利用读出信号实时反馈的深度强化学习实现快速量子门设计

Emily Wright物理与天文学系
维多利亚大学维多利亚，加拿大
emilywright1@uvic.ca

Rog´erio deSousa物理与天文学系 维多利亚大学维多利亚，加拿大 rdesousa@uvic.ca

arXiv:2305.01169v1 [quant-ph] 2 May 2023

摘要——**高保真量子门的设计颇具挑战，因为它需要优化两个相互竞争的目标，即提高量子门速度并尽量减少泄漏出量子比特子空间的情况。我们提出一种深度强化学习算法，利用两个智能体同时应对速度和泄漏方面的挑战。第一个智能体根据奖励学习策略，构造量子比特的同相控制脉冲；这些奖励用于补偿较短的门时间。在构造完整脉冲的过程中，系统会在中间时间步提供奖励，使智能体能够探索更短脉冲的优化空间。第二个智能体则确定一个正交相位脉冲，以抑制泄漏。两个智能体都使用来自含噪硬件的实时数据进行训练，因此能够实现无模型量子门设计，并适应不可预测的硬件噪声。为减轻测量分类误差的影响，智能体直接使用探测量子比特所获得的读出信号进行训练。我们通过在IBM硬件上设计不同时长的 X 门和 X门平方根，展示了概念验证实验。仅经过 200 次训练迭代，我们的算法就能够构造出速度最高达到 IBM 默认量子门两倍的新型控制脉冲，同时在态保真度和泄漏率方面达到相当的性能。随着自定义控制脉冲时长增加，其性能开始超过默认量子门。提高量子门操作的速度和保真度，有望提升量子模拟、量子化学及其他算法在近期和未来量子器件上的电路深度。**

索引词—**超导量子比特、最优控制、强化学习**

本研究通过加拿大自然科学与工程研究委员会（NSERC）的 Discovery（RGPIN-2020-04328）、CREATE（543245-2020）和 CGS M 项目获得资助。此项工作由加拿大自然科学与工程研究委员会（CRS NG）通过其 Discovery（RGPIN-2020-04328）、CREATE（543245-2020）和 CGS M 项目提供支持。

I. 引言

硬件噪声、制造差异以及不完善的逻辑门，是大规模执行可靠量子计算的最大障碍 [1]。目前，IBM 等公司生产的量子计算机上的量子门操作，是通过基于简单三能级模型进行解析计算的绝热门导数消除（DRAG）脉冲来实现的 [2]。由于门时间较长、模型不完善，以及处理器参数（例如量子比特频率）随时间发生变化，这些量子门操作的保真度会受到影响。为了应对这些波动而进行频繁校准的成本很高，即使经过适当校准，控制脉冲形状仍未达到最优，因而容易产生误差。在容错量子计算领域，退相干和从计算子空间中发生的泄漏尤其值得关注，因为纠正这些问题需要大量额外资源，而且可能显著影响某些纠错码的阈值 [3]–[6]。因此，设计更快速且具有更高保真度的量子门具有重要的现实意义。

现有的量子门集设计策略要么采用解析方法 [2], [7]–[10] ，要么基于数值模拟，而这类模拟需要精确的硬件物理模型和噪声模型 [11]–[17]。对于大规模量子处理器，难以对系统进行全面表征，因此无法采用基于模型的控制技术。强化学习（RL） [18]–[20]是一种量子门设计的替代方法，无需预先了解硬件模型即可运行。强化学习及其变体已在数值模拟环境中应用于众多量子控制问题 [21]–[37]。这类设置展现了强化学习的潜力，但也受到与其他优化方法相同的建模限制。近期，已有少数实验直接在含噪硬件上使用强化学习 [38]–[41]。

本文提出一种新的强化学习算法，用于设计快速量子门。与现有方案相比，我们的算法具有多项优势，包括：

1) 通过奖励控制脉冲中的中间步骤，实现更快量子门的设计；2) 通过直接基于读出信号进行训练，而非对状态进行分类，降低测量误差的影响，

3）使用双智能体架构缓解泄漏，4）利用低测量开销和实时反馈加速训练，以及 5）通过直接在硬件上进行训练，考虑量子处理器中的真实噪声。

此外，我们通过在经过校准的 DRAG 脉冲上进行预训练来初始化智能体，以捕获有关系统动力学的信息。作为概念验证，我们优化了持续时间不同的 X 和 $\sqrt{X}$ 门；不过，只需修改奖励函数和状态空间，我们的算法便可轻松扩展到双量子比特门，从而完成一个通用门集。

本文的结构如下：在第 II 节中，我们介绍一般的强化学习算法；在第 III节中，我们回顾相关文献；在第 IV节中，我们描述用于快速量子门设计的深度强化学习算法；在第 V 节中，我们呈现实验结果；在第 VI节中，我们介绍将我们的算法扩展到双量子比特门的方法。

II. 强化学习算法

本节将详细介绍强化学习算法。强化学习是一种 Machine Learning 算法，其中智能体与系统交互，并根据观测结果迭代更新控制策略。整个过程被建模为一个受控马尔可夫决策过程（MDP）。设 $S \subset \mathrm{R}^n$ 为系统的状态空间，U 为可能动作的集合。在每一步中，系统处于某个状态 s_j，智能体根据策略决定采取某个动作 u_j。策略 $\pi(\cdot|s_j)$ 是在给定当前状态的条件下，对 U 中可能动作的条件概率分布。系统通过随机转移核 $\mathcal{T}(\cdot|s_j, u_j)$ 转移到下一个状态 s_{j+1}。观测到下一个状态后，智能体获得相应的奖励 $r_j(s_j, u_j, s_{j+1})$。控制器的目标是最大化无限时域折扣期望奖励

$$J_\beta(s_0, \pi) = E_{s_0}^{\mathcal{T},\pi}\left[\sum_{j=0}^{\infty} \beta^j r_j(s_j, u_j, s_{j+1})\right] \tag{1}$$

在可行策略集合 π 上，其中 $0 < \beta < 1$ 是折扣因子，$E^{\mathcal{T},\pi}_{s_0}$ 表示在策略 π 下，对初始状态 s_0 和转移核 $\mathcal{T}$ 取期望。

标准的强化学习算法是 Q学习 [19]。Q学习通过跟踪在当前状态 s_j 下采取动作 u_j 的“价值”来进行。Q值存储在一个以状态和动作作为索引的表中。给定一个初始表 Q_0，每个状态-动作对的价值都根据贝尔曼方程进行更新

$$Q_{j+1}(s_j, u_j) = Q_j(s_j, u_j) + \alpha_j(s_j, u_j)\Big[r(s_j, u_j, s_{j+1}) + \beta \max_{v\in\mathbb{U}} Q_j(s_{j+1}, v) - Q_j(s_j, u_j)\Big] \tag{2}$$

随着智能体探索环境 [19]。系数 α_j 是一个称为学习率的超参数，它决定智能体适应环境变化的速度。在满足学习率的温和条件时，该算法会收敛到一个记作 Q_* 的不动点，该不动点满足

$$Q_*(s, u) = E\left[r(s, u, s') + \beta \max_{v\in\mathbb{U}} Q_*(s', v)\Big|s, u\right]. \tag{3}$$

