Feedback-based Steering for Quantum State Preparation

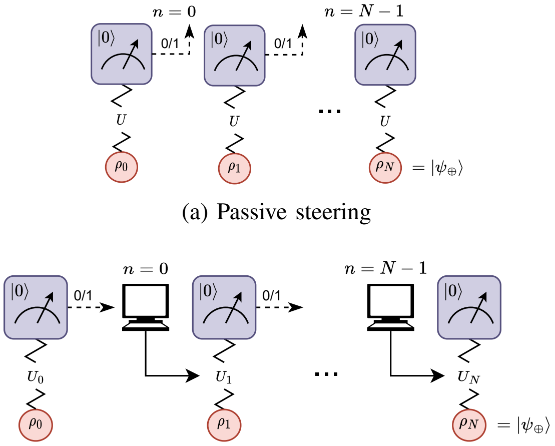
Daniel Volya, Zhixin Pan and Prabhat Mishra
University of Florida, Gainesville, Florida, USA

***Abstract*—State preparation is an essential component in quantum information science. A recently developed steering protocol utilizes a sequence of generalized measurements on a detector to steer a quantum system towards a desired state. However, it is designed as an open-loop technique that requires accurate modeling of the overall quantum system and can be prone to errors. To address this challenge, we propose a closed-loop control technique that introduces feedback to the steering protocol, providing robustness to noise and faster state convergence. We introduce two strategies for feedback: (1) a gradient-based active steering protocol that changes the detector-system coupling conditioned on the detector's readout and (2) tuning the fixed detector-system coupling via model-free reinforcement learning. We study the effectiveness of these strategies under various noise models, including both incoherent and decoherent noise, and discuss potential applications in quantum technologies.**

***Index Terms*—Quantum computing, quantum measurement, quantum steering, state preparation, quantum control.**

I. Introduction

The ability to initialize quantum systems to a desired state, commonly referred to as state preparation, is a fundamental requirement for realizing quantum technologies. This process is vital for a range of applications, such as quantum computing where the correct application of gates to implement specific functionalities relies on the initial known state. Similarly, quantum communication protocols require the preparation of an initial entangled state, quantum memory relies on state preparation to read and write quantum states, and quantum sensors must be in a known state in order to accurately measure external stimuli. A well-established and successful approach to solve the task of state preparation is via quantum optimal control. These methods rely on the Schrödinger equation, its differentiability, and gradient-based optimization methods, to improve the control fields that manipulate the quantum system. One such technique is Gradient Ascent Pulse Engineering (GRAPE) [1]–[3], which evaluates and ascends along the gradients to reach a desired state (or a quantum gate). However, many techniques such as GRAPE are open-loop, also referred to as non-feedback or passive, where the actions are independent of the process output [4]. This makes them susceptible to parameter uncertainties and noise processes that can arise in devices, leading to inaccurate predictions of the underlying quantum dynamics [5]–[7].



(a) Passive steering

(b) Active steering

Fig. 1: Overview of the steering protocol where a system couples with a detector. Measurements of the detector induce a backaction on the system. Readouts of the detector may be simply ignored (passive steering) or processed by a classical computer to make decisions about the coupling U (active steering). The goal is to reach a desired target state $|\psi_\oplus\rangle$, with an arbitrary initial state ρ_0. (a) Passive steering is an open-loop protocol and utilizes a sequence of generalized measurements on a detector to steer a quantum system towards a desired state. (b) Active steering is a closed-loop protocol and involves a detector coupled with our system via a tunable unitary U_i $(0 \leq i < N)$ conditioned on the readouts of the detector. As the detector is measured, it induces a backaction on the system taking the state from ρ_i to ρ_{i+1}. After measurement, the detector is reset back to a known state (i.e. $|0\rangle$). The readout results are processed by a classical computer, which selects the next unitary coupling parameter U_{i+1}.

To address these challenges closed-loop control strategies, also known as active, are employed by performing measurements on the quantum system to incorporate feedback. These strategies have led to important developments in a variety of challenging tasks, including quantum error correction [8], [9], state preparation [10]–[12], stabilization [13], the Zeno effect [14], and removing entropy from a system [15]. However, a downside of feedback-based active strategies is that they are typically slower as each sequence of measurements must be conducted in the weak regime, meaning that a weak coupling of a detector with a system is necessary. While a weak coupling between the system and detector results in slower evolution, it is necessary to prevent the system from collapsing, enabling repeated measurements for obtaining feedback. The weak coupling is associated with slow state convergence which can leave the system exposed to its environment for longer periods, and increases the likelihood of cascading errors. Furthermore, in

This work was partially supported by the NSF grant CCF-1908131.

contemporary quantum devices [16] it is crucial to optimize the utilization of quantum resources which makes slow evolution resulting from weak coupling undesirable.

A novel approach for state preparation based on quantum *steering protocol* takes advantage of the backaction induced by measuring an entangled bipartite state [17]. In the case of *passive steering*, this protocol repeatedly applies a fixed operation on a coupled detector-system while performing measurements on the detector to steer the system towards a desired target state as illustrated in Figure 1(a). The system state is prepared through a series of measurements, where the readouts of the measurements' are ignored. The measurement readouts can be considered for halting the repetition when the desired outcome is achieved [18]. This steering protocol offers a significant advantage: with a particular form of strong coupling between the detector-system and a single measurement, a state may be prepared instantly. If a strong coupling is unattainable, repeated coupling and measurements still result in convergence to the desired state [19], [20]. Recent experimental results on cloud-accessible quantum computers demonstrate the effectiveness of this steering protocol [21]. While promising, the outcomes also highlight the limitations of an open-loop control steering protocol, where noise can lead to uncertainties in state preparation.

To overcome such limitations, closed-loop control can be employed utilizing the feedback. Recent theoretical work [18] introduces feedback control to allow for real-time adjustments in the steering protocol – *active steering*. The goal of work [18] was to accelerate the rate of convergence to a desired state. While the proposed faster protocol optimizes the usage of quantum resources, it does not explicitly account for system uncertainties or noise processes. In this paper, we seek to address this concern, and introduce feedback strategies that yield stability even in the presence of noise. Specifically, we propose the following two strategies:

1) A gradient-based active steering approach where the final state fidelity is optimized for all quantum trajectories, as shown in Figure 1(b).
2) A model-free reinforcement learning (RL) approach that aims to improve passive steering (Figure 1(a)). This is useful in the case when labeled data is scarce or expensive to obtain.

We demonstrate the effectiveness of each method on simple, yet realistic, noise models. This paper is structured as follows. First, we review the background and related works on quantum feedback strategies, as well as recent developments in measurement-induced steering of quantum systems. Next, we briefly outline the relevant background information and terminology used throughout the paper. We then present our methodology for gradient-based optimization of the active steering protocol followed by RL-based quantum steering. Finally, we present experimental results of our approach and summarizes our findings.

II. Related Work

In a closed system without external disturbances or noises, one can always prepare a target state by applying appropriate unitary operations to the system. However, in open quantum systems, it is difficult to maintain coherence because of unavoidable coupling with the environment. Therefore, the main challenge in quantum state preparation is obtaining and protecting a target state in the presence of decoherence caused by environmental noise. In this section, we review some recent works on quantum control and outline the general measurement-induced steering protocol.

A. Quantum State Engineering

Quantum feedback can be broadly categorized into two main types: measurement-based feedback and coherent feedback, each presenting distinct advantages and challenges depending on the specific application, as well as the desired level of control and error tolerance. Measurement-based feedback requires collapsing the quantum state through measurement, followed by classical information processing [22]. However, this method can introduce noise and errors that negatively impact feedback control and the overall performance of the quantum technology. Furthermore, the loss of quantum coherence due to state collapse may constrain the precision and control that can be achieved. In contrast, coherent feedback [23] preserves the quantum nature of the system and its feedback loop through continuous interactions between the system and a quantum controller, such as a detector. While this approach can provide enhanced control and reduced noise, it also poses challenges in maintaining coherence and preventing decoherence in the involved quantum systems, making implementation more complex. In this paper, we focus on steering, which, in essence, combines elements of both methods, wherein measurements on a detector supply feedback while simultaneous interaction with the system and detector drives its evolution. The remainder of this section examines related studies, and comments on the differences and similarities with steering.

