0, 2\pi)$ has to be taken into account. This task poses an additional challenge to calculating Eq. (3), which is computationally expensive. We instead estimate the value of S from

$$S := \left| \sum_{k=1}^{K} \frac{e^{i\theta_k}}{K} \right| \tag{7}$$

where $\theta_k = \phi_k - \widetilde{\phi}_k$ and $K = 10N^2$. This particular choice of K has been shown to deliver the estimate of S that converges in the previous work [36]. The samples are also training data $\{\phi_k\}$ chosen uniformly from $[0, 2\pi)$ to avoid the problem of overfitting.

The adaptive phase estimation scheme works by refining the estimate $\widetilde{\phi}$ through tweaking Φ according to the outcome. In the event of small probability η that the photon is lost, the automated system is instructed to do nothing in the absence of information to update the estimate. We assume a small loss in which case we optimize without loss but test the performance accounting for loss. If the test succeeds, we adopt the policy; if the test fails, we repeat the optimization process.

4.1.2. Shaping the frequencies of transmon system

The goal for quantum-gate design is to generate a set of pulse sequences $\boldsymbol{\varepsilon}_\tau$ such that the transformation of the three transmons over time τ is approximately the Toffoli gate. The pulses are affected by Gaussian distortion, thereby smoothing out the constant piecewise functions. Disturbances are included by adding frequency noises to the pulses after the optimization to test for robustness. For both noises considered, the assumption of unitary transformation holds, and the problem of finding the control pulses can be treated as an optimization problem. In this work, however, we formulate this problem for supervised learning as the learning approach can be implemented regardless of whether the transformation is unitary or not.

When the transformations of the quantum state are unitary, i.e., the quantum system is perfectly isolated from the environment, the control pulses can be optimized over a set of basis states. That is because the quantum state remains a pure state, and due to linearity in quantum mechanics, any transformation on the superposition of the basis states are equal to the superposition of the transformed basis. Hence, optimizing the gate's performance over the basis states is equivalent to the optimization of the problem over the entire state space.

This observation is no longer true when the state becomes mixed, as is the case when the quantum system is coupled to the environment, and the transformations are represented by quantum channels. In this case, optimization over the basis states would not guarantee that the gate's performance is maximized over the entire space of all input states. Hence, the optimization process has to take into account other input states as well. Finding a feasible policy, in this case, is non-trivial as there are infinitely many superposition states. Machine learning becomes a promising approach for generating the control pulses.

The task of supervised learning is to discover a model given a set of input–output data (a training set), which approximates the true function that maps the inputs data to their corresponding outputs. By casting the Toffoli gate as a model for we wish to learn the parameters $\boldsymbol{\varepsilon}_\tau$, we formulate the Toffoli gate design as a regression problem. We use the basis states and their corresponding output states in Table 1 as the training set. The time τ for the gate operation is found through trial and error such that the confidence in the model, quantified by the intrinsic fidelity (5) exceeds 0.9999 as required by fault-tolerant quantum computing.

4.2. Noise-resistant global optimization heuristics

DE is able to find feasible solutions in high-dimensional search space for a set of test problems [45] and for adaptive phase estimation [37]. However, when DE is employed for the problem of noisy phase estimation for N up to 100, we observe that DE does not perform as well as PSO and, in fact, fails to deliver better than SQL scaling. To devise a noise-resistant global optimization algorithm for our scheme, we use the mean value $\bar{S}$ instead of S (7) to determine the performance of a policy. This strategy is one of the many strategies proposed in the literature to create noise-resistant DE [90,91] and is found to work best for our problem.

The principle behind the use of mean objective value is as follows: if noise is added to the fitness function, the process of averaging recovers the true objective value. The optimization using this value is, therefore, a close approximation to the noiseless optimization. The major drawback of this approach is that computing the objective function multiple times makes the procedure computationally expensive. Therefore, determining the smallest sample size of $\{S\}$ necessary to recover S is crucial. To this end, we employ the heuristic applied to PSO in the previous work [89]. The method updates $\bar{S}$ by computing one new sample of S every iteration until a better offspring is generated.

The sample size for computing S is then determined by the probability for DE to generate an offspring that is better than the parent. This probability decreases as the candidates converge on the optimal value. As a result, the sample size grows automati-

cally as the optimization progresses. The computational resources are allocated towards candidates that are close to optimal, which is a favorable strategy as the differences between objective values are dominated by noise in this case. Large sample size enables accurate estimations of the candidates' objective values.

In the particular case of adaptive phase estimation, the phase noise is not additive in S due to the exponential dependence. The mean value computed from J samples, therefore, does not converge on the objective value of the noiseless case but an estimate of S using sample size JK. This method, thereby, provides a better estimate of imprecision than for the sample size of K for adaptive phase estimation including phase noise.

4.3. Improving scalability

In this section, we explain two techniques for achieving scalable learning algorithms. One of the technique is to create an accept–reject criterion, allowing the algorithm to run for as long as it is necessary to generate a feasible policy. This technique is applied to the adaptive phase estimation, where successful policies from small values of N are used to identify a region with a feasible policy for $N + 1$. Another algorithm is devised for quantum-gate design that alternates between optimizing in subspaces and overall space with self-adaptive DE.