满足条件的策略 π

$$\max_{u\in\mathbb{U}} Q_*(s, u) = Q_*(s, \pi(s)) \tag{4}$$

是一个最优策略（参见 [18] 中的定理 4 和 [19] 中的主定理）。

Q学习算法最初是为有限的动作空间和状态空间设计的。对于大型和/或连续状态空间，将 Q 值存储在表中并不可行。为克服这一挑战，可以对状态空间进行量化 [42]，或者使用函数近似 [43]。虽然采用量化或函数近似的 Q 学习并不能保证收敛，但大量实验证据表明，它可以用于解决量子控制问题 [21]–[41]。在本研究中，我们使用神经网络近似 Q 表，这种策略被称为“深度强化学习”（DRL） [44], [45]。状态 s_j 是神经网络的输入，输出则是动作空间 U 上的概率分布 $\pi(\cdot|s_j)$。神经网络由一组参数 θ 表示，这些参数的更新方式旨在近似贝尔曼方程 [46]。

III. 用强化学习进行量子门设计

介绍强化学习之后，我们现在回顾其迄今在量子门设计中的应用。人们提出了许多强化学习算法，用于解决从哈密顿量工程 [23] 到量子计量学 [22]等领域中的量子控制问题。理论算法利用模拟环境，全面获取量子系统 [21]–[30]的状态信息。在这些方案中，有几种专门针对酉门设计 [25], [47], [48]。在所有情况下，智能体都能获取精确的酉算符，该算符描述了系统的薛定谔演化。在 [25], [47] 中使用了简化的哈密顿量模型，而 [48] 则模拟了 gmon环境，该环境模仿含噪控制驱动并纳入泄漏错误。智能体根据门不保真度获得奖励，而门不保真度在实验中无法获取。在模拟中，与其他量子门合成策略相比，这些算法能够提高门保真度 [25], [47], [48]并缩短门时间 [25], [48]。然而，这些方案无法兼容在真实硬件上进行训练，因此必然会受到模型偏差的影响。更贴近实际的量子控制强化学习设置只能获取保真度和/或某些可观测量的期望值 [31]–[37]。具体到门设计，Shindi等人提出在 [31]中通过一系列输入态来探测量子门。奖励中纳入了输出态与目标态之间的平均保真度。这类算法在实验中需要进行大量重复平均，因此也只能用于数值模拟。

下一步是利用随机测量结果或物理可观测量的低样本估计量 [38]–[41]进行强化学习。最近，一些实验已直接在含噪量子硬件 [38]，[39]上开展强化学习。Baum等人在一台超导计算机上训练了一个深度强化学习智能体，用于抗误差门集设计 [39]。该智能体能够学习新颖的脉冲形状，速度比业界标准 DRAG 门快至三倍，而且每门错误率略低

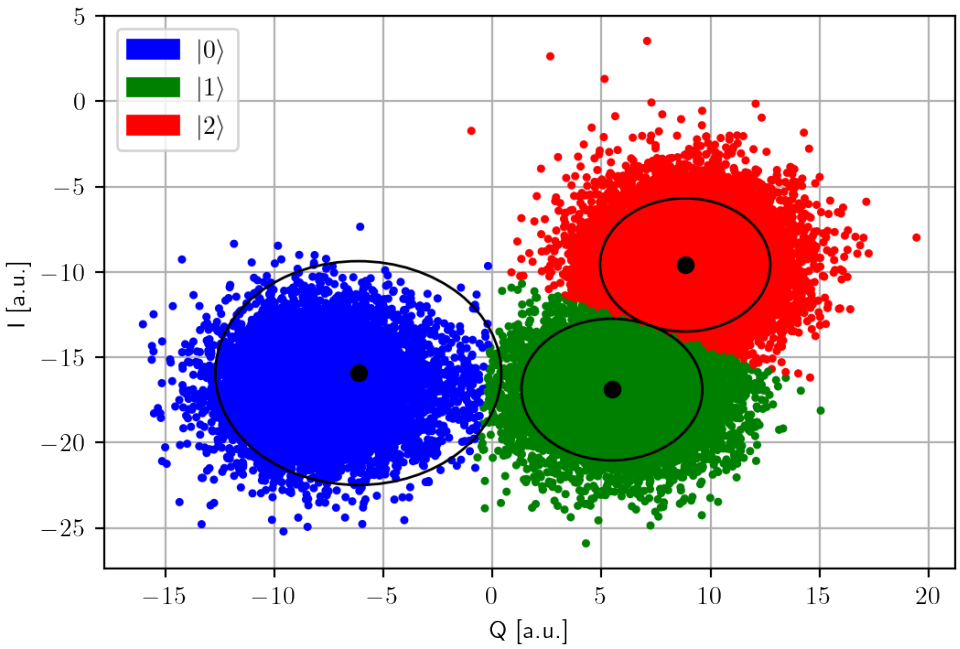


图 1：IBM Lima 上不同量子比特态的读出信号图（每种状态各测量 10000 次）。黑色圆圈表示每个簇的均值和标准差。信号重叠表明区分不同状态的难度。

且改进效果在无需重新校准的情况下最多维持了 25 天。搜索空间被限制为用于单量子比特门和双量子比特门的 8 段和 10 段分段分段常数操作。尽管搜索空间很小，优化算法的效率仍然不高。训练分批次完成，奖励是根据完整量子态层析估算的保真度加权均值。随后，Reuer 等人开发了一种基于延迟的深度强化学习智能体，通过现场可编程门阵列实现，能够利用微秒时间尺度的实时反馈 [38]。他们通过一项超导量子比特上的量子态制备实验，展示了该方法的有效性。本文展望了实时反馈在量子控制中普及的未来。我们在 [39] 的工作基础上进行了改进：通过提高脉冲形状的灵活性、直接依据读出信号进行训练以减小测量误差的影响、专门针对泄漏错误，并依靠实时反馈加快学习过程。

IV. 用于快速量子门设计的深度强化学习算法

现在，我们详细介绍用于设计快速量子门的深度强化学习算法。在这里，我们针对超导传输子量子比特具体阐述该算法，但也可以轻松将其调整到其他硬件平台，例如囚禁离子 [49],量子点 [50], 或中性原子 [51]。在这种情况下，量子系统由一个色散耦合到超导谐振器、并通过电容耦合的电压驱动线进行控制的 transmon量子比特组成（参见附录 A 以了解详细信息）。电压驱动包络

$$c(t) = \begin{cases} c^x(t)\cos(\omega_d t) + c^y(t)\sin(\omega_d t) & 0 < t < t_g \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

由单个驱动频率 ω_d 下的两个独立正交分量控制 $c^x(t)$ 和 $c^y(t)$ 构成，持续时间为量子门操作 t_g。我们旨在设计一个分段常数控制脉冲，包含至多 N_{seg} 个等长分段 $\tau = t_g/N_{\text{seg}}$，其中 t_g 为量子门的最长持续时间。