1) Gradient Ascent Pulse Engineering with Feedback: Drawing inspiration from model-free reinforcement learning, feedback-GRAPE was developed to incorporate the response to strong stochastic measurements while performing direct gradient ascent optimization of quantum dynamics [24]. In addition to the conventional optimization of control parameters for the dynamics, feedback-GRAPE accounts for the probabilistic state collapse resulting from measurements, which in turn provide feedback to the system. This innovative approach combines elements of both optimization and feedback control, enabling more robust and efficient control of quantum systems. Notably, our work also optimizes control parameters for quantum dynamics, akin to feedback-GRAPE. However, we consider strong measurements on a coupled detector rather than direct measurements on the system, and instead exploit the measurement backaction to evolve the system state. Furthermore, unlike traditional GRAPE, which optimizes pulses for continuous dynamics, we focus on optimizing specific discrete gates (or quantum circuits) (see Section III-A), providing a general perspective that does not require detailed knowledge of the quantum device. Nonetheless, when the control Hamiltonians for the system of interest are universal, we may take pulses into account by using the chain rule.

2) State Engineering via Dissipation: Quantum-state engineering driven by dissipation has emerged as a promising paradigm for manipulating and controlling quantum systems by harnessing the dissipative dynamics that typically arise from interactions between the system and its environment [25]–[27]. While dissipation has traditionally been considered detrimental due to its potential to cause decoherence and loss of information, researchers have started recognizing its potential as a valuable resource for generating specific quantum states or performing quantum operations. Seminal works in this field have led to the development of various techniques, such as engineered reservoirs and tailored control sequences, to guide quantum systems towards desired target states or operations. In these scenarios, the dissipative environment is assumed to be Markovian and is modeled using Lindbladian dynamics. Similarly, steering incorporates a detector that serves as an environment, and as the detector state is freshly prepared, it also displays Markovian behavior. A notable benefit of steering is that, in contrast to the dissipation case where the environment simply becomes entangled with the system, the measurement of a detector enables feedback control possibilities, presenting new avenues for managing quantum systems. Moreover, the jump operators in the Lindblad equation for steering are designed based on the selected detector-system coupling, unlike in the dissipative case where the environment itself provides its own jump operators.

3) Quantum Error Correction: Other state engineering and feedback strategies have also been proposed, including feedback based on quantum error correction codes [28]. These strategies have shown promise for achieving fault-tolerant quantum computing and stabilizing quantum states in the presence of noise. However, the implementation of quantum error correction codes is typically resource-intensive, requiring a large number of physical qubits and significant overhead for error detection and correction. Furthermore, feedback based on quantum error correction codes can be sensitive to the choice of code and the specific error model, and the performance of the code may degrade as the error rate increases.

B. Measurement-induced Steering

Quantum steering, as first coined by Schrödinger, refers to the peculiar property of quantum mechanics whereby an entangled quantum state may be steered from one state to another due to an experimenter's act of measurement [29]. While the phenomena of quantum entanglement has appeared in a number of application in the field of quantum information, it has recently been utilized in developing a novel protocol for state preparation [17]. The protocol assumes that our system of interest, described by the density matrix ρ_s, is allowed to couple with a detector with the density matrix ρ_d. Secondly, the protocol assumes that the detector can be quickly reset to a predefined pure state, and that it is first initialized to this state. We label this state to be $|0\rangle$. With these assumptions, the protocol prepares an arbitrary target state $|\psi_\oplus\rangle$ via a repetition of the following steps:

1) At the n-th step, couple the detector and system yielding a composite state described by the density matrix $\rho_{d-s}^{n+1} = U(\rho_d \otimes \rho_s^n)U^\dagger$.
2) a) *blind* measurement of the detector resulting in a system state
$$\rho_s^{n+1} = \mathrm{Tr}_d \left[\rho_{d-s}^{n+1}\right] = \mathrm{Tr}_\mathrm{d} \left[U \rho_d \otimes \rho_s^n U^\dagger\right] . \tag{1}$$
   b) *non-blind:* projective measurement (Π_r) of the detector qubit resulting in a readout r, where the system state now depends on r
$$\rho_s^{n+1} = \mathrm{Tr}_\mathrm{d} \left[\frac{\Pi_r \rho_{d-s}^{n+1} \Pi_r^\dagger}{\mathrm{tr}(\rho_\mathrm{d-s}^{n+1} \Pi_\mathrm{r})}\right] . \tag{2}$$
3) a) *passive protocol:* continue to next step.
   b) *active protocol:* make a decision for the choice of coupling U, dependent on the measurement outcome, that appears in next iteration (or decide to terminate).
4) Reinitialize the detector to the simple pure state $\rho_d = |0\rangle$, and return to step 1.

In the paper, we discuss strategies $1 \rightarrow 2(a) \rightarrow 3(a)$ and $1 \rightarrow 2(b) \rightarrow 3(b)$ which we simply refer to as passive steering and active steering respectively. With a sufficient number of iterations, the backaction induced by our detector steers our system state to the desired target state $|\psi_\oplus\rangle$. The initial state may be arbitrary pure or mixed. However, the success of the protocol crucially depends on the coupling operators U, which controls the kind of backaction experienced by the system. In the passive case, the operator U is chosen [17] such that the fidelity of the system improves with each iteration:
$$\langle\psi_\oplus| \rho_s^n |\psi_\oplus\rangle < \langle\psi_\oplus| \rho_s^{n+1} |\psi_\oplus\rangle . \tag{3}$$
Because the operator U remains fixed, this is the simplest case of the protocol. Furthermore, in the blind variation, readouts of measurement are ignored and hence does not require any additional processing. However, a fixed predetermined operator U may increase problems with noise. In Section IV-B, we present a model-free reinforcement framework that learns an U which will satisfy Equation 3 in the presence of noise.

However, in the active scenario, the operators U_i are determined dynamically and do not need to satisfy Equation 3. The only requirement is that at the end of protocol we arrive at our target state:
$$\langle\psi_\oplus| \rho_s^n |\psi_\oplus\rangle = 1. \tag{4}$$

The active protocol allows us to gain feedback and tune U_i for the next iteration. This provides an interesting opportunity to modify the behavior of the system without directly measuring it. However, a strategy to choosing U_i is a difficult [18] and open problem. In Section IV, we outline our gradient-based method that optimizes the choice of U_i. For further details on the variations of the protocol, we refer to [18], [21].

III. Background

In this section, we briefly introduce the terminology and concepts used throughout the paper. We consider unitary parametrization, noise models of quantum devices, and ingredients for reinforcement learning.

A. Unitary Parametrization

As briefly mentioned in Section II-A1, we consider a general quantum system, and therefore assume the system is capable of preparing an arbitrary quantum evolution (e.g. an arbitrary quantum gate). Any valid quantum gate on an N dimensional system, excluding an irrelevant phase, lives in the Lie group $SU(N)$. Matrices in $SU(N)$ have size $N \times N$, are unitary, and have a determinant equal to one. From a group theoretical perspective, actions within the $SU(N)$ group can be represented by $N^2 - 1$ generators that are represented as $N \times N$ Hermitian and traceless matrices, which generate infinitesimal rotations. The finite rotations are generated with generalized Euler angles [30]. A convenient choice for these matrices are the generalized Gell-mann matrices λ_i. Hence, a special unitary matrix U can be specified by Euler angle coordinates $\vec{\theta}$ in the Gell-mann basis $\vec{\lambda}$

$$U = e^{i\vec{\theta}\cdot\vec{\lambda}}. \tag{5}$$

The generators generally do not commute, so ordered summation is assumed in Equation 5. An important property is that these Euler angles are strictly real, which simplifies differentiation with respect to the angles and for presenting a reinforcement learning strategy.