4.3.1. Adaptive phase estimation at N > 90

In this section, we discuss accept–reject criteria and how this technique leads to DE delivering successful policies up to $N = 100$. Although DE can generate successful policies for $N > 45$, which is the limitation observed when PSO is used [37], the variances also display stagnation when $N > 90$. To generate policies from a search space that scales up to 100 dimensions, we implement a criterion to the noise-resistant DE to ensure that only successful policies are accepted.

The stagnation is a manifestation of the algorithm not being able to converge to a successful solution in the time limit imposed. Previously the algorithm accepted a policy after a fixed number of iterations regardless of whether the population converges. However, as the dimension of the search space increases, so does the time for the population to converge. Eventually, the algorithm fails to deliver a policy that passes the test. We change the criterion for accepting a policy from a fixed number of iterations to only if V_H is within a distance corresponding to a confidence interval of 0.98 from the inverse power-law line. Thus, we guarantee that the policy from our algorithm always delivers a power-law scaling better than SQL.

The acceptable error δ_y for $N > 93$ is calculated from the statistics of V_H. The Holevo variance V_H are collected from $N = \{4, 5, \ldots, 93\}$, in which we accept the policies after a fixed number of iterations. A linear equation is determined from $\{y_i\} = \{\log V_H(N)\}$ and $\{x_i\} = \{\log N\}$, and is used to predict the next data point y'. The acceptable error from this predicted value is calculated using the previously stored data and the best value of V_H at iteration G from a statistical formula, namely [92,93]

$$\delta_y = t^*_{n'-2} \sqrt{\frac{\sum_{i=1}^{n'}(y'_i - y_i)^2}{n' - 2} \left(\frac{1}{n'} + \frac{(x' - \bar{x})^2}{\sum_{i=1}^{n'}(x_i - \bar{x})^2} \right)}, \tag{8}$$

where n' is the number of data points, $x' = \log N$ for which the error is calculated, and $\bar{x}$ is the average of $\{x_i\}$. The value $t^*_{n'-2}$ is the quantile of the Student's t distribution for $n' - 2$ data points, which we approximate using a normal distribution. The policy is accepted if

$$\left| \log V_H(N) - y' \right| \leq \delta_y, \tag{9}$$

or the optimization continues.

The noise-resistant DE variant, including accept–reject criterion, works as follows[1].

Step 1 Initialize the population of size N_P randomly.
Step 2 Evaluate the objective function for each candidate *twice*, and store the mean objective value and the sample size.
Step 3 Generate a donor $\boldsymbol{D}_i(G)$ for each of candidate $\boldsymbol{V}_i(G)$, where G is the iterative time step, from three other candidates $\{\boldsymbol{V}_{i, 1}(G), \boldsymbol{V}_{i, 2}(G), \boldsymbol{V}_{i, 3}(G)\}$ chosen randomly. For element j of the donor $D_i(G)^{(j)}$,

$$D_i(G)^{(j)} = \begin{cases} V_{i,1}^{(j)}(G) + F \cdot (V_{i,2}^{(j)}(G) - V_{i,3}^{(j)}(G)), & \text{if } r \leq C_r, \\ V_i(G)^{(j)}, & \text{else,} \end{cases} \tag{10}$$

where F is the mutation rate, $C_r \in [0, 1]$ is the crossover rate, and $r \in [0, 1]$ is a random number.
Step 4 Evaluate the mean objective value for each of the new candidates from two samples.
Step 5 Compare and select the candidate for $G + 1$ using the mean objective value,

$$\boldsymbol{V}_i(G + 1) = \begin{cases} \boldsymbol{D}_i(G) & \text{if } \bar{S}(\boldsymbol{D}_i(G)) > \bar{S}(\boldsymbol{V}_i(G)), \\ \boldsymbol{V}_i(G) & \text{else.} \end{cases} \tag{11}$$

Step 6 Evaluate the objective function once, and update the mean value and the sample size.
Step 7 Repeat steps 3–6 until the criterion to terminate the algorithm is met.
Step 8 Compute the objective value of the entire population 10 more times before selecting the candidate with the highest mean objective value as the solution.

The computational complexity of the algorithm is polynomial [37], but it has a high degree, and therefore it is important to identify the performance critical parts of the implementation. We establish that over 90% of the execution time is spent on generating random numbers one by one. The random number generation is primarily used in estimating the Holevo variance as the computation involves simulations of the adaptive measurement procedure. Generating random numbers as they are needed is not efficient on contemporary hardware. The operations can be vectorized to use the single-instruction multiple-data architectures of the central and the graphics processing units (GPUs). Abstracting the random number generation routines and introducing a buffer, we are able to vectorize the respective operations. We study two approaches: one relies on the CPU, using the Intel Vector Statistical Library (VSL), the other on graphics processing units. Eventually, the VSL-based vectorized solution proves to be scalable.

4.3.2. SuSSADE for quantum-gate design

In this section, we turn to another technique to improve DE's scalability and devise the subspace-selective self-adaptive differential evolution (SuSSADE). SuSSADE combines two technique to improve convergence of the algorithm to a successful policy: the self-adaptive heuristic search and the reduction of the search space size. The algorithm alternates between optimizing over the entire search space and one of the subspaces randomly selected for each iteration.