我们的深度强化学习算法由两个神经网络组成，依次决定控制脉冲的 x分量和 y正交分量。双智能体结构使我们能够减轻泄漏错误，同时生成更快的量子门。在每个时间步，智能体根据系统状态的实时反馈，选择下一个脉冲段的振幅。在构造脉冲的整个过程中（而非仅在结束时），智能体都会获得奖励，因此有望设计出更快的量子门。我们通过在量子硬件上进行训练来消除模型偏差，使深度强化学习算法能够应对量子比特中的真实噪声，以及经典驱动线引入的过旋转等其他误差。我们直接根据探测量子比特所获得的观测信号进行训练。该信号包含两个分量，我们分别用 I 表示“同相”分量，用 Q 表示“正交分量”。以往的量子门设计算法根据信号在 $|0\rangle$ 与 $|1\rangle$ 布居在(I, Q)平面上的位置，对其进行不精确估算。图1展示了在 IBM Lima 量子计算机上采集的读出数据，其中$|0\rangle$、 $|1\rangle$和 $|2\rangle$ 测量结果的位置彼此重叠。我们的算法通过避免将

信号分类为 $|0\rangle$ 或 $|1\rangle$，降低了测量误差的影响。由于无需执行完整的量子态层析，测量开销也较低。

网络参数 θ^x, θ^y 会在训练过程中定期更新。对于 $i = 0, \ldots, N_{\text{iter}} - 1$，每次训练迭代包含 $N_{\text{ep}} \leq N_{\text{seg}}$ 个回合，其计数索引为 j。比值 $N_{\text{seg}}/N_{\text{ep}}$ 为一个整数，用于设定在构建完整波形之前神经网络参数更新的次数。我们使用第二个索引 $k = k(i, j)$来跟踪波形片段。第 j 个输入第一个神经网络的状态为 $s_j^x = (\langle I_j \rangle, \langle Q_j \rangle, k)$，由经过 N_{shot} 次测量得到的平均 I, Q信号和当前片段索引k组成。智能体的输出是动作空间 U^x 上的概率分布，我们将其记为 $\pi_{\theta_i^x}(\cdot|s_j^x)$，以表示其对神经网络参数 θ_i^x的依赖。智能体根据 $\pi_{\theta_i^x}(\cdot|s_j^x)$采样动作 $u_j^x \in \mathrm{U}^x$。第二个智能体的状态 $s_j^y = (u_j^x, \mathcal{L}_j)$由第一正交分量上的振幅$u_j^x$ 和泄漏布居组成

$$\mathcal{L}_j = |\langle 2| U_{k+1} |0\rangle|^2 \tag{6}$$

从 (I, Q) 平面估计得出，其中 U_{k+1} 表示通过波形的前 $k+1$ 个分段演化量子比特的操作。现在，智能体根据第二个网络的输出 $\pi_{\theta_i^y}(\cdot|s_j^y)$ 采样一个动作 $u_j^y \in \mathrm{U}^y$。因此，控制脉冲的第 k 个分段形式为

$$c(t) = u_j^x \cos(\omega_d t) + u_j^y \sin(\omega_d t) \tag{7}$$

对于 $k\tau \leq t < (k+1)\tau$。根据硬件参数，我们将 x 正交分量振幅限制在集合 $\mathrm{U}^x = \{0.00, 0.01, \ldots, 0.19, 0.20\}$ 中，并将 y 正交分量振幅限制在 $\mathrm{U}^y = \{-0.10, -0.09, \ldots, 0.09, 0.10\}$ 中。

为了训练智能体，我们为每个网络定义一个奖励函数。令 $c_T = (\langle I_T \rangle, \langle Q_T \rangle) \pm \sigma_T$ 表示将目标门作用于基态后（经实验校准）的预期平均测量结果。在训练期间，我们惩罚观测到的测量信号与 $(\langle I_T \rangle, \langle Q_T \rangle)$ 之间的距离，该量同时反映了未能将量子比特导入目标状态以及向更高能级泄漏的情况。我们使用如下奖励函数

$$r_j^x\left(s_{j+1}^x, c_T\right) = \min\left\{1 - \lambda k, \frac{\sigma_T}{\|(\langle I_{j+1}\rangle, \langle Q_{j+1}\rangle) - (\langle I_T\rangle, \langle Q_T\rangle)\|} - \lambda k\right\} \tag{8}$$

用于训练第一个网络，其中项 λk 会以系数 $\lambda \in \mathbb{R}$ 对控制脉冲的长度进行惩罚。对于第二个网络，我们通过奖励来补偿较低的泄漏布居

$$r_j^y\left(s_{j+1}^y\right) = \max\{0, \mathcal{L}_{\max} - \mathcal{L}_{j+1}\} \tag{9}$$

其中 $\mathcal{L}_{\max}$ 设定了允许泄漏的上限。我们的深度强化学习算法总结如算法 1所示。

通过在执行任何量子门操作之前测量量子比特，估算初始状态 s_0^x 和泄漏布居 $\mathcal{L}_0$。通过对智能体进行预训练，基于行业标准 DRAG 门生成初始策略 $\pi_{\theta_0}^x, \pi_{\theta_0}^y$。DRAG 脉冲为高斯脉冲，并在第二个分量上包含一个导数项。

算法 1 用于快速量子门设计的深度强化学习

要求： 初始状态 s_0^x、泄漏 $\mathcal{L}_0$ 和参数 θ_0^x, θ_0^y

设置 $k = 0$
对于 每次迭代 $i = 0, \ldots, N_{\text{iter}} - 1$ **执行**
  对于每个回合 $j = 0, \ldots, N_{\text{ep}} - 1$ **执行**
    如果 $k = N_{\text{seg}}$ **then**
      重新初始化波形 $k = 0$
      重新初始化状态 $s_j^x = s_0^x$
      重新初始化泄漏 $\mathcal{L}_j = \mathcal{L}_0$
    结束 if
    根据策略 $\pi_{\theta_i^x}(\cdot|s_j^x)$选择下一个动作 u_j^x
    设置 $s_j^y = (u_j^x, \mathcal{L}_j)$
    根据策略 $\pi_{\theta_i^y}(\cdot|s_j^y)$ 选择下一个动作 u_j^y
    通过波形的前 $k+1$ 个分段演化量子比特
    对量子比特 N_{shot} 进行测量$\langle I_{j+1}\rangle, \langle Q_{j+1}\rangle$次
    更新 $k \to k+1$
    设置 $s_{j+1}^x = (\langle I_{j+1}\rangle, \langle Q_{j+1}\rangle, k)$
    计算奖励 r^x j
    估算泄漏布居 $\mathcal{L}_{j+1}$
    计算奖励 r_j^y
    将量子比特重置为 $|0\rangle$
    结束循环 $\left(s_0^x, u_0^x, r_0^x, s_1^x, u_1^x, r_1^x, \ldots, r_{N_{\text{ep}}-1}^x\right)$ 发送至
    第一个网络
    更新神经网络参数 $\theta_i^x \to \theta_{i+1}^x$
    将轨迹 $\left(s_0^y, u_0^y, r_0^y, s_1^y, u_1^y, r_1^y, \ldots, r_{N_{\text{ep}}-1}^y\right)$ 发送至
    第二个网络
    更新神经网络参数 $\theta_i^y \to \theta_{i+1}^y$
结束循环