B. Weyl Chamber

Any two qubit gate $U \in SU(4)$ can be expressed according to the Cartan decomposition [31], [32]

$$U = k_1 \exp\left[\frac{i}{2}\left(c_1\sigma_x^d\sigma_x^s + c_2\sigma_y^d\sigma_y^s + c_3\sigma_z^d\sigma_z^s\right)\right]k_2 \tag{6}$$

where $\sigma_x, \sigma_y, \sigma_z$ are Pauli matrices for detector and system, and $k_{1,2} \in SU(2)\otimes SU(2)$ are single-qubit local operations. By taking into account symmetries, the coefficients can be limited to $c_1 \in [0, \pi]$ and $c_2, c_3 \in [0, \pi/2]$. They may be interpreted as coordinates in a three-dimensional space, where all possible two-qubit gates are points in a quarter pyramid known as the Weyl-chamber. These coefficients express the non-local "entangling" part of the gate U. The chamber is depicted in Figure 8.

An additional characterization of two-qubit gates (or any bipartite system) is via entanglement power [33]. Entanglement power considers how much entanglement is produced by a gate U on average, acting on a set of unentangled states. In the two-qubit case, the entanglement power can be expressed in terms of the coordinates c_1, c_2 and c_3 [34],

$$\begin{aligned} e_p(U) = &\frac{1}{18}[3 - \cos 2c_1 \cos 2c_2 + \cos 2c_2 \cos 2c_3 \\ &+ \cos 2c_3 \cos 2c_1]. \end{aligned} \tag{7}$$

The values are bound to $0 \leq e_p \leq 2/9$. Perfect entanglers, ones that produce a maximally entangled state from some product state, are in $1/6 \leq e_p \leq 2/9$.

C. Noise Models

In this paper, we consider two general noise models [7].

1) Decoherent noise: Decoherent noise is a type of noise that arises in quantum systems due to interactions with their environment, causing the system to lose its coherence and become entangled with the environment. This leads to errors in quantum operations and measurements, which can significantly affect the accuracy and reliability of quantum computations. We will consider depolarizing noise, which is a general type decoherent noise.

Depolarizing noise is a process that randomly changes the state of a qubit with a certain probability, causing it to lose its coherence over time. This type of noise is characterized by the depolarizing parameter p, which represents the probability that a qubit will experience a random Pauli rotation around one of its axes. In other words, depolarizing noise can be expressed as either maintaining its current state or becoming mixed. If the state of a qubit is given by ρ then depolarizing noise will map the state to

$$\mathcal{E}(\rho) = (1-p)\rho + \frac{p}{N}\mathbb{I} \tag{8}$$

where $\mathbb{I}/N$ is mixed state of the detector. In this paper, we use depolarizing noise to model faulty detector initialization.

2) Incoherent errors: Incoherent noise is another type of noise that can affect quantum systems. Unlike decoherent noise, which is due to the coupling of the system with its environment, incoherent noise arises from fluctuations within the quantum system itself. Examples of incoherent noise include random variations in the amplitude or phase of the qubits or gates, as well as errors in the initialization or readout of qubits. We use incoherent noise to model errors in the steering operator U. Specifically, to model incoherent noise, we pick a set of R random unitary matrices using the Gaussian Unitary Ensemble (GUE) [35], [36]. This is equivalent to generating through randomly chosen generalized Euler angles from the Haar measure [30]

$$\{U_j\}_{j=1}^R = \{e^{i\vec{\theta}_j\cdot\vec{\lambda}}\}_{j=1}^R.$$

We associate a probability to each unitary matrix so that

$$\{p_j\}_{j=1}^R \quad \text{s.t.} \quad \sum_{j=1}^R p_j = 1.$$

At each instance, a noisy unitary matrix U_{noisy} is generated

$$U_{\text{noisy}} = UU_{\text{rand}}, \quad U_{\text{rand}} \in \{U_j\}_{j=1}^R \tag{9}$$

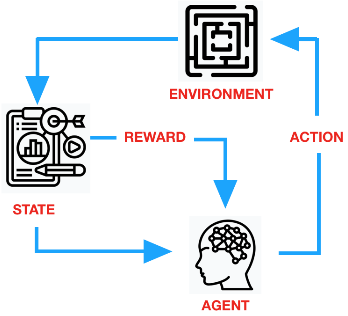


Fig. 2: An overview of reinforcement learning.

with probability p_r. Additionally, we add a parameter ϵ that tunes the strength of U_{rand}. In this paper, we use incoherent errors to model faulty steering operators U_i.

D. Machine Learning Algorithms

Machine learning (ML) algorithms have received considerable attention for various domains in recent years due to their scalability in handling tasks [37]. Broadly speaking, ML algorithms can be categorized into two major types, supervised and unsupervised learning. Supervised learning is a type of ML where the algorithm extract features from ground-truth labeled data, and is often preferred when the goal is to learn a mapping function from input to output data. In cases where we have sufficient amount of data available, supervised learning algorithms can often achieve high levels of accuracy and performance. However, supervised learning is not applicable in situations where labeled data is scarce or expensive to obtain. In these cases, unsupervised learning is more suitable.

Unsupervised learning is a type of machine learning where the algorithm learns patterns and relationships from unlabeled data, without the need for explicit supervision or guidance from a human. One of the typical unsupervised learning algorithm is reinforcement learning (RL), which trains an agent to continuously learn decision-making behaviors by interacting with an environment and receiving feedback rewards, as outlined in Section III-E. This self-learning nature enables RL to perform well in scenarios where it is difficult or impractical to provide sufficient amount of labeled training data. Another advantage of RL is interpretability and explainability. Because the algorithm is trained in an iterative “feedback and update” manner, it is often easier to understand why the RL model is making its predictions and how the model gradually keeps improving itself. However, RL requires continuously adapting and improving itself over time. In this way, RL is relatively less stable compared to supervised learning algorithms, and is often hard to train for applications where the environment is noisy or highly unpredictable.

E. Unsupervised Reinforcement Learning

A key challenge in feedback-based passive steering boils down to finding a suitable unitary matrix U as the operator. To address this challenge, Reinforcement Learning (RL) is applied to our framework. RL has emerged as a promising approach to excavate optimal solution in a large problem space, as demonstrated by its successful application in various domains [38]–[40]. RL is more similar to human learning, where acquisition process involves exploration, trial and error, and feedback from the environment, which gradually teaches human the policies of interacting with the world. Similarly, RL algorithms learn to discover optimal strategies by constantly adjusting the ML model’s behavior based on feedback from the environment, through a series of attempts and iterations. An overview of RL framework is shown in Figure 2. It consists of five core components: *Agent*, *Environment*, *Action*, *State* and *Reward*.