The efficacy of the algorithm for a particular landscape and dimension is determined by the search parameters (F, C_r), set to constant values in traditional DE. Instead of determining optimal values of the parameters through trial and error, which is infeasible when the problem has a large size such as in quantum-gate design, we implemented an algorithm to adapt the value of F and C_r

---

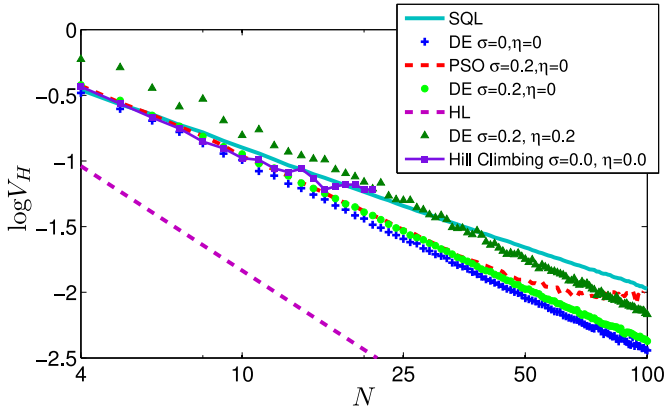[1] Code available at http://panpalitta.github.io/phase_estimation/.

Fig. 3. Logarithm of Holevo variance from adaptive interferometric-phase estimation. The interferometer includes small phase noise of width σ and loss rate η. Three algorithms are used to generate the feedback policies: DE, PSO, and stochastic hill-climbing. This image is a rescaled version of Fig. 1 in Ref. [60].
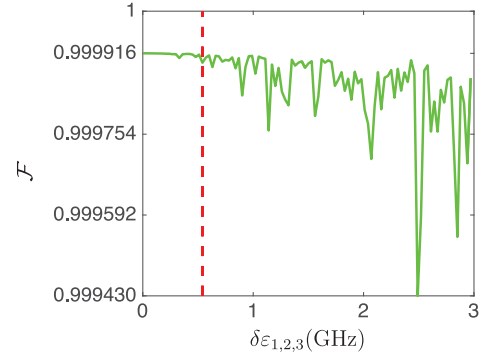

Fig. 4. Intrinsic fidelity $\mathcal{F}$ of the Toffoli gate as a function of $\delta\varepsilon_{1,2,3}$, corresponding to the noise level on each of the three transmons. The vertical red dotted line denotes the threshold, such that $\mathcal{F} > 0.9999$ on the left of the line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

between iterations. This self-adaptive approach [94] has also been used to enhance the performance of DE for the high-dimensional optimization problems and reads as follows.

At iteration G, the algorithm determines the mutation rate F and crossover rate C_r for $G+1$ from

$$F_{G+1} = \begin{cases} F_l + r_1 \cdot F_u & \text{if } r_2 < \kappa_1 \\ F_G & \text{otherwise} \end{cases} \tag{12}$$

and

$$C_{r_{G+1}} = \begin{cases} r_3 & \text{if } r_4 < \kappa_2 \\ C_{r_G} & \text{otherwise,} \end{cases} \tag{13}$$

where r_j, $j \in 1, 2, 3, 4$ are random numbers uniformly sampled from $(0, 1]$. The value of F_l and F_u are predetermined to be 0.1 and 0.9 respectively. The adaptive rate κ_1 and κ_2 are both set to 0.1.

In addition to adapting the search parameters, the algorithm also randomly decides whether the optimization is performed on the entire search space or a subspace. The switching rate $\mathcal{S}$ determines how often this switch occurs and is set by the user. DE is applied on the whole domain if a random number $r < \mathcal{S}$ for $r \in [0, 1]$. Otherwise, the optimization is applied to a smaller domain of the problem. Further details of the SuSSADE optimization algorithm[2] can be found in [50].

5. Results

5.1. Adaptive quantum-enhanced metrology

By applying the method of creating noise-resistant to DE, we are able the obtain a policy that delivers the scaling of $V_H \propto N^{-1.421}$ when the width of the Gaussian distribution $\sigma = 0.2$ rad, and the probability of losing a photon $\eta = 0.2$ are included. This result shows a scaling exceeding N^{-1} expected from SQL, which is given for the ideal interferometer as a benchmark in Fig. 3. The SQL data is generated using a non-entangled N-particle state. The HL shown is an extrapolation using the intercept from the SQL data and is included for the purpose to providing a possible benchmark for the scheme.

Although both the SQL and the HL are reported in the literature for the mean-squared error $\Delta\widetilde{\phi}$ [6], we use the same benchmark for Holevo variance V_H. This follows from the approximation of V_H at low error $\left|\phi_k - \widetilde{\phi}_k\right| \ll 1$. Under this condition, the sharpness in

Eq. (7) is approximated by a series expansion, and through this approximation, V_H is found to be the mean-squared error.

The accept–reject criterion applied to $N > 93$ enables the scheme to show the attain the power-law up to $N = 100$ (Fig. 3). The limitation at 100 photons is due to the computational time and the rounding error in the generation of the large multiparticle entangled state. The time required to find a policy under the accept–reject criterion from 94 to 100 photons is between 1.5 and 3 h per data point.

Policies that are found using stochastic hill climbing break down at 20 photons even for ideal phase estimation. The noise-resistant PSO shows the breakdown at 45 photons, consistent with the previous result [37]. We did not apply the accept–reject criterion to PSO as the computational time would have exceeded the time used by DE at the same number of N and hence not considered worth an investment.