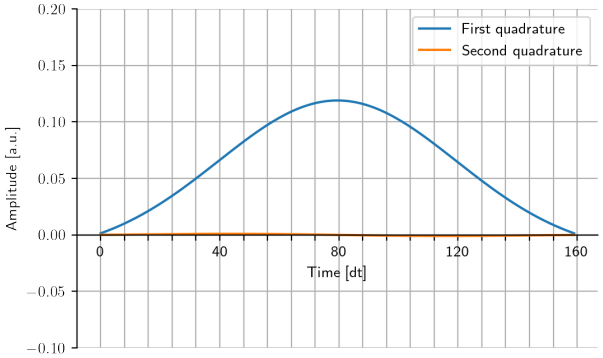
正交分量。也就是说，$c^x(t) = \Omega_G(t)$ 以及 $c^y(t) = \gamma\Omega_G'(t)$ 对于高斯包络 Ω_G 和系数 $\gamma \in \mathbb{R}$ [2]。DRAG 脉冲旨在基于一个简单的三能级模型降低泄漏。在实验中，对 Ω_G 的振幅和 γ 因子进行校准，以应对噪声随时间发生的变化。我们的初始策略利用校准的DRAG脉冲捕获有关系统动力学的信息；然而，我们的算法不应被视为对其他研究中提出的DRAG脉冲进行优化 [2], [52]。下一节将介绍实验结果，表明我们的智能体学会了新颖的脉冲形状，在保真度和泄漏率方面能够胜过 DRAG 门。有关预训练的更多细节，请参阅附录 B。

V. 实验结果

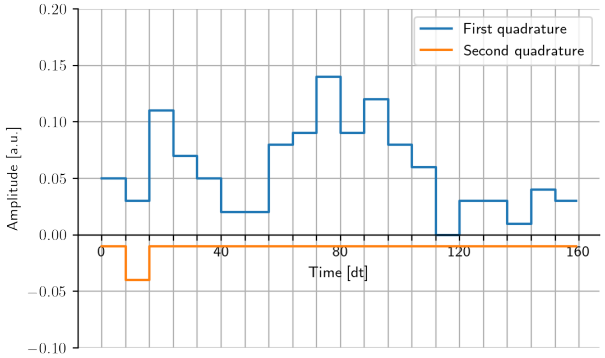
我们在 IBM Lima 量子计算机上使用 Qiskit Pulse 库设计 X 和 $\sqrt{X}$门，以进行概念验证 [53], [54]。为了评估量子门 U的效果，我们测量泄漏率 $\mathcal{L}$，其定义见（6），以及态保真度

$$\mathcal{F} = |\langle \psi_T| U |0\rangle|^2 \tag{10}$$

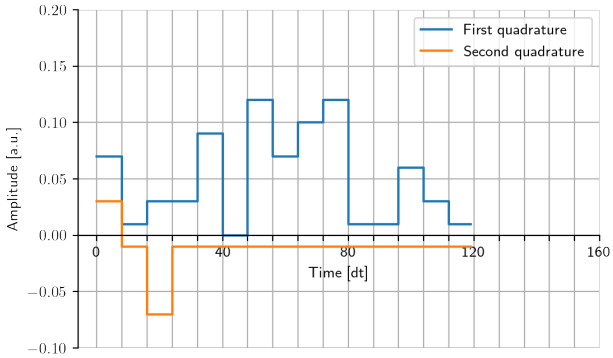
其中，$|\psi_T\rangle$ 是目标态。我们将快速量子门与 DRAG 门进行基准比较；DRAG 门是许多量子计算机所采用的行业标准，包括由



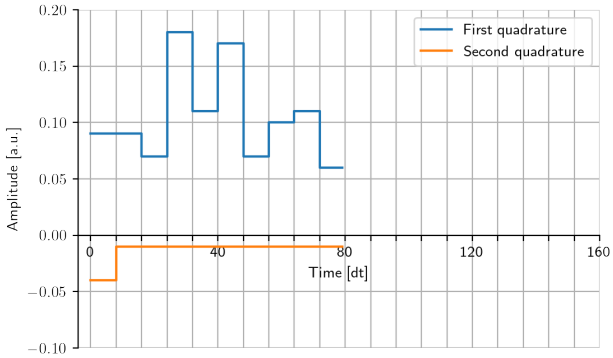
(a) DRAG X 门。



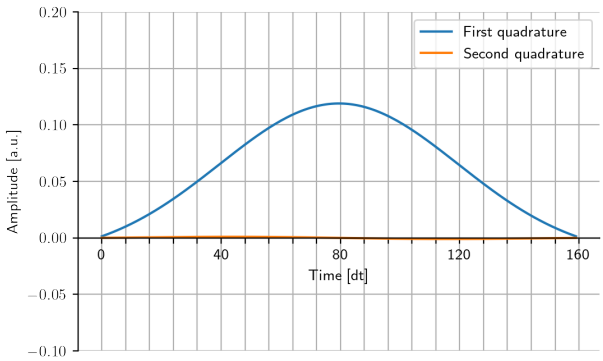
(b) 优化后的 X 门。



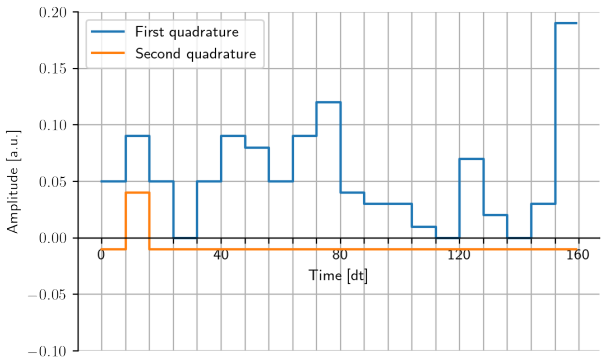
(c) 优化后的 X 门，速度快 1.33× 倍。



(d) 优化后的 X 门，速度快 2× 倍。



(e) DRAG X 门在 30 天后校准。

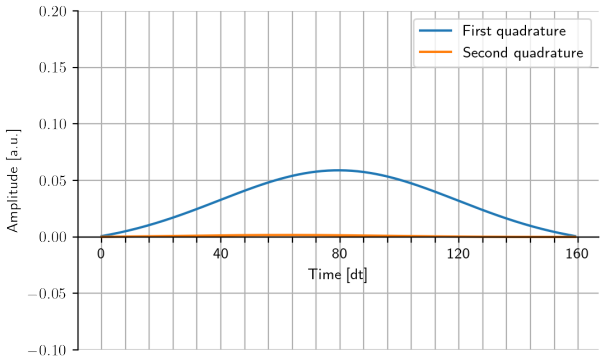


(f) 优化后的 X 门在 30 天后校准。

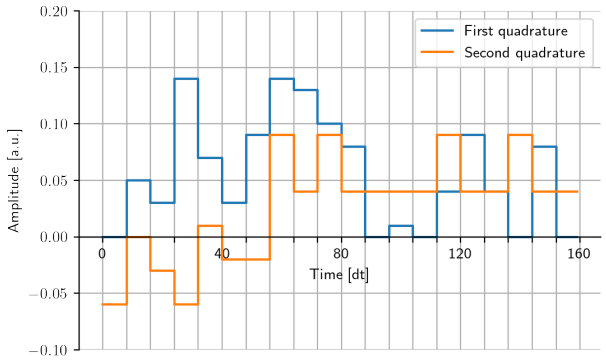
Gate	**Duration [ns]**	**Fidelity [%]** (± 0.6)	**Leakage [%]** (± 0.3)
DRAG X	35.6	91.7	4.2
Optimized X	35.6	92.2	3.5
Optimized X 1.33× faster	26.7	92.2	4.3
Optimized X 2× faster	17.8	90.4	3.8
DRAG X calibrated after 30 days	35.6	88.9	5.3
Optimized X calibrated after 30 days	35.6	89.2	4.3