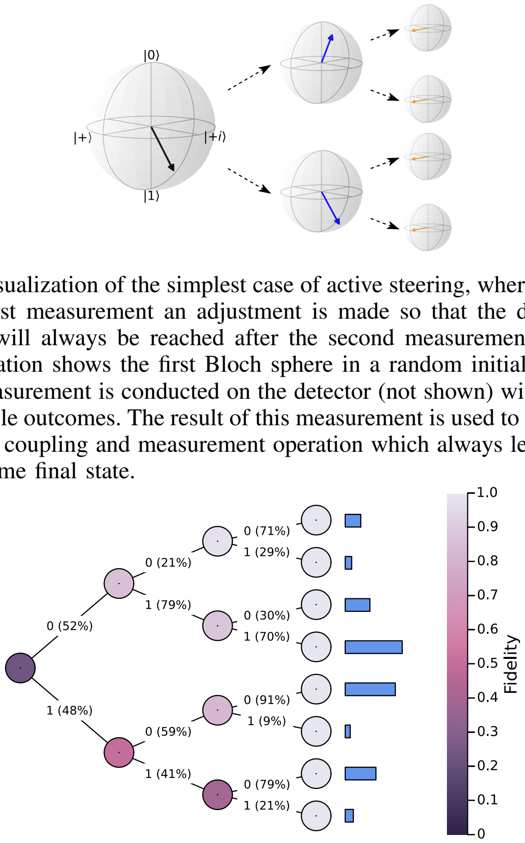
The *agent* is the entity responsible for making decisions and taking actions in the environment. In the context of optimization problems, an agent can be viewed as a set of test initials to be optimized. The *environment* is the system that the agent interacts with. For example, the environment typically refers to the objective function that needs to be optimized. An *action* is a particular decision or choice made by the agent that affects the environment. For example, an action could correspond to mutating a particular set of input parameters to evaluate the objective function. A *state* is a description of the environment that is perceived by the agent. For example, a state could include information about the current status of the entire system, as well as any other relevant variables or parameters. The *reward* is a feedback signal from the environment that reflects the effect of the agent’s latest action. For example, the reward is typically defined as the improvement in the objective function after applying the current set of input parameters. The goal of the agent is to maximize the expected reward over time, by learning to select actions that lead to better performance.

IV. Feedback-based Steering Protocol

In this section we outline the two strategies for introducing feedback into the steering protocol, with the goal of noise resilience. The first strategy is to consider active steering, and allow changes in the steering operator based on the previous history of readouts from the detector. We formulate this as an optimization problem, where a classical computer must derive the optimal steering operators at each step: *gradient-based active steering*. The second strategy seeks to improve the previously studied passive steering [17], [21] where the detector readouts are ignored and the steering operators remain fixed. We add feedback by considering the final fidelity of the system state, and modify the steering operator via a reinforcement learning: *reinforcement learning passive steering*.

A. Gradient-based Active Steering

As introduced in Section II-B, the active steering protocol consists of a detector-system that is coupled via a parameterized unitary evolution U_θ^i with parameters θ (Equation 5) and at a step i. Measurements are conducted on the detector, which produces readout outcomes r_i and resulting in a backaction on the system. The act of measurement may be



(a) Visualization of the simplest case of active steering, where after the first measurement an adjustment is made so that the desired state will always be reached after the second measurement. The illustration shows the first Bloch sphere in a random initial state. A measurement is conducted on the detector (not shown) with two possible outcomes. The result of this measurement is used to devise a new coupling and measurement operation which always leads to the same final state.

(b) Each node represents an application of a coupling unitary followed by a measurement on the detector (not shown). The edges represent the readout result of measurement (0/1) along with the associated probability. The color of an edge represents the fidelity of the system with respect to a desired target state. A histogram is shown with a cumulative probability distribution of all trajectories. In this example, with 3 measurements the system goes from an arbitrary initial state to our desired target state.

Fig. 3: The evolution of a quantum system subject to the gradient-based active feedback strategy.

defined in terms of a projection operator Π_r, and will result in a conditional system state

$$\rho_s^{i+1} = \begin{cases} \mathrm{Tr}_d[\Pi_0 \rho_{d-s} \Pi_0 / p_0], & p_0 = \mathrm{Tr}[\rho_{\mathrm{d-s}} \Pi_0] \\ \mathrm{Tr}_d[\Pi_1 \rho_{d-s} \Pi_1 / p_1], & p_1 = \mathrm{Tr}[\rho_{\mathrm{d-s}} \Pi_1] \end{cases} \tag{10}$$

where detector readouts are a 0 or a 1 with probability p_0 and p_1, respectively.

Our goal is to select U_i, and utilize them as a feedback mechanism such that given path will steer to a desired state. Specifically, in the context of quantum control, the control parameters are the unitary coupling operators that are applied at different steps, and depend on all previous readout outcomes which we utilize to provide feedback,

$$U_\theta^i(r_n, r_{n-1}, \ldots, r_1). \tag{11}$$

For the ease of illustration, we simplify the notation to denote $U_\theta^j(r)$ to refer to the operator being dependent on all readout outcomes up to r_j. These feedback control are differentiable, depending on parameters θ, which may be optimized via gradient-based optimization techniques. Although the first unitary U^1 does not depend on any previous readouts, it can still be optimized. Note that although the unitary operators are differentiable with respect to the parameters θ, their dependence on measurement readouts requires extra care. The key insight is to note that the measurements are not conducted on the system, but rather on the detector. Furthermore, the probabilities of different readout outcomes depend on all the previously applied unitary operators which is carried the evolution of detector state. To account for this dependence, the evaluations of gradients with respect to θ is done by summing all readout paths. Therefore, when a gradient is computed with respect to θ, it also accounts for the dependency on the probability that arises from the application of the first unitary and measurement of the detector, $U_\theta^0(\rho_S^0 \otimes |0\rangle\langle 0|) U_\theta^{0^\dagger}$.

Our goal is to minimize the overall cumulative error $\mathcal{J}$, which is referred to as the cost. In our case, the cost is defined in terms of the final fidelity of our system with respect target state. Hence, for a given sequence of measurement readouts, we define

$$\mathcal{J}(r) = 1 - \langle \psi_\oplus | \, \rho_s^n(r) \, | \psi_\oplus \rangle \tag{12}$$

where ρ_S^n is the final system state after applying n-repetitions of the protocol and which is dependent on all prior readouts. As described previously, to account for all readout paths, the cost is defined as a sum

$$\langle \mathcal{J} \rangle_r = \sum_r P^n(r) \mathcal{J}(\rho_s^n) \tag{13}$$

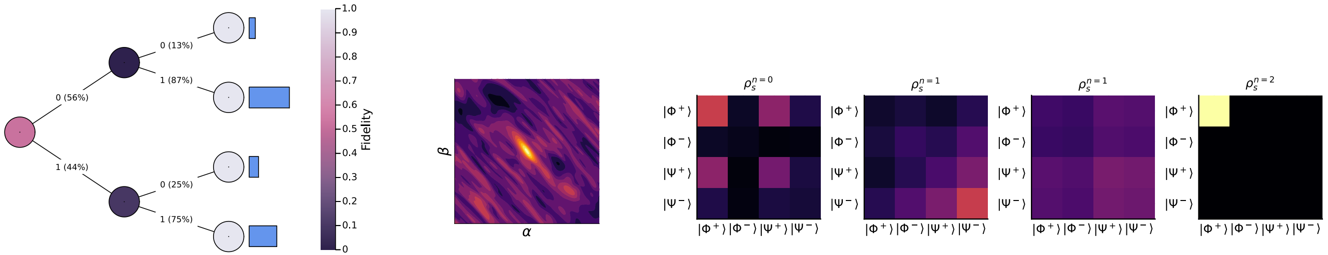
where $P^n(r)$ is the cumulative probability of the path defined through previous readouts r and for the iteration n. In other words, the cost is a sum of all final fidelities (step n) for all possibles sequences of readout outcomes. As we take the gradient with respect to the parameters θ, we note that the derivative does not act on the cumulative probabilities $P^n(r)$. This is due to the protocol being independent to the initial states ρ_s^0. Therefore, the gradient simply defined as

$$\frac{\partial \langle \mathcal{J}(r) \rangle_r}{\partial \theta} = \left\langle \frac{\partial \mathcal{J}(r)}{\partial \theta} \right\rangle_r . \tag{14}$$

In general, the evaluation of the gradient with respect to parameters θ can be done in two ways: using the analytical expression to obtain expressions for the gradients that may be evaluated numerically, or using automatic differentiation. We opt to implement gradient evaluation via automatic differentiation frameworks. This approach is particularly useful when the time evolution described by U can consist of many building blocks, such as a parameterized quantum circuit.