5.2. Quantum gate design

Our machine learning approach to designing a three-qubit gate succeeds in generating policies for the design of a high-fidelity Toffoli gate. The resultant fidelity exceeds 0.9999, which exceeds the threshold fidelity for the fault-tolerant quantum computing. The gate operation time is found to be 26 ns. Therefore, the number of learning parameters add to the total of 81. Although our machine learning technique has optimized the shape of the tranmons' frequencies over a piecewise-error-function, we have shown [50] that the algorithm does not rely on the shape of the pulse but on the number of learning parameters to generate successful policy for the gate design. Our quantum Toffoli gate operates as fast as a two-qubit entangling quantum gate under the same experimental conditions. The policies are also robust against the random uniform noise on the control pulses. The threshold of the frequency $\delta\varepsilon$ of which the intrinsic fidelity remains above 0.9999 is well above the practical noise (up to 100 kHz) of the control devices (Fig. 4).

6. Conclusion

In this work, we report on two examples of applying machine-learning algorithms to quantum control, namely the adaptive phase estimation and quantum-gate design. We employ reinforcement learning to adaptive phase estimation including noise and loss. We are able to attain enhanced precision better than SQL up to 100 photons using a noise-resistant variant of DE and accept–reject criterion. The supervised-learning technique using SuSSADE enables us to perform single-shot high-fidelity three-qubit gates that are as

---

[2] Code available at https://github.com/ezahedin/DE_high-dimensional.

fast as an entangling two-qubit gate under the same experimental constraints.

The methods we employed do not require explicit knowledge of the system's dynamics, although the convexity of the objective functions, the dimension of the problems and the presence of noise have to be taken into account in order to generate a feasible policy. We minimize the runtime of the algorithms by vectorizing the random number generation and employing GPUs and VSL. This technique mostly affects the simulation of the quantum system as the simulations are the most time- and resource-consuming components of the current algorithms.

In principle, the simulation in the learning algorithms can be replaced by signals from experimental setup or simulations of other quantum control schemes. This work can be used as the basis to develop learning algorithms for solving other quantum control problems, such as estimating more than one unknown parameters, which has an application in the characterization of quantum information processing devices, controlled quantum-state transfer in a spin chain [95], and quantum error correction [96].

Acknowledgment

References

[1] C.M. Caves, Quantum information science: emerging no more, in: P. Kelley, G. Agrawal, M. Bass, J. Hecht, C. Stroud (Eds.), OSA Century of Optics, OSA, Washington, D. C., 2015, pp. 320–323.

[2] P.W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, SIAM J. Comput. 26 (5) (1997) 1484–1509, doi:10.1137/S0097539795293172.

[3] L.K. Grover, A fast quantum mechanical algorithm for database search, in: G.L. Miller (Ed.), Proceedings of the Twenty-eighth Annual ACM Symposium on Theory of Computing, STOC '96, ACM, New York, 1996, pp. 212–219, doi:10.1145/237814.237866.

[4] C.H. Bennett, G. Brassard, Quantum cryptography: public key distribution and coin tossing, Proceedings of the 1984 IEEE International Conference on Computers, Systems and Signal Processing, 175, IEEE, Piscataway, New Jersey, 1984, pp. 8–12.

[5] V. Scarani, H. Bechmann-Pasquinucci, N.J. Cerf, M. Dušek, N. Lütkenhaus, M. Peev, The security of practical quantum key distribution, Rev. Mod. Phys. 81 (2009) 1301–1350, doi:10.1103/RevModPhys.81.1301.

[6] V. Giovannetti, S. Lloyd, L. Maccone, Advances in quantum metrology, Nat. Photonics 5 (4) (2011) 222–229, doi:10.1038/nphoton.2011.35.

[7] G. Tóth, I. Apellaniz, Quantum metrology from a quantum information science perspective, J. Phys. A: Math. Theor. 47 (42) (2014) 424006.

[8] C.A. Ryan, M. Laforest, J.C. Boileau, R. Laflamme, Experimental implementation of a discrete-time quantum random walk on an NMR quantum-information processor, Phys. Rev. A 72 (6) (2005) 062317.

[9] 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 (2) (2005) 296–305.

[10] R. Ruskov, K. Schwab, A.N. Korotkov, Squeezing of a nanomechanical resonator by quantum nondemolition measurement and feedback, Phys. Rev. B 71 (23) (2005) 235407.

[11] S. Mancini, D. Vitali, P. Tombesi, Optomechanical cooling of a macroscopic oscillator by homodyne feedback, Phys. Rev. Lett. 80 (1998) 688–691, doi:10.1103/PhysRevLett.80.688.

[12] A. Hopkins, K. Jacobs, S. Habib, K. Schwab, Feedback cooling of a nanomechanical resonator, Phys. Rev. B 68 (23) (2003) 235328.

[13] A. Assion, T. Baumert, M. Bergt, T. Brixner, B. Kiefer, V. Seyfried, M. Strehle, G. Gerber, Control of chemical reactions by feedback-optimized phase-shaped femtosecond laser pulses, Science 282 (1998) 919–922.

[14] D. Meshulach, Y. Silberberg, Coherent quantum control of two-photon transitions by a femtosecond laser pulse, Nature 396 (6708) (1998) 239–242.