（g）保真度和泄漏率汇总。

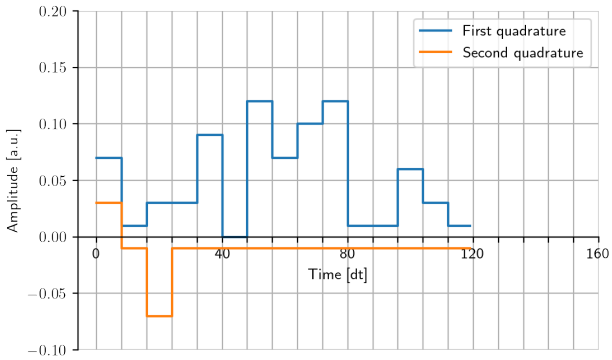
图 2：展示 IBM Lima 量子计算机上单量子比特门的默认控制脉冲和优化控制脉冲，以及各自对应的态保真度和泄漏率。 x 轴上的时间单位以 $dt = 0.222222$ ns 为单位，这是一个依赖于设备的参数，用于指定波形发生器的最大采样率。图 (a) 为一个 X 门的 DRAG 脉冲。图 (b) 为一个由我们的深度强化学习算法学习得到的 X 门的脉冲，持续时间为 35.6 ns。图 (c) 为一个由我们的深度强化学习算法学习得到的 X 门的脉冲，持续时间为 26.7 ns。图 (d) 为一个由我们的深度强化学习算法学习得到的 X 门的脉冲，持续时间为 17.8 ns。图 (e) 为训练 30 天后，由我们的深度强化学习算法校准的一个 X 门的脉冲。图 (f) 为训练 30 天后校准的一个 X 门的 DRAG 脉冲。图 (g) 为汇总这些量子比特门所达到的态保真度（10）和泄漏率（6）的表格。



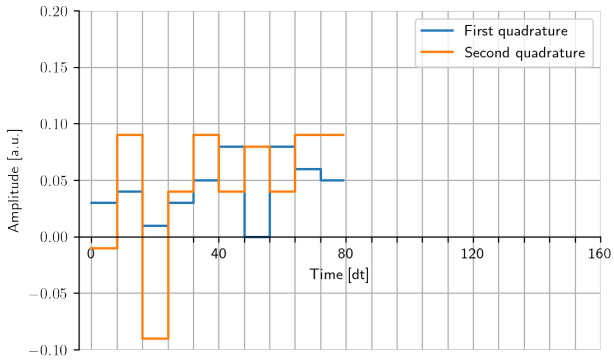
(a) DRAG $\sqrt{X}$ 门。



(b) 优化后的 $\sqrt{X}$ 门。



(c) 优化后的 $\sqrt{X}$ 门 1.33× 更快。



(d) 优化后的 $\sqrt{X}$ 门 2× 更快。

Gate	Duration [ns]	Fidelity [%] (± 0.6)	Leakage [%] (± 0.3)
DRAG $\sqrt{X}$	35.6	96.8	2.5
Optimized $\sqrt{X}$	35.6	97.5	2.4
Optimized $\sqrt{X}$ 1.33× faster	26.7	96.8	2.9
Optimized $\sqrt{X}$ 2× faster	17.8	96.5	3.5

(e) 保真度和泄漏率汇总。

图 3：IBM Lima 量子计算机上单量子比特门的默认和优化控制脉冲图，以及相应的态保真度和泄漏率。 x轴上的时间单位为 $dt = 0.222222$ ns，这是一个依赖于设备的参数，用于指定波形发生器的最大采样率。图 (a) 为一个 $\sqrt{X}$ 门的 DRAG 脉冲。图 (b) 为由我们的深度强化学习算法学习得到的一个 $\sqrt{X}$ 门脉冲，持续时间为 35.6 ns。图 (c) 为由我们的深度强化学习算法学习得到的一个 $\sqrt{X}$ 门脉冲，持续时间为 26.7 ns。图 (d) 为由我们的深度强化学习算法学习得到的一个 $\sqrt{X}$ 门脉冲，持续时间为 17.8 ns。图 (e) 汇总了这些门实现的态保真度（10）和泄漏率（6）。

由 IBM 提供。训练后的智能体可以创建分段数从 1 到 N_{seg} 的任意长度的门。我们测试了长度为 20 个分段（$t_g \approx 35.6$ ns）的门（与 DRAG 门长度相同）、15 个分段（$t_g \approx 26.7$ ns）的门，以及 10 个分段（$t_g \approx 17.8$ ns）的门。仅经过 200 次训练迭代后，较快的门便开始达到默认门的性能，其保真度和泄漏率相同或略差。我们优化的全长度门开始实现比校准的DRAG脉冲更高的态保真度和更低的泄漏率。优化后的 X 和 $\sqrt{X}$ 门分别如图2 和 3所示分别与相应的 DRAG 脉冲一同展示。各个门的保真度和泄漏率汇总于图 2g 和 3e。

我们还测试了智能体随时间变化的稳健性。图2f 展示了一个在智能体完成训练 30 天后创建的新优化 X 门。与校准的DRAG脉冲相比，它同样具有略高的保真度和更低的泄漏。这表明，30 天后我们的智能体无需任何额外训练。在初始训练期之后，我们的算法可用于高效校准超导量子比特上的门。

由于硬件访问受限，我们不会将智能体的训练迭代次数增加到 200 次以上，也不会对学习率、神经网络结构、分段数量或其他超参数进行优化。在最初的 200 次训练迭代后，我们的算法尚未完全收敛。我们预计，在优化算法超参数（包括增加训练迭代次数）后，能够取得更显著的优势。

VI. 推广到双量子比特门

对状态空间和奖励函数稍作修改后，我们的算法便可推广至双量子比特门的设计。在超导传输子量子比特中，双量子比特相互作用通过交叉共振（CR）脉冲产生（即以目标量子比特的共振频率驱动控制量子比特）[55]。该 CR 脉冲对应于门 $Z\otimes X$，即在目标量子比特上进行 X 旋转，其方向取决于控制量子比特的状态。行业标准 CR 门采用带高斯上升沿的圆角方波脉冲，并在第二个正交分量上进行导数脉冲校正 [56]。为避免杂散串扰，会同时向目标量子比特施加抵消音调。Baum等人使用强化学习设计了一种改进的 CR 门，无需额外的这一脉冲 [39]。我们采用相同的策略，使智能体能够利用动作空间$\mathbb{U}^x$ 和$\mathbb{U}^y$，学习改进的脉冲形状 $\tilde{s}_j^x = (\langle I_j^1\rangle, \langle Q_j^1\rangle, \langle I_j^2\rangle, \langle Q_j^2\rangle, k)$ 来自单量子比特情形的状态包含每个量子比特的平均读出信号。我们使用奖励

$$\tilde{r}_j^x(\tilde{s}_{j+1}^x, c_T^1, c_T^2) = \frac{1}{2}\left(r_j^x(\langle I_{j+1}^1\rangle, \langle Q_{j+1}^1\rangle, k, c_T^1) + r_j^x(\langle I_{j+1}^2\rangle, \langle Q_{j+1}^2\rangle, k, c_T^2)\right) \tag{11}$$