1) Example: Single-Qubit State Preparation: In the simplest case, we assume a qubit is coupled with a detector. An example quantum trajectory of the qubit is shown Figure 3a and Figure 3b for preparing a superposition state

$$|\psi_\oplus\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle + |1\rangle \right). \tag{15}$$



(a) The quantum trajectories produced via measurements of the detector. Each node shows the fidelity of the two-qubit state for the desired $|\Phi^+\rangle$.

(b) The landscape along a two-dimensional slice of the parameter space (three coupling operators U_i) for obtaining the Bell state.

(c) The density matrix of the two-qubit system, ρ_s in the Bell basis as a colormap (absolute value of the matrix elements.) $\rho_s^{n=0}$ is a randomly-chosen initial state and $\rho_s^{n=2}$ is the final state for all four possible outcomes in (a). The two intermediate states in (a) correspond to $\rho_s^{n=1}$, sharing low overlap with the desired Bell state.

Fig. 4: Preparing a two-qubit Bell state, $|\psi_\oplus\rangle = |\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$, via feedback-based steering. The detector (not shown) is a 2-level system that couples with two qubits. In two steps ($N = 2$), resulting in 3 steering operators (dimension 8×8), the protocol was able to prepare the target state.

While it is possible to prepare the state in one iteration ($N = 1$), we will later show that this is susceptible to noise as this is equivalent to passive steering (no feedback is possible). Figure 3a shows the next simplest case ($N = 2$) by visualizing the qubit's state on the Bloch sphere. Figure 3b now shows the convergence of state fidelity using three measurements ($N = 3$).

2) Example: Two-qubit State Preparation: To illustrate the effectiveness of the protocol, we consider a system consisting of two qubits. While the number of detectors required in the passive protocol is three [17], we are able to use a single detector to prepare an arbitrary two-qubit state. The noiseless results for the two-qubit steering are shown in Figure 4. The protocol was able to prepare a specific entangled state starting from a random initial state. The density matrix elements given with respect to the Bell basis: $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$, $|\Phi^-\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$, $|\Psi^+\rangle = (|01\rangle + |10\rangle)/\sqrt{2}$, $|\Phi^-\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$. At the final iteration, the matrix elements spanned by $\{|\Phi^-\rangle, |\Psi^+\rangle, |\Psi^-\rangle\}$ have zero value, while the element spanned by our desired target state $|\psi_\oplus\rangle = |\Phi^+\rangle$ is unity.

B. Reinforcement Learning Passive Steering

Based on the challenges and workflow discussed in Section III-E, we propose a learning paradigm shown in Figure 5 to map the objects in quantum steering onto the five key components of reinforcement learning: *agent*, *environment*, *action*, and *reward*. The *agent*, which interacts with the environment, is chosen as the Euler angles used to compose the operator unitary matrix U. The *environment* is represented as the entire quantum system that receives the composed operator U to perform quantum steering. The state values record all the basic information of the interaction between the current operator U and the quantum system to evaluate the reward computation.

The *action* space is defined as all possible mutations to the Euler angles, which produce an updated U that is subsequently applied to the steering simulation. The reward-based optimization step enables the reinforcement learning model to learn a sophisticated strategy to update initial operators. However, the vast action space of Euler angles, which is encoded as a vector of real number, makes it impractical for encoding and simulation. To address this challenge, we assign a Gaussian distribution to each of the entrees. This produces the offset to the current entrees, and the action is chosen randomly at each step based on the parameterized distribution. This guarantees the coverage of all possible actions. Moreover, the expectation of the current can be precisely computed, making it possible to apply policy gradient to optimize the parameters (μ_i, σ_i) during the training phase.

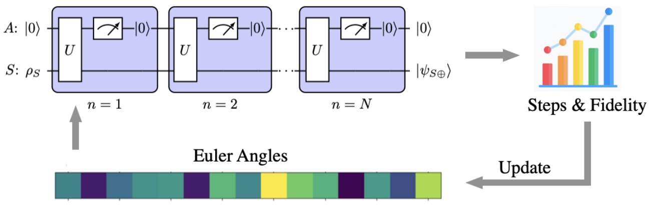


Fig. 5: The reinforcement learning based U generator.

The *reward* is the most important feedback information from the environment that describes the effect of the latest action. It often refers to the benefit of performing the current operation. In our framework, we apply policy gradient, a stochastic approach, to compute and optimize the reward evaluation. To achieve this, the policy is represented by a function, denoted as $\pi_\theta(a|s)$, where s denotes the state and a denotes the action. The parameter θ represents the policy function. $\pi_\theta(a|s)$ is the probability of selecting action a given the state s. The objective function, which is dependent on the policy, determines the value of the reward. Gradient descent is applied to optimize θ and achieve the best reward. The complete reward function is defined as:

$$J(\theta) = \sum_{s \in S} d^\pi(s) \sum_{a \in \mathcal{A}} \pi_\theta(a|s) Q^\pi(s,a)$$

S and $\mathcal{A}$ denote the sets of all states and actions, respectively. $d^\pi(s)$ represents the stationary distribution of Markov chain for the policy function π_θ, which is the on-policy state distribution under π. This implies that the reinforcement learning model continuously travels along the Markov chain's states until it eventually reaches a steady state probability distribution. Formally, this can be expressed as $d^{\pi}(s) = lim_{t\to\infty} P(s_t = s|s_0, \pi_\theta)$. $Q^{\pi}(s,a)$ represents the one-step reward. In this work, temporary Euler angles are applied to manipulate quantum steering for several iterations, and the improvement in terms of fidelity is recorded as the Q value. However, the sets S and $\mathcal{A}$ are uncountable, and it is also impossible to run infinite iterations to obtain an accurate value for d^{π}. Therefore, we approximate $d^{\pi}(s)$ by applying the current policy for ten iterations, i.e., $d^{\pi}(s) = P(s_{10} = s|s_0, \pi_\theta)$. According to the policy gradient theorem [41], the gradient computation can be expressed as:

$$\nabla J(\theta) = \nabla \sum_{s\in S} d^{\pi}(s) \sum_{a\in\mathcal{A}} \pi_\theta(a|s) Q^{\pi}(s,a) \tag{16}$$

$$\propto \sum_{s\in S} d^{\pi}(s) \sum_{a\in\mathcal{A}} \nabla\pi_\theta(a|s) Q^{\pi}(s,a) \tag{17}$$

Since we are using Gaussian distributions to manipulate the Euler angles, θ in our case is $\mathcal{G} = \mathcal{N}_1(\mu_1,\sigma_1), \mathcal{N}_2(\mu_2,\sigma_2), ..., \mathcal{N}_{15}(\mu 15, \sigma_{15})$. In this context, the action involves adding an offset ϵ to the corresponding Euler angle entry. This can be expressed as:

$$\nabla\pi_\theta(a|s) = \nabla\pi_{\mathcal{G}}(s + \{\epsilon_i\}|s), \epsilon_i \sim \mathcal{N}_i(\mu_i, \sigma_i)$$

In actual computation, we apply logarithmic loss for the ease of computation, and by putting all these together, the policy gradient computation can be accommodated as following formula in our case:

$$\nabla J(\theta) = \mathbb{E}_\pi[Q^{\pi}(s,a) \nabla_{\mathcal{G}} ln(\pi_{\mathcal{G}}(s + \{\epsilon_i\}|s))] \tag{18}$$

$$\epsilon_i \sim \mathcal{N}_i(\mu_i, \sigma_i), \;\; i = 1, 2, ..., 15 \tag{19}$$

and the overall training process of proposed RL model is presented in Algorithm 1.