[15] P. Brumer, M. Shapiro, Laser control of molecular processes, Ann. Rev. Phys. Chem. 43 (1) (1992) 257–282.

[16] D.J. Tannor, S.A. Rice, Control of selectivity of chemical reaction via control of wave packet evolution, J. Chem. Phys. 83 (1985) 5013.

[17] D. Dong, I.R. Petersen, Quantum control theory and applications: a survey, IET Control Theory Appl. 4 (12) (2010) 2651–2671.

[18] R.S. Sutton, A.G. Barto, R.J. Williams, Reinforcement learning is direct adaptive optimal control, IEEE Control Syst. 12 (2) (1992) 19–22, doi:10.1109/37.126844.

[19] L.P. Kaelbling, M.L. Littman, A.W. Moore, Reinforcement learning: a survey, J. Artif. Intell. Res. 4 (1996) 237–285, doi:10.1613/jair.301.

[20] C.P. Koch, J.P. Palao, R. Kosloff, F. Masnou-Seeuws, Stabilization of ultracold molecules using optimal control theory, Phys. Rev. A 70 (2004) 013402, doi:10.1103/PhysRevA.70.013402.

[21] G. Jäger, D.M. Reich, M.H. Goerz, C.P. Koch, U. Hohenester, Optimal quantum control of Bose–Einstein condensates in magnetic microtraps: comparison of gradient-ascent-pulse-engineering and Krotov optimization schemes, Phys. Rev. A 90 (2014) 033628, doi:10.1103/PhysRevA.90.033628.

[22] B. Rowland, J.A. Jones, Implementing quantum logic gates with gradient ascent pulse engineering: principles and practicalities, Philos. Trans. R. Soc. Lond. A: Math. Phys. Eng. Sci. 370 (1976) (2012) 4636–4650, doi:10.1098/rsta.2011.0361.

[23] M.A. Armen, J.K. Au, J.K. Stockton, A.C. Doherty, H. Mabuchi, Adaptive homodyne measurement of optical phase, Phys. Rev. Lett. 89 (2002) 133602, doi:10.1103/PhysRevLett.89.133602.

[24] H.M. Wiseman, D.W. Berry, S.D. Bartlett, B.L. Higgins, G.J. Pryde, Adaptive measurements in the optical quantum information laboratory, IEEE J. Sel. Top. Quantum Electron. 15 (6) (2009) 1661–1672, doi:10.1109/JSTQE.2009.2020810.

[25] P. Cappellaro, Spin-bath narrowing with adaptive parameter estimation, Phys. Rev. A 85 (3) (2012) 030301, doi:10.1103/PhysRevA.85.030301.

[26] M. Mirrahimi, R. Van Handel, Stabilizing feedback controls for quantum systems, SIAM J. Control Optim. 46 (2) (2007) 445–467, doi:10.1137/050644793.

[27] R. Vijay, C. Macklin, D.H. Slichter, S.J. Weber, K.W. Murch, R. Naik, A.N. Korotkov, I. Siddiqi, Stabilizing Rabi oscillations in a superconducting qubit using quantum feedback, Nature 490 (7418) (2012) 77–80.

[28] H.M. Wiseman, R.B. Killip, Adaptive single-shot phase measurements: the full quantum theory, Phys. Rev. A 57 (3) (1998) 2169–2185, doi:10.1103/PhysRevA.57.2169.

[29] D.W. Berry, H.M. Wiseman, Optimal states and almost optimal adaptive measurements for quantum interferometry, Phys. Rev. Lett. 85 (24) (2000) 5098–5101, doi:10.1103/PhysRevLett.85.5098.

[30] H.M. Wiseman, S. Mancini, J. Wang, Bayesian feedback versus Markovian feedback in a two-level atom, Phys. Rev. A 66 (2002) 013807, doi:10.1103/PhysRevA.66.013807.

[31] S. Roy, I.R. Petersen, E.H. Huntington, Robust adaptive quantum phase estimation, New J. Phys. 17 (6) (2015) 063020.

[32] M. Tiersch, E.J. Ganahl, H.J. Briegel, Adaptive quantum computation in changing environments using projective simulation, Sci. Rep. 5 (2015) 12874, doi:10.1038/srep12874.

[33] L. Banchi, N. Pancotti, S. Bose, Quantum gate learning in engineered qubit networks: Toffoli gate with always-on interactions, NPJ Quantum Inf. 2 (2016) 16019, doi:10.1038/npjqi.2016.19.

[34] M. August, X. Ni, Using recurrent neural networks to optimize dynamical decoupling for quantum memory, Phys. Rev. A 95 (2017) 012335, doi:10.1103/PhysRevA.95.012335.

[35] P.B. Wigley, P.J. Everitt, A.v.d. Hengel, J.W. Bastian, M.A. Sooriyabandara, G.D. McDonald, K.S. Hardman, C.D. Quinlivan, P. Manju, C.C.N. Kuhn, I.R. Petersen, A. Luiten, J.J. Hope, N.P. Robins, M.R. Hush, Fast machine-learning online optimization of ultra-cold-atom experiments, Sci. Rep. 6 (2016) 25890, doi:10.1038/srep25890.

[36] A. Hentschel, B.C. Sanders, Machine learning for precise quantum measurement, Phys. Rev. Lett. 104 (6) (2010) 063603, doi:10.1103/PhysRevLett.104.063603.