其中 r_j^x 定义见（8），而 c_T^1，c_T^2 分别是第一个和第二个量子比特的预期测量位置。对于第二个控制正交分量，将两个量子比特的泄漏布居相加，从而得到与单量子比特情形相同的状态和奖励。该 CR门构成一个通用门集 $\left\{CR,\ R_Z(\theta),\ X,\ \sqrt{X}\right\}$ 在与任意虚拟 Z旋转 [57] 以及我们在概念验证中展示的单量子比特门结合使用时。

VII. 结论

我们创建了一种深度强化学习算法，利用含噪硬件的读出信号进行实时反馈，以快速设计量子门。与需要精确模型的其他量子门合成策略不同，强化学习是一种无模型方法，因为这些精确模型无法捕捉量子比特的随机动力学。我们的双智能体架构使我们能够同时致力于减少泄漏和设计更快的量子门，从而降低退相干。我们提出的算法直接利用探测量子比特所获得的读出信号进行训练，从而减小测量分类误差的影响。由于测量开销低，我们的算法可以在硬件上训练，而非通过数值模拟进行训练。

我们在 IBM 的硬件上，使用时长不同的 X 和 $\sqrt{X}$ 门进行了概念验证。我们的智能体提出了新颖的控制脉冲，其速度是行业标准 DRAG 门的两倍，并且在仅经过 200 次训练迭代后，在保真度和泄漏方面达到了相当的性能。我们的智能体还创建了与 DRAG 门时长相同的门，在保真度和泄漏方面实现了小幅改进。我们还展示了训练好的智能体能够长期保持稳健。

到目前为止，我们获取硬件的机会有限，并且仅使用一组特定的超参数运行了该算法。众所周知，通过对超参数进行微调，强化学习（RL）的性能能够得到显著提升 [58]。我们预计，优化算法的超参数将带来更为显著的优势。

我们的概念验证是在 transmon量子比特上完成的；然而，我们提出的深度强化学习算法具有通用性，只需调整状态空间和动作空间（例如，针对囚禁离子量子计算 [49]，可使用光子计数和激光脉冲），即可应用于其他量子硬件。我们的深度强化学习算法为快速量子门设计所生成的改进型量子门操作，为近中期及未来量子器件的更广泛应用开辟了道路，同时降低了退相干和泄漏，而这些问题无法通过大多数纠错算法加以修复。

附录 A transmon硬件

我们将深度强化学习算法应用于快速量子门设计，目标是一个transmon量子比特，该量子比特由超导电路量子化能谱中最低的两个能级构成 [59]。transmon量子比特通过电容耦合驱动线进行控制。量子电路分析得出transmon哈密顿量 [60]

$$\mathcal{H} = \omega_q a^\dagger a + \frac{\alpha}{2} a^\dagger a^\dagger a a - ic(t)(a^\dagger - a) \tag{12}$$

其中 a、$a^\dagger$ 为产生算符和湮灭算符，ω_q 为量子比特频率，α 为非简谐性，且 $c(t)$ 为（5）中引入的电压包络。

transmon量子比特通过色散耦合连接到用于测量的超导谐振器。使用微波信号探测量子比特，该信号从谐振器散射并得到放大。最终的观测结果由数字化信号的同相 I 和正交分量 Q 的时间轨迹组成。量子比特的状态可以根据信号在 $(I,\ Q)$平面中的位置推断出来。在含噪硬件上，各位置通常会在 $(I,\ Q)$平面中重叠，因此需要使用决策器（例如线性判别分析器）[61]来预测状态（见图 1）。

附录 B 预训练

在本节中，我们介绍智能体的预训练。首先，我们将一个校准的DRAG脉冲离散化为一个 N_{seg} 分段常值脉冲，其中每个分段的振幅分别取U^x 和U^y 中距离最近的动作，分别对应第一和第二正交分量。对于每个 k，我们测量波形前 $k+1$ 个分段之后的平均值（$\langle I\rangle$，$\langle Q\rangle$）信号和泄漏 $\mathcal{L}$。每个智能体根据神经网络在给定输入状态时输出的策略与以期望振幅为中心、定义在相应动作空间上的高斯分布之间的均方误差（MSE）获得奖励

预训练在完整脉冲上进行若干次迭代。由于奖励和动作不依赖于系统状态，因此可以在更新网络参数之前运行所有电路，预训练能够高效执行。预训练结束时，智能体偏好DRAG脉冲。尽管智能体随后会学习新的脉冲形状，但探索是从解空间中的近最优区域开始的。