Algorithm 1: RL Training Process

```
Input  : System Qubit ($S$), ancilla detector ($D$)
         Model Parameter ($\mathcal{G}$),number of epochs ($k$)
Output: Optimal Model Parameter $\mathcal{G}^*$
Initialize $S, D, \mathcal{G}, k$, learning rate $\alpha$, decay ratio $\gamma$
Initialize random Euler angles $\mathcal{E}$
$i = j = 0$
repeat
    Initialize Reward: $R = 0$
    $\mathcal{E}$ = $act(\mathcal{E}$, $\mathcal{G})$
    $U = createUnitary(E)$
    repeat
        $fidelity$ = Simulate($S$, $D$, $\mathcal{E}$)
        $R' = R' + \gamma \cdot (1 - fidelity)$
    until $j \geq 10$;
    $R = R + R'$
    Update parameter : $\mathcal{G} = \mathcal{G} + \alpha\nabla_\theta J(R)$
until $i \geq k$;
Return $\mathcal{G}$
```

In summary, we described two complementary strategies for active steering in this section: gradient-based and reinforcement learning. The choice of the strategy depends on the specific configuration and the availability of labeled data. If labeled data (Euler angles-fidelity mappings) is available, then gradient-based method is beneficial due to the high stability and resistance towards noise. However, if labeled data is scarce or non-existent, or if interpretability is important, then reinforcement learning is more appropriate.

V. Experiments

In this section, we evaluate our gradient-based active steering (*GB + Active*) strategy and our reinforcement learning (*RL + Passive*) strategy for preparing quantum states under different noise assumptions.

A. Experimental Setup

The experimental setup involves the use of Qiskit, an open-source quantum computing software development framework, along with a custom Julia library for performing gradient-based optimization of feedback-based quantum steering. We implement noise models in our custom library as well as in Qiskit for reinforcement learning. Our custom library is open-source [42], and utilizes Tullio's [43] flexible Einstein notation to perform operations on tensors while simultaneously providing gradients using automatic differentiation. The objective is to optimize the parameters of steering protocol using quasi-Newton methods, namely limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm [44], to maximize the fidelity of obtaining a specific target state while considering the effects of noise. Our free parameter is the number of iterations N of the protocol. Therefore, for each N, we find an optimal solution for all steering operators.

The reinforcement learning model in our approach was conducted on a host machine with Intel i7 3.70GHz CPU, 32 GB RAM and RTX 3090 256-bit GPU. We choose Python code using PyTorch with cudatoolkit (10.0) to implement the machine learning framework. The total training process consisted of 500 epochs.

We investigate the performance of a feedback-based steering protocol for preparing an arbitrary target state of a qubit coupled to a detector. The protocol involves a closed-loop control technique that utilizes feedback to steer the system towards the target state. Specifically, we consider a detector that is coupled to the system – a qubit or two-qubits – and employ our two strategies for closed-loop control technique: gradient-based active steering that changes the detector-system coupling after each measurement, and tuning the fixed detector-system coupling via model-free reinforcement learning passive steering. We demonstrate the results of the approaches on decoherent and incoherent noise models.

B. Decoherent Noise on the Detector

In this model, we assume that the detector can not be initialized to a perfect pure state $|0\rangle$, as discussed in

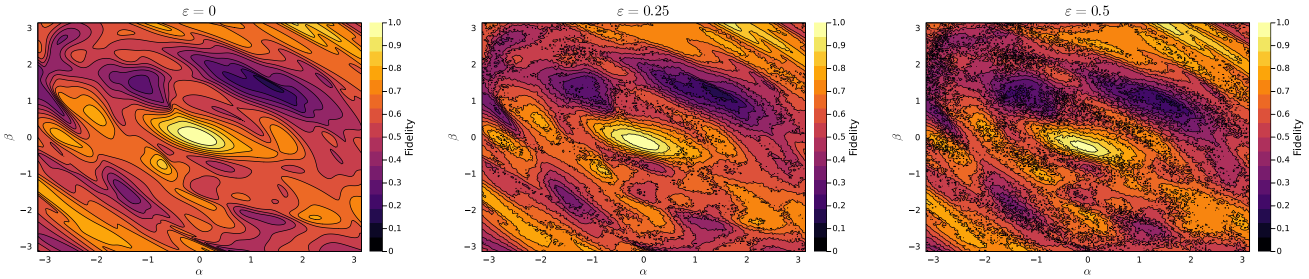


Fig. 6: The fidelity landscape with respect to a two-dimensional slice of the parameter space. Gradient-based optimization must find the global maxima of the fidelity landscape in order to determine the optimal values for the parameters. The fidelity landscape provides insights into the sensitivity of the system to changes in the parameters. Namely, the feedback-based steering protocol is locally stable, containing a range of parameters that yield high fidelity. While locally stable, the addition of noise introduces islands of local maximas.

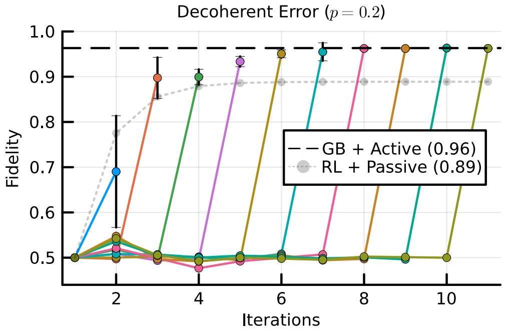


Fig. 7: The average fidelity of the gradient-based active steering versus reinforcement learning passive steering in the presence of faulty detector initialization. The depolarizing error for the detector initialization is fixed to $p = 0.2$. The reinforcement learning protocol incorrectly converges to a fidelity of 89% while the gradient-based protocol converges to 96% in 10 iterations. Each color line indicates an experiment performed with a specific number of iterations N. The error bars are taken from 100 samples of each experiment.

Section II-B. Instead, the detector is assumed to undergo decoherence, and will be in a mixed state. The extent of noise is parameterized by p, as defined in Equation 8.

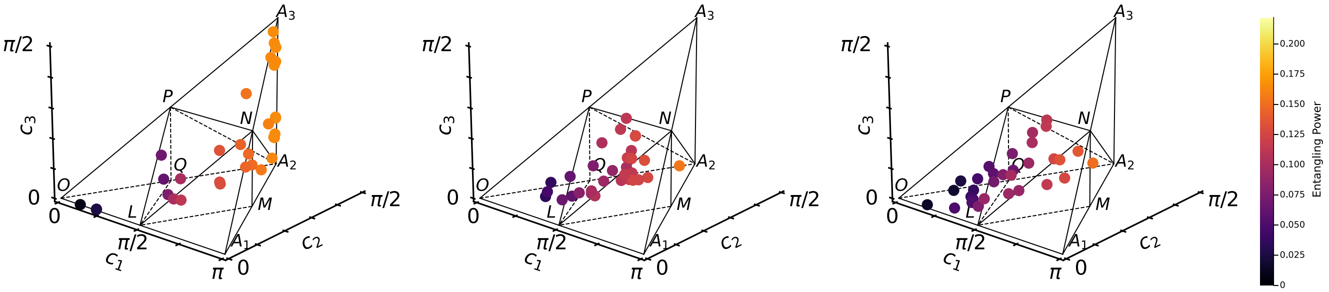
We found that the gradient-based active steering was able to overcome a faulty detector and achieve high fidelity over longer iterations. This was achieved by optimizing the steering operator based on feedback information obtained from noisy measurements. As shown in Figure 7, as the number of iterations increased, the protocol was able to refine the steering operator further, resulting in 96% fidelity in 10 iterations of the protocol. In comparison, the reinforcement learning passive steering plateaued at 89% fidelity. The optimizer attempts to minimize entropy from the composite system as a result of a mixed detector state via gates that have high entanglement power, as shown in Figure 8a.