[37] N.B. Lovett, C. Crosnier, M. Perarnau-Llobet, B.C. Sanders, Differential evolution for many-particle adaptive quantum metrology, Phys. Rev. Lett. 110 (22) (2013) 220501, doi:10.1103/PhysRevLett.110.220501.

[38] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, Singapore, 2006.

[39] E. Magesan, J.M. Gambetta, A.D. Córcoles, J.M. Chow, Machine learning for discriminating quantum measurement trajectories and improving readout, Phys. Rev. Lett. 114 (2015) 200501, doi:10.1103/PhysRevLett.114.200501.

[40] M. Guţă, W. Kotłowski, Quantum learning: asymptotically optimal classification of qubit states, New J. Phys. 12 (12) (2010) 123032.

[41] E. Zahedinejad, S. Schirmer, B.C. Sanders, Evolutionary algorithms for hard quantum control, Phys. Rev. A 90 (2014) 032310, doi:10.1103/PhysRevA.90.032310.

[42] N. Srinivas, K. Deb, Muiltiobjective optimization using nondominated sorting in genetic algorithms, Evol. Comput. 2 (3) (1994) 221–248.

[43] C.J. Bardeen, V.V. Yakovlev, K.R. Wilson, S.D. Carpenter, P.M. Weber, W.S. Warren, Feedback quantum control of molecular electronic population transfer, Chem. Phys. Lett. 280 (1–2) (1997) 151–158.

[44] R. Storn, K. Price, Differential evolution: a simple and efficient heuristic for global optimization over continuous spaces, J. Global Optim. 11 (4) (1997) 341–359, doi:10.1023/A:1008202821328.

[45] J. Vesterstrom, R. Thomsen, A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems, Proceedings of the 2004 Congress on Evolutionary Computa-

tion (CEC2004), 2, IEEE, Piscataway, New Jersey, 2004, pp. 1980–1987, doi:10.1109/CEC.2004.1331139.

[46] K.V. Price, R.M. Storn, J.A. Lampinen, Benchmarking differential evolution, in: Differential Evolution, in: Natural Computing Series, Springer, Berlin, 2005, pp. 135–187, doi:10.1007/3-540-31306-0_3.

[47] H.M. Wiseman, Adaptive phase measurements of optical modes: going beyond the marginal q distribution, Phys. Rev. Lett. 75 (25) (1995) 4587–4590, doi:10.1103/PhysRevLett.75.4587.

[48] H.M. Wiseman, R. Killip, Adaptive single-shot phase measurements: a semiclassical approach, Phys. Rev. A 56 (1) (1997) 944–957, doi:10.1103/PhysRevA.56.944.

[49] R. Okamoto, M. Iefuji, S. Oyama, K. Yamagata, H. Imai, A. Fujiwara, S. Takeuchi, Experimental demonstration of adaptive quantum state estimation, Phys. Rev. Lett. 109 (2012) 130404, doi:10.1103/PhysRevLett.109.130404.

[50] E. Zahedinejad, J. Ghosh, B.C. Sanders, High-fidelity single-shot Toffoli gate via quantum control, Phys. Rev. Lett. 114 (2015) 200502, doi:10.1103/PhysRevLett.114.200502.

[51] E. Zahedinejad, J. Ghosh, B.C. Sanders, Designing high-fidelity single-shot three-qubit gates: a machine learning approach, Phys. Rev. Applied 6 (2016) 054005, doi:10.1103/PhysRevApplied.6.054005.

[52] A.C. Doherty, S. Habib, K. Jacobs, H. Mabuchi, S.M. Tan, Quantum feedback control and classical control theory, Phys. Rev. A 62 (2000) 012105, doi:10.1103/PhysRevA.62.012105.

[53] C. Altafini, F. Ticozzi, Modeling and control of quantum systems: an introduction, IEEE Trans. Autom. Control 57 (8) (2012) 1898–1917, doi:10.1109/TAC.2012.2195830.

[54] J. Watrous, Lecture Notes on Theory of Quantum Information, University of Waterloo, 2011.

[55] A.S. Holevo, Quantum evolutions and channels, in: Quantum Systems, Channels, Information: A Mathematical Introduction, in: De Gruyter Studies in Mathematical Physics, 16, Walter de Gruyter, Berlin, 2012, pp. 103–131, doi:10.1515/9783110273403.

[56] J. Dolbeault, P. Felmer, J. Mayorga-Zambrano, Compactness properties for trace-class operators and applications to quantum mechanics, Monatshefte für Math. 155 (1) (2008) 43–66, doi:10.1007/s00605-008-0533-5.

[57] H.E. Brandt, Quantum measurement with a positive operator-valued measure, J. Opt. B: Quantum Semiclassical Optics 5 (3) (2003) S266.

[58] W.H. Zurek, Decoherence and the transition from quantum to classical — revisited, vol. 48Progress in Mathematical Physics, Birkhäuser Basel, Basel, pp. 1–31, 10.1007/978-3-7643-7808-0_1.

[59] G.M. D'Ariano, M.G.A. Paris, M.F. Sacchi, Quantum tomography, Adv. Imaging Electron Phys. 128 (2003) 206–309.