参考文献

[1] J. Preskill，《NISQ时代及未来的量子计算》，Quantum，第2卷，第79页，2018年8月。 [在线]。获取地址：https: //doi.org/10 22331 2018 08 06 79 . /q- - - - F. Motzoi、J. M. Gambetta、P. Rebentrost 和 F. K. Wilhelm，《用于消除弱非线性量子比特中泄漏的简单脉冲》，《物理评论快报》，第103卷，第110501页，2009年9月。 [在线]。获取地址：https://link.aps.org/doi/10 1103 103 110501 . /PhysRevLett. . M. McEwen、D. Kafri 等，《消除超导量子纠错中由泄漏引起的相关错误》，《自然通讯》，第12卷，第1期，第1761页，2021年。 [在线]。获取地址：https://www.nature.com/articles/s41467-021-21982-y [4] C. J. Wood 和 J. M. Gambetta，《泄漏错误的量化与表征》，《物理评论A》，第97卷，第032306页，2018年3月。 [在线]。获取地址：https://link.aps.org/doi/10.1103/PhysRevA.97.032306[5] M. Suchara、A. W. Cross 和 J. M. Gambetta，《环面码中的泄漏抑制》，arXiv，2014年。 [在线]。获取地址：https://arxiv.org/ abs/1410.8562[6] A. G. Fowler，《拓扑码中应对量子比特泄漏》，《物理评论A》，第88卷，第042308页，2013年10月。 [在线]。获取地址：https://link.aps.org/doi/10 1103 88 042308 . /PhysRevA. . N. Khaneja、R. Brockett 和 S. J. Glaser，《自旋系统中的时间最优控制》，《物理评论A》，第63卷，第032308页，2001年2月。 [在线]。获取地址：https://link.aps.org/doi/10 1103 63 032308 . /PhysRevA. . U. Boscain、G. Charlot、J.-P. Gauthier、S. Guerin 和 H.-R. Jauslin，《激光诱导的二能级或三能级量子系统布居转移中的最优控制》，《数学物理学杂志》，第43卷，第5期，第2107–2132页，2002年5月。 [在线]。获取地址：https://hal.science/hal-00983019 。[9] A. Pechen、N. Il'in、F. Shuang 和 H. Rabitz，《通过冯·诺依曼测量进行量子控制》，《物理评论A》，第74卷，第052102页，2006年11月。 [在线]。获取地址：https://link.aps.org/doi/10.1103/PhysRevA.74.052102[10] S. Montangero、T. Calarco 和 R. Fazio，《约瑟夫森电荷量子比特的鲁棒最优量子门》，《物理评论快报》，第99卷，第17期，2007年10月。 [在线]。获取地址：https://doi.org/10.1103%2Fphysrevlett. 99.170501[11] V. F. Krotov，《最优控制理论的全局方法》。美国马萨诸塞州波士顿：Birkh¨auser Boston，1993年，第74–121页。[12] N. B. Dehaghani 和 F. L. Pereira，《基于庞特里亚金极大值原理的高保真量子态传输》，2022年。 [在线]。获取地址：https://arxiv.org/abs/2203 04361 . W. Zhu 和 H. Rabitz，《一种针对正定算符期望值的量子最优控制快速单调收敛迭代算法》，《化学物理学杂志》，第109卷，第2期，第385–391页，1998年。 [在线]。获取地址：https: //pubs.aip.org/aip/jcp/article/109 2 385 476865 / / / D. D'Alessandro 和 M. Dahleh，《二能级量子系统的最优控制》，IEEE《自动控制汇刊》，第46卷，第6期，第866–876页，2001年。 [在线]。获取地址：https://ieeexplore.ieee.org/document/928587 。[15] N. Khaneja、T. Reiss、C. Kehlet、T. Schulte-Herbr¨uggen 和 S. J. Glaser，《耦合自旋动力学的最优控制：通过梯度上升算法设计核磁共振脉冲序列》，《磁共振杂志》，第172卷，第2期，第296–305页，2005年。[在线]。获取地址：https://www.sciencedirect.com/science/article/abs/pii/S1090780704003696

[16] S. Machnes、E. Ass´emat、D. Tannor 和 F. K. Wilhelm，"实用量子比特控制脉冲的可调谐、灵活且高效优化"，《物理评论快报》，第120卷，第150401页，2018年4月。 [在线]。可访问：https://link.aps.org/doi/10 1103 120 150401 。/PhysRevLett。[17] P. Doria、T. Calarco 和 S. Montangero，"多体量子动力学的最优控制技术"，《物理评论快报》，第106卷，第19期，2011年5月。 [在线]。可访问：https://doi.org/10.1103% 2 106 190501 Fphysrevlett。J. N. Tsitsiklis，"异步随机逼近与Q学习"，Machine Learning，第16， 3， 185 202, 卷，第期，第页–。1994 [ ] 10 1023 在线。可访问：https://link.springer.com/article/。/A: 1022689125041[19] C. Watkins 和 P. Dayan，"Q学习"，Machine Learning，第8卷，第3期，第279–292页，1992年7月。 [在线]。可访问：https://link.springer。com/article/10.1007/BF00992698[20] C. Szepesvari 和 M. L. Littman，"基于价值函数的强化学习算法的统一分析"，《神经计算》，第11卷，第2017–2060页，1999年。 [在线]。可访问：https://pubmed.ncbi.nlm.nih。gov/10578043/ [21] F. Metz 和 M. Bukov，"使用张量网络和强化学习进行自校正量子多体控制"，2022年。 [在线]。可访问：https://arxiv.org/abs/2201 11790 。Y. Qiu、M. Zhuang、J. Huang 和 C. Lee，"用于量子计量学的深度强化学习高效且鲁棒的纠缠生成"，《新物理学杂志》，第24卷，第8期，第083011页，2022年8月。 [ ] 10 1088 2 1367 2630 2 8285 在线。可访问：https://doi.org/。% F - % Fac P. Peng、X. Huang、C. Yin、L. Joseph、C. Ramanathan 和 P. Cappellaro，"深度强化学习用于量子哈密顿量工程"，2021 [ ] 2102 13161 。在线。可访问：https://arxiv.org/abs/。Z. An、H.-J. Song、Q.-K. He 和 D. L. Zhou，"使用强化学习对多能级耗散量子系统进行量子最优控制"，《物理评论A》，第103卷，第012404页，2021年1月。 [在线]。可访问：https://link.aps.org/doi/10 1103 103 012404 。/PhysRevA。S. Daraeizadeh、S. P. Premaratne 和 A. Y. Matsuura，"通过深度强化学习为半导体量子点设计高保真多量子比特门"，载于2020年IEEE量子计算与工程国际会议（QCE）。IEEE，2020年10月。 [ ] 10 1109 2 49297 2020 00014 在线。可访问：https://doi.org/。% Fqce。。H. Ma、D. Dong、S. X. Ding 和 C. Chen，"基于课程的深度强化学习用于量子控制"，2020年。 [在线]。可访问：https://arxiv.org/abs/2012 15427 。J.-J. Chen 和 M. Xue，"通过深度强化学习智能体操纵自旋动力学"，2019年。 [在线]。可访问：https://arxiv.org/abs/1901 08748 。R. Porotti、D. Tamascelli、M. Restelli 和 E. Prati，"通过深度强化学习实现量子态的相干输运"，《通讯物理》，第2卷，第1期，2019年6月。 [在线]。可访问：https: //doi.org/10.1038%2Fs42005-019-0169-x [29] X.-M. Zhang、Z. Wei、R. Asad、X.-C. Yang 和 X. Wang，"强化学习何时在量子控制中脱颖而出？关于态制备的比较研究"，NPJ Quantum Information，第5卷，2019年10月。 [在线]。可访问：https://www.nature.com/articles/s41534-019-0201-8#citeas[30]C. Chen、D. Dong、H.-X. Li、J. Chu 和 T.-J. Tarn，"基于保真度的概率Q学习用于量子系统控制"，《IEEE神经网络与学习系统汇刊》，第25卷，第5期，第920–933页，2014年5月。 [在线]。可访问：https://doi.org/10 1109 2 2013 2283574 。% Ftnnls。。O. Shindi、Q. Yu、P. Girdhar 和 D. Dong，"使用深度强化学习进行无模型量子门设计与校准"，arXiv，2023年。 [ ] 2302 02371 在线。可访问：https://arxiv.org/abs/。V. V. Sivak、A. Eickbusch、H. Liu、B. Royer、I. Tsioutsios 和 M. H. Devoret，"使用强化学习进行无模型量子控制"，Physical Review X，第12卷，第011059页，2022年3月。 [在线]。可访问：https://link.aps.org/doi/10 1103 12 011059 。/PhysRevX。。S.-F. Guo、F. Chen 等，"强化学习辅助量子相变中的更快态制备"，《物理评论快报》，第126卷，第6期，2021年2月。 [在线]。可访问：https: //doi.org/10 1103 2 126 060401 。% Fphysrevlett。。E.-J. Kuo、Y.-L. L. Fang 和 S. Y.-C. Chen，"通过深度强化学习进行量子架构搜索"，arXiv，2021年。 [在线]。可访问：https://arxiv.org/abs/2104 07715 。M. M. Wauters、E. Panizon、G. B. Mbeng 和 G. E. Santoro，"强化学习辅助的量子优化"，Physical