In contrast, the reinforcement learning strategy was not able to extract the entropy and obtain lower fidelity. This is because the steering operator remained fixed (passive) with each iteration, and the reinforcement learning algorithm has no mechanism to modify it to account for the noise. In other words, because the system state is an average of all readout outcomes, the entropy remains fixed. Meanwhile, the gradient-based active steering can optimize each operator and produce an outcome that one quantum trajectory occurs with high probability, whereas the unwanted entropy is spread across the remaining low-probability trajectories. Our results demonstrate the effectiveness of feedback-based strategies in quantum systems, as they can adapt and optimize in real-time to overcome various sources of noise and errors. This has important implications for the development of quantum technologies, as it provides a way to mitigate the effects of noise and errors in real-time, improving the reliability and robustness of quantum devices.

C. *Incoherent Noise*

In addition to investigating the protocol's effectiveness in overcoming decoherent noise on the detector, we also tested its ability to handle incoherent noise. Specifically, we assumed the steering operator was perturbed with a randomly noisy unitary selected from a fixed set, as outlined in Equation 9. We tested the protocol on different strengths of noise. Figure 6 shows a two-dimensional slice of the landscape corresponding to different noise strengths that the gradient-based optimizer needs to traverse. From the landscape we note two key properties: (1) the protocol is resilient to small perturbations to the steering unitaries – in other words, a quantum device has leeway in implementing a unitary operator; (2) an increase in incoherent noise strength, corresponds to a growth in the number of local minima and maxima, which lowers the extent of perturbation resilience. Figure 9 shows that the feedback-based steering protocol is able to achieve high fidelity even in the presence of incoherent noise. Even with an increase of noise, the protocol was able to reach and maintain a high fidelity of 99%. In comparison, reinforcement learning passive steering obtains significantly lower fidelity and with high variance.

The gradient-based active steering protocol's effectiveness in handling incoherent noise is due to its ability to adapt and optimize the steering operator based on feedback information obtained from the noisy measurements. By adjusting the steering operator in real-time to account for the perturbations



(a) Decoherence of detector ($p = 0.25$) (b) Incoherent gate noise ($\epsilon = 0.25$) (c) Incoherent gate noise ($\epsilon = 0.5$)

Fig. 8: Weyl chambers representing the non-local parts of the optimized steering operators U_i under different noise parameters. The number of iterations in the protocol is fixed to $N = 5$. The color of each gate represents its entanglement power which is related to the coordinates in the Weyl chamber [45]. In the presence of decoherence, majority of the gates require high amounts of entanglement power as the protocol attempts to dispel the entropy from composite detector-system. The gates in the presence of incoherent noise require less entanglement power, as the entropy of the composite detector-system remains fixed (but changes for the system). For higher noise levels, the variation in the gates, and the entanglement power, increases.

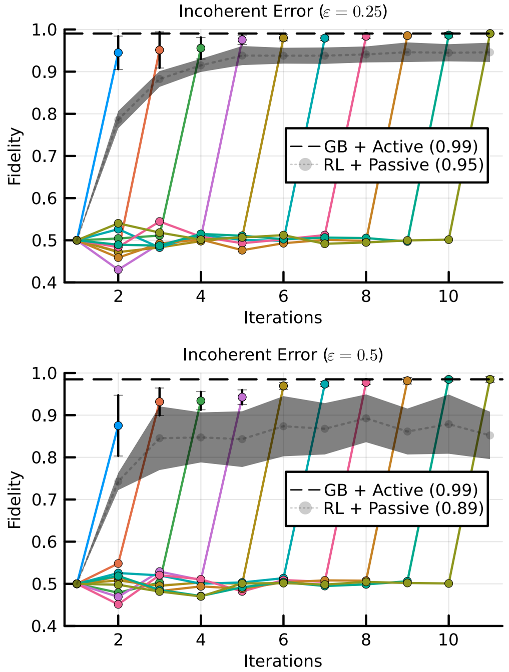


Fig. 9: Resilience of feedback-based steering in the presence of incoherent gate noise. Each steering operator is perturbed by a unitary that is randomly selected from a discrete set of unitary operators. The strength of the perturbation is given in terms of ϵ. Each color line indicates an experiment performed with a specific number of iterations N. The error bars are taken from 100 samples of each experiment. With a sufficient number of iterations, the protocol is able to overcome the noise and approaches a state fidelity 99%. At lower noise strengths, the protocol achieves higher state fidelity in fewer iterations. Meanwhile, the reinforcement learning passive steering approach is unable to reach a suitable average fidelity and suffers from fidelity variation that scales with the strength of noise.

introduced by the incoherent noise, the protocol was able to mitigate its effects and maintain high final fidelity. In particular, the protocol converges to the desired target state with low variance. Figure 8b and Figure 8c shows the gates in the Weyl chamber. At low noise levels, the majority of gates are perfect entanglers. At higher noise levels, the gates tend to spread across the Weyl chamber, necessitating a variety of entanglement operations. Unlike in the decoherent case, high entangling power gates are minimal. We note that it may be difficult to achieve certain types of entangling gates on noisy devices. But, a trade-off may be made by penalizing unwanted gates in the loss function.

VI. Conclusion

In this work, we have introduced a framework for feedback-based steering, featuring two primary strategies: gradient-based optimization and reinforcement learning. These methods optimize the detector-system coupling, such that the system is steered toward desired state due to the measurement backaction of the detector. Our findings indicate that gradient-based active steering is an effective approach for state preparation in quantum systems, even in the presence of noisy measurements and incoherent noise. This insight holds significant implications for the advancement of quantum technologies, as it offers a robust control mechanism for contemporary quantum architectures.

The active steering strategy necessitates substantial computational resources for higher-dimensional systems. In contrast, reinforcement learning-based passive steering is capable of learning a specific objective with limited resources. However, it achieves lower fidelity under the influence of noise, underscoring the importance of incorporating an active mechanism to modify the detector-system coupling. We emphasize the need for future research to investigate and integrate both algorithms, leveraging their respective strengths as outlined in this paper. Our research lays a solid groundwork for feedback-based steering and highlights its potential in preparing quantum states amidst noise.

REFERENCES

[1] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, "Optimal control of coupled spin dynamics: Design of NMR pulse sequences by gradient ascent algorithms," *Journal of Magnetic Resonance*, vol. 172, no. 2, pp. 296–305, Feb. 2005.

[2] N. Anders Petersson and F. Garcia, "Optimal Control of Closed Quantum Systems via B-Splines with Carrier Waves," *SIAM J. Sci. Comput.*, vol. 44, no. 6, pp. A3592–A3616, Dec. 2022.

[3] Daniel, "RustyBamboo/Pulses.jl: V0.0.1," Zenodo, Dec. 2022.

[4] D. Volya and P. Mishra, "Quantum data compression for efficient generation of control pulses," in *Asia and South Pacific Design Automation Conference (ASPDAC)*, 2023, pp. 216–221.

[5] F. F. Floether, P. de Fouquieres, and S. G. Schirmer, "Robust quantum gates for open systems via optimal control: Markovian versus non-Markovian dynamics," *New J. Phys.*, vol. 14, no. 7, p. 073023, Jul. 2012.

[6] D. Dong and I. R. Petersen, "Quantum control theory and applications: A survey," *IET Control Theory &amp; Applications*, vol. 4, no. 12, pp. 2651–2671, Dec. 2010.

[7] D. Volya and P. Mishra, "Impact of noise on quantum algorithms in NISQ systems," in *IEEE International Conference on Computer Design (ICCD)*, 2020.

[8] M. D. Reed, L. DiCarlo, S. E. Nigg, L. Sun, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, "Realization of three-qubit quantum error correction with superconducting circuits," *Nature*, vol. 482, no. 7385, pp. 382–385, Feb. 2012.