[60] P. Palittapongarnpim, P. Wittek, B.C. Sanders, Controlling adaptive quantum-phase estimation with scalable reinforcement learning, in: M. Verleysen (Ed.), Proceedings of the Twenty-fourth European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Ciaco, Belgium, 2016, pp. 327–332.

[61] M. Zwierz, C.A. Pérez-Delgado, P. Kok, Ultimate limits to quantum metrology and the meaning of the Heisenberg limit, Phys. Rev. A 85 (4) (2012) 042112, doi:10.1103/PhysRevA.85.042112.

[62] D.W. Berry, H.M. Wiseman, J.K. Breslin, Optimal input states and feedback for interferometric phase estimation, Phys. Rev. A 63 (5) (2001) 053804, doi:10.1103/PhysRevA.63.053804.

[63] C.M. Caves, Quantum-mechanical noise in an interferometer, Phys. Rev. D 23 (8) (1981) 1693–1708, doi:10.1103/PhysRevD.23.1693.

[64] B.P. Abbott, et al., Observation of gravitational waves from a binary black hole merger, Phys. Rev. Lett. 116 (6) (2016) 061102, doi:10.1103/PhysRevLett.116.061102.

[65] J. Borregaard, A.S. Sørensen, Near-Heisenberg-limited atomic clocks in the presence of decoherence, Phys. Rev. Lett. 111 (9) (2013) 090801, doi:10.1103/PhysRevLett.111.090801.

[66] J. Hilgevoord, The standard deviation is not an adequate measure of quantum uncertainty, Am. J. Phys. 70 (10) (2002) 983. http://dx.doi.org/10.1119/1.1503380.

[67] G.W. Forbes, M.A. Alonso, Measures of spread for periodic distributions and the associated uncertainty relations, Am. J. Phys. 69 (3) (2001) 340–347. http://dx.doi.org/10.1119/1.1317562.

[68] P.W. Shor, Fault-tolerant quantum computation, Proceedings of the Thirty-seventh Annual Symposium on Foundations of Computer Science, IEEE, IEEE, Piscataway, New Jersey, 1996, pp. 56–65.

[69] Y.R. Sanders, J.J. Wallman, B.C. Sanders, Bounding quantum gate error rate based on reported average fidelity, New J. Phys. 18 (1) (2016) 012002.

[70] A. Barenco, C.H. Bennett, R. Cleve, D.P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J.A. Smolin, H. Weinfurter, Elementary gates for quantum computation, Phys. Rev. A 52 (5) (1995) 3457–3467, doi:10.1103/PhysRevA.52.3457.

[71] M. Möttönen, J.J. Vartiainen, V. Bergholm, M.M. Salomaa, Quantum circuits for general multiqubit gates, Phys. Rev. Lett. 93 (13) (2004) 130502, doi:10.1103/PhysRevLett.93.130502.

[72] A. Fedorov, L. Steffen, M. Baur, M.P. da Silva, A. Wallraff, Implementation of a Toffoli gate with superconducting circuits, Nature 481 (7380) (2012) 170–172.

[73] T. Monz, K. Kim, W. Hänsel, M. Riebe, A.S. Villar, P. Schindler, M. Chwalla, M. Hennrich, R. Blatt, Realization of the quantum Toffoli gate with trapped ions, Phys. Rev. Lett. 102 (2009) 040501, doi:10.1103/PhysRevLett.102.040501.

[74] M.D. Reed, L. DiCarlo, S.E. Nigg, L. Sun, L. Frunzio, S.M. Girvin, R.J. Schoelkopf, Realization of three-qubit quantum error correction with superconducting circuits, Nature 482 (7385) (2012) 382–385.

[75] B.P. Lanyon, M. Barbieri, M.P. Almeida, T. Jennewein, T.C. Ralph, K.J. Resch, G.J. Pryde, J.L. O'Brien, A. Gilchrist, A.G. White, Simplifying quantum logic using higher-dimensional Hilbert spaces, Nat. Phys. 5 (2) (2009) 134–140.

[76] J.A. Schreier, A.A. Houck, J. Koch, D.I. Schuster, B.R. Johnson, J.M. Chow, J.M. Gambetta, J. Majer, L. Frunzio, M.H. Devoret, S.M. Girvin, R.J. Schoelkopf, Suppressing charge noise decoherence in superconducting charge qubits, Phys. Rev. B 77 (2008) 180502, doi:10.1103/PhysRevB.77.180502.

[77] J. Leigh, in: Control Concepts: A Non-Mathematical Introduction, second ed, Institution of Engineering and Technology, London, UK, 2004.

[78] M.A. Haidekker, Introduction to linear feedback controls, in: M.A. Haidekker (Ed.), Linear Feedback Controls, Elsevier, Oxford, UK, 2013, pp. 1–13. http://dx.doi.org/10.1016/B978-0-12-405875-0.00001-2.

[79] K.J. Aström, R.M. Murray, Introduction, Princeton University Press, Princeton, NJ, pp. 1–26.

[80] V. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 1995.

[81] L. Devroye, L. Györfi, G. Lugosi, A Probabilistic Theory of Pattern Recognition, Springer, New York, 1996.

[82] P.M. Long, The complexity of learning according to two models of a drifting environment, Mach. Learn. 37 (3) (1999) 337–354, doi:10.1023/A:1007666507971.

[83] R.S. Sutton, A.G. Barto, Problem, in: Reinforcement Learning: An Introduction, in: Adaptive Computation and Machine Learning, MIT, Massachusetts, 1998.