Review Res.，第2卷，第033446页，2020年9月。 [在线]。可访问：https://link.aps.org/doi/ 10 1103 2 033446 。/PhysRevResearch。。[36] A. Garcıa-Saez 和 J. Riu，“通过强化学习对量子近似优化算法进行连续控制的量子可观测量”，arXiv，第abs/1911.09682卷，2019年。 [在线]。可访问：https://arxiv.org/abs/ 1911 09682 。[37] M. Bukov、A. G. R. Day、D. Sels、P. Weinberg、A. Polkovnikov 和 P. Mehta，“量子控制不同阶段中的强化学习”，Physical Review X，第8卷，第031086页，2018年9月。 [在线]。可访问：https://link.aps.org/doi/ 10 1103 8 031086 。/PhysRevX。。[38] K. Reuer、J. Landgraf 等，“实现一个发现量子系统实时反馈控制策略的深度强化学习智能体”， 2022 [ ] 2210 16715 。在线。可访问：https://arxiv.org/abs/。[39] Y. Baum、M. Amico 等，“在超导量子计算机上进行抗误差门集设计的实验性深度强化学习”，PRX Quantum，第2卷，第4期，2021年11月。 [在线]。可访问：https://doi.org/ 10 1103 2 2 040324 。% Fprxquantum。。[40] S. Borah、B. Sarma、M. Kewming、G. J. Milburn 和 J. Twamley，“使用深度强化学习对双阱非线性势进行基于测量的反馈量子控制”，《物理评论快报》，第127卷，第190403页，2021年11月。 [在线]。可访问：https://link.aps.org/doi/ 10 1103 127 190403 。/PhysRevLett。。[41] M. Bukov，“用于自主制备Floquet工程态的强化学习：反演量子卡皮察振子”，Physical Review B，第98卷，第22期，2018年12月。 [在线]。可访问：https://doi.org/ 10 1103 2 98 224305 。% Fphysrevb。。[42] R. Gray 和 D. Neuhoff，“量化”，《IEEE信息论汇刊》，第44卷，第6期，第2325–2383页，1998年。[43] A. Gadjiev 和 A. Ghorbanalizadeh，“通过k-正线性算子序列逼近解析函数”，Journal of Approximation Theory，第162卷，第6期，第1245–1255页，2010年。[在线]。可访问：https://www.sciencedirect.com/science/article/pii/ S0021904510000249[44] D. Bertsekas 和 J. N. Tsitsiklis，《神经动态规划》。Athena Scientific，1996年。[45] Y. Li，“深度强化学习：综述”，arXiv，2018年。 [ ] 1701 07274 在线。可访问：https://arxiv.org/abs/。[46] R. J. Williams，“用于连接主义强化学习的简单统计梯度跟踪算法”，Machine Learning，第8卷，第3期，第229–256页，1992年。 [在线]。可访问：https://link.springer.com/article/ 10.1007/BF00992696[47] Z. An 和 D. L. Zhou，“用于量子门控制的深度强化学习”，EPL（Europhysics Letters），第126卷，第6期，第60002页，2019年7月。 [在线]。可访问：https://doi.org/10.1209%2F0295-5075%2F126%2F60002[48] M. Y. Niu、S. Boixo、V. N. Smelyanskiy 和 H. Neven，“通过深度强化学习实现通用量子控制”，NPJ Quantum Information，第5卷，第33页，2019年4月。 [在线]。可访问：https: //www.nature.com/articles/s41534-019-0141-3#Bib1[49] C. D. Bruzewicz、J. Chiaverini、R. McConnell 和 J. M. Sage，“囚禁离子量子计算：进展与挑战”，Applied Physics Reviews，第6卷，第2期，第021314页，2019年。 [在线]。可访问：https://doi.org/ 10 1063 2 1 5088164 。% F。L. M. K. Vandersypen 和 M. A. Eriksson，“使用半导体自旋进行量子计算”，Physics Today，第72卷，第8期，第38–45页，2019年。 [ ] 10 1063 3 4270 在线。可访问：https://doi.org/。/PT。。[51] L. Henriet、L. Beguin 等，“使用中性原子进行量子计算”，Quantum，第4卷，第327页，2020年。 [在线]。可访问：https://doi.org/ 10 22331 2 2020 09 21 327 。% Fq- - - - M. Werninghaus、D. J. Egger、F. Roy、S. Machnes、F. K. Wilhelm 和 S. Filipp，“通过最优控制抑制快速超导量子比特门中的泄漏”，NPJ Quantum Information，第7卷，第1期，2021年1月。 [ ] 10 1038 2 41534 020 00346 2 在线。可访问：https://doi.org/。% Fs - - - Qiskit贡献者，“Qiskit：用于量子计算的开源框架”，2023年。 [在线]。可访问：https://qiskit.org/ [54] D. C. McKay、T. Alexander 等，“用于OpenQASM和OpenPulse实验的Qiskit后端规范”，arXiv，2018年。 [在线]。可访问：https://arxiv.org/abs/ 1809 03452 。G. S. Paraoanu，“超导量子比特的微波诱导耦合”，Physical Review B，第74卷，第140504页，2006年10月。 [在线]。可访问：https://link.aps.org/doi/ 10 1103 74 140504 。/PhysRevB。。[56] S. Sheldon、E. Magesan、J. M. Chow 和 J. M. Gambetta，“在交叉共振门中系统调谐串扰的过程”，《物理评论A》，第93卷，页码060302，2016年6月。 [在线]。获取地址：https://link.aps.org/doi/ 10 1103 93 060302 。/PhysRevA.。[57] D. C. McKay、C. J. Wood、S. Sheldon、J. M. Chow 和 J. M. Gambetta，《用于量子计算的高效 z 门》，《物理评论A》，第96卷，页码022330，2017年8月。 [在线]。获取地址：https://link.aps.org/doi/ 10 1103 96 022330 。/PhysRevA.。[58] M. Feurer 和 F. Hutter，《超参数优化》。Cham：施普林格国际出版公司，2019年，第3–33页。[59] J. Koch、T. M. Yu 等人，《源自库珀对盒的电荷不敏感量子比特设计》，《物理评论A》，第76卷，第4期，2007年10月。 [ ] 10 1103 2 76 042319 在线。获取地址：https://doi.org/。% Fphysreva.。[60] P. Krantz、M. Kjaergaard、F. Yan、T. P. Orlando、S. Gustavsson 和 W. D. Oliver，《超导量子比特指南：量子工程师必读》，Applied Physics Reviews，第6卷，第2期，页码021318，2019年6月。 [在线]。获取地址：https://doi.org/ 10 1063 2 1 5089550 。% F。E. Magesan、J. M. Gambetta、A. D. Corcoles 和 J. M. Chow，《用于区分量子测量轨迹并改进读出的机器学习》，《物理评论快报》，第114卷，页码200501，2015年5月。 [在线]。获取地址：https://link.aps.org/doi/10. 1103/PhysRevLett.114.200501 [61]