[9] A. D. King, J. Carrasquilla, J. Raymond, I. Ozfidan, E. Andriyash, A. Berkley, M. Reis, T. Lanting, R. Harris, F. Altomare, K. Boothby, P. I. Bunyk, C. Enderud, A. Fréchette, E. Hoskinson, N. Ladizinsky, T. Oh, G. Poulin-Lamarre, C. Rich, Y. Sato, A. Y. Smirnov, L. J. Swenson, M. H. Volkmann, J. Whittaker, J. Yao, E. Ladizinsky, M. W. Johnson, J. Hilton, and M. H. Amin, "Observation of topological phenomena in a programmable lattice of 1,800 qubits," *Nature*, vol. 560, no. 7719, pp. 456–460, Aug. 2018.

[10] M. Takita, A. W. Cross, A. D. Córcoles, J. M. Chow, and J. M. Gambetta, "Experimental Demonstration of Fault-Tolerant State Preparation with Superconducting Qubits," *Phys. Rev. Lett.*, vol. 119, no. 18, p. 180501, Oct. 2017.

[11] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets," *Nature*, vol. 549, no. 7671, pp. 242–246, Sep. 2017.

[12] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis, "Scalable Quantum Simulation of Molecular Energies," *Phys. Rev. X*, vol. 6, no. 3, p. 031007, Jul. 2016.

[13] B. P. Lanyon, C. Hempel, D. Nigg, M. Müller, R. Gerritsma, F. Zähringer, P. Schindler, J. T. Barreiro, M. Rambach, G. Kirchmair, M. Hennrich, P. Zoller, R. Blatt, and C. F. Roos, "Universal Digital Quantum Simulation with Trapped Ions," *Science*, vol. 334, no. 6052, pp. 57–61, Oct. 2011.

[14] G. A. Paz-Silva, A. T. Rezakhani, J. M. Dominy, and D. A. Lidar, "Zeno Effect for Quantum Computation and Control," *Phys. Rev. Lett.*, vol. 108, no. 8, p. 080501, Feb. 2012.

[15] D. Basilewitsch, J. Fischer, D. M. Reich, D. Sugny, and C. P. Koch, "Fundamental bounds on qubit reset," *Phys. Rev. Res.*, vol. 3, no. 1, p. 013110, Feb. 2021.

[16] J. Preskill, "Quantum Computing in the NISQ era and beyond," *Quantum*, vol. 2, p. 79, Aug. 2018.

[17] S. Roy, J. T. Chalker, I. V. Gornyi, and Y. Gefen, "Measurement-induced steering of quantum systems," *Phys. Rev. Research*, vol. 2, no. 3, p. 033347, Sep. 2020.

[18] Y. Herasymenko, I. Gornyi, and Y. Gefen, "Measurement-driven navigation in many-body Hilbert space: Active-decision steering," Oct. 2022.

[19] P. Kumar, K. Snizhko, and Y. Gefen, "Engineering two-qubit mixed states with weak measurements," *Phys. Rev. Res.*, vol. 2, no. 4, p. 042014, Oct. 2020.

[20] P. Kumar, K. Snizhko, Y. Gefen, and B. Rosenow, "Optimized steering: Quantum state engineering and exceptional points," *Phys. Rev. A*, vol. 105, no. 1, p. L010203, Jan. 2022.

[21] D. Volya and P. Mishra, "State Preparation on Quantum Computers via Quantum Steering," *arXiv:2302.13518*, Mar. 2023.

[22] D. Ristè, C. C. Bultink, K. W. Lehnert, and L. DiCarlo, "Feedback Control of a Solid-State Qubit Using High-Fidelity Projective Measurement," *Phys. Rev. Lett.*, vol. 109, no. 24, p. 240502, Dec. 2012.

[23] N. Yamamoto, "Coherent versus Measurement Feedback: Linear Systems Theory for Quantum Information," *Phys. Rev. X*, vol. 4, no. 4, p. 041029, Nov. 2014.

[24] R. Porotti, V. Peano, and F. Marquardt, "Gradient Ascent Pulse Engineering with Feedback," Mar. 2022.

[25] B. Kraus, H. P. Büchler, S. Diehl, A. Kantian, A. Micheli, and P. Zoller, "Preparation of entangled states by quantum Markov processes," *Phys. Rev. A*, vol. 78, no. 4, p. 042307, Oct. 2008.

[26] F. Verstraete, M. M. Wolf, and J. Ignacio Cirac, "Quantum computation and quantum-state engineering driven by dissipation," *Nature Phys*, vol. 5, no. 9, pp. 633–636, Sep. 2009.

[27] H. Weimer, M. Müller, I. Lesanovsky, P. Zoller, and H. P. Büchler, "A Rydberg quantum simulator," *Nature Phys*, vol. 6, no. 5, pp. 382–388, May 2010.

[28] C. Ahn, A. C. Doherty, and A. J. Landahl, "Continuous quantum error correction via quantum feedback control," *Phys. Rev. A*, vol. 65, no. 4, p. 042301, Mar. 2002.

[29] E. Schrödinger, "Discussion of Probability Relations between Separated Systems," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 31, no. 4, pp. 555–563, Oct. 1935.

[30] T. Tilma and E. C. G. Sudarshan, "Generalized Euler angle parametrization for SU(N)," *J. Phys. A: Math. Gen.*, vol. 35, no. 48, p. 10467, Nov. 2002.

[31] Y. Makhlin, "Nonlocal Properties of Two-Qubit Gates and Mixed States, and the Optimization of Quantum Computations," *Quantum Information Processing*, vol. 1, no. 4, pp. 243–252, Aug. 2002.

[32] J. Zhang, J. Vala, S. Sastry, and K. B. Whaley, "Geometric theory of nonlocal two-qubit operations," *Phys. Rev. A*, vol. 67, no. 4, p. 042313, Apr. 2003.

[33] P. Zanardi, C. Zalka, and L. Faoro, "Entangling power of quantum evolutions," *Phys. Rev. A*, vol. 62, no. 3, p. 030301, Aug. 2000.

[34] S. Balakrishnan and R. Sankaranarayanan, "Entangling power and local invariants of two-qubit gates," *Phys. Rev. A*, vol. 82, no. 3, p. 034301, Sep. 2010.

[35] E. P. Wigner, "Characteristic vectors of bordered matrices with infinite dimensions i," *The Collected Works of Eugene Paul Wigner: Part A: The Scientific Papers*, pp. 524–540, 1993.

[36] M. L. Mehta, *Random matrices*. Elsevier, 2004.

[37] Z. Pan and P. Mishra, "Automated test generation for hardware trojan detection using reinforcement learning," in *Asia and South Pacific Design Automation Conference (ASPDAC)*, 2021, pp. 408–413.

[38] A. Goldie and A. Mirhoseini, "Placement optimization with deep reinforcement learning," *CoRR*, vol. abs/2003.08445, 2020.

[39] C. Wu *et al.*, "Explore deep neural network and reinforcement learning to large-scale tasks processing in big data," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 33, no. 13, pp. 1 951 010:1–1 951 010:29, 2019.

[40] S. Khairy, R. Shaydulin, L. Cincio, Y. Alexeev, and P. Balaprakash, "Reinforcement-learning-based variational quantum circuits optimization for combinatorial problems," *CoRR*, vol. abs/1911.04574, 2019.

[41] R. S. Sutton, D. McAllester, S. Singh, and Y. Mansour, "Policy gradient methods for reinforcement learning with function approximation," *Advances in neural information processing systems*, vol. 12, 1999.

[42] D. Volya, "Measurement-induced quantum steering with feedback." https://github.com/RustyBamboo/MIQSback.jl, 2023.

[43] M. Abbott, D. Aluthge, N3N5, S. Schaub, C. Elrod, C. Lucibello, and J. Chen, "Mcabbott/Tullio.jl: V0.3.5," Zenodo, Sep. 2022.

[44] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Mathematical Programming*, vol. 45, no. 1, pp. 503–528, Aug. 1989.

[45] S. Balakrishnan and R. Sankaranarayanan, "Measures of operator entanglement of two-qubit gates," *Phys. Rev. A*, vol. 83, no. 6, p. 062320, Jun. 2011.