[84] S. Ertekin, L. Bottou, C.L. Giles, Nonconvex online support vector machines, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2) (2011) 368–381, doi:10.1109/TPAMI.2010.109.

[85] J. Kennedy, R. Eberhart, Particle swarm optimization, Proceedings of the 1995 IEEE International Conference on Neural Networks, 4, IEEE, Piscataway, New Jersey, 1995, pp. 1942–1948, doi:10.1109/ICNN.1995.488968.

[86] C. Chen, D. Dong, H.X. Li, J. Chu, T.J. Tarn, Fidelity-based probabilistic q-learning for control of quantum systems, IEEE Trans. Neural Netw. Learn. Syst. 25 (5) (2014) 920–933, doi:10.1109/TNNLS.2013.2283574.

[87] H. Ma, C. Chen, D. Dong, Differential evolution with equally-mixed strategies for robust control of open quantum systems, Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, Piscataway, NJ, 2015, pp. 2055–2060, doi:10.1109/SMC.2015.359.

[88] M. van Otterlo, M. Wiering, Reinforcement learning and Markov decision processes, Adaptation, Learning, and Optimization, vol. 12 Springer, Berlin, pp. 3–42, 10.1007/978-3-642-27645-3_1.

[89] A. Hentschel, B.C. Sanders, Efficient algorithm for optimizing adaptive quantum metrology processes, Phys. Rev. Lett. 107 (23) (2011) 233601, doi:10.1103/PhysRevLett.107.233601.

[90] S. Das, A. Konar, U.K. Chakraborty, Improved differential evolution algorithms for handling noisy optimization problems, Proceedings of the 2005 IEEE Congress on Evolutionary Computation, 2, IEEE, Piscataway, New Jersey, 2005, pp. 1691–1698, doi:10.1109/CEC.2005.1554892.

[91] L. Barone, A. Di Pietro, L. While, Applying evolutionary algorithms to problems with noisy, time-consuming fitness functions, in: G. Greenwood (Ed.), Proceedings of the 2004 Congress on Evolutionary Computation (CEC2004), 2, IEEE, Oregon, 2004, pp. 1254–1267, doi:10.1109/CEC.2004.1331041.

[92] X. Yan, X.G. Su, Simple Linear Regression, World Scientific, Singapore, pp. 9–40.

[93] N.H. Bingham, J.M. Fry, Linear Regression, Springer Undergraduate Mathematics, Springer, London, pp. 1–32, 10.1007/978-1-84882-969-5_3.

[94] J. Brest, V. Zumer, M.S. Maucec, Self-adaptive differential evolution algorithm in constrained real-parameter optimization, Proceedings of the 2006 IEEE International Conference on Evolutionary Computation, IEEE, Piscataway, New Jersey, 2006, pp. 215–222, doi:10.1109/CEC.2006.1688311.

[95] J. Gong, P. Brumer, Controlled quantum-state transfer in a spin chain, Phys. Rev. A 75 (2007) 032331, doi:10.1103/PhysRevA.75.032331.

[96] C. Ahn, A.C. Doherty, A.J. Landahl, Continuous quantum error correction via quantum feedback control, Phys. Rev. A 65 (2002) 042301, doi:10.1103/PhysRevA.65.042301.
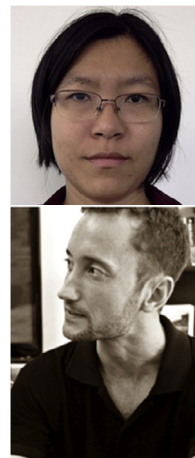
Pantita Palittapongarnpim is a doctoral student in physics at the University of Calgary working on quantum-enhanced metrology. She received her Bachelor of Science degree from Mahidol University in Bangkok, Thailand and a Master of Science degree in experimental quantum optics from the University of Calgary.

Peter Wittek is a research fellow working on quantum machine learning, computational methods in quantum correlations and quantum simulations, and scalable learning algorithms on supercomputers. Trained as a mathematician and computer scientist, he received his Ph.D. from the National University of Singapore. He has been involved in major EU research projects and obtained several national and industry grants. Currently, he works in the Quantum Information Theory group in ICFO – The Institute of Photonic Sciences. He has been affiliated with the University of Borås since 2010. He did longer research terms at several institutions, including the Indian Institute of Science, Barcelona Supercomputing Center, Tsinghua University, the Centre for Quantum Technologies at the National University of Singapore, and the Quantum Information Group at the University of Tokyo.

Ehsan Zahedinejad is a researcher at 1QBit following his doctoral and post-doctoral work at the University of Calgary. His doctoral research was on supervised learning for quantum-gate design.

Shakib Vadaie is a doctoral student in physics at the University of Calgary. His research applies reinforcement learning to quantum control, specifically in the task of designing quantum gates. He received a Master of Science degree in physics from Sharif University of Technology, Tehran, Iran.

Barry C. Sanders is Director of the University of Calgary Institute for Quantum Science and Technology, a 1000-Talent Chair at the University of Science and Technology of China and a Senior Fellow of the Canadian Institute for Advanced Research. He is a Fellow of the American Physical Society, the Optical Society of America and the Institute of Physics (U.K.) and is Editor-in-Chief of the New Journal of Physics.