

# Reinforcement Learning for Efficient Resource Allocation Between Bayesian Estimation and Operation in Quantum Computer Limited by Low-Frequency Noise

Jan A. Krzywda\* and Evert van Nieuwenburg

*(aQa<sup>L</sup>) Applied Quantum Algorithms, Lorentz Institute and Leiden Institute of Advanced Computer Science,  
Leiden University, P.O. Box 9506, 2300 RA Leiden, The Netherlands*

## I. INTRODUCTION

**Intro** Coupling between the qubit and its environment often leads to the fluctuations of qubit Hamiltonian parameters, which deteriorate its control and cause decoherence mechanisms. While fast fluctuations causes qubit relaxation (related to  $T_1$  time), and uncorrelated dephasing ( $T_2$  time), many solid-state devices are limited by the slow fluctuations of the Hamiltonian parameters.

**[Mitigation techniques]** Treating such temporarily correlated noise with standard Quantum Error Correction [1] and Quantum Error Mitigation techniques [2] is not straightforward, and often requires additional decohering techniques, i.e. dynamical decoupling [3], or Pauli Twirling [4]. However, these methods are not always efficient, and can be costly in terms of resources. Alternative approach to mitigate effects of slowly fluctuating parameters, is to track and correct them in real-time.

Tracking of Hamiltonian parameters can be realised by directly measuring the drift [5] or using Bayesian approach [6]. In the latter, the knowledge about them is updated from shot data of an experiment with known probability distribution [7], which has proven to improve performance of spin qubits affected by brown [8] and pink (1/f) noise [9]. Although for constant and unknown parameters, the optimal estimation methods of Hamiltonian learning has been developed [10] and implemented [11], for dynamical case optimal strategy depends on the interplay between noise dynamics and data acquisition rate. For instance there exists a natural tension between increasing precision by more estimation shots and decreasing accuracy due to changes in estimated parameter during estimation period.

On the other hand investing significant part of shot-budget in estimation, is naturally limiting resources for running any meaningful quantum algorithm. Therefore, the optimal strategy for the qubit control is to allocate resources between the estimation and the qubit operation, in order to maximize the number of clean runs of the algorithm. This setup naturally resamples exploration (estimation) and exploitation (qubit operation) trade-off, which is a typical problem in Reinforcement Learning (RL) [12].

In this work we design the algorithms for efficient resource allocation between the Bayesian estimation and the qubit operation, in order to maximize the qubit performance. As a proof-of-principle algorithm, that could

be also used for the qubit calibration, we concentrate here on the phase-flip gate, i.e. dynamically adjust waiting time to realise a  $\pi$ -rotation around the rotation axis affected by low-frequency noise. We develop two algorithms, that are on the opposite spectrum of complexity. In the first one, we map the problem of flipping phase into a multi-arm bandit, where arms are the waiting time. The agent is responsible for choosing the arm, that maximizes the successful flips in a finite time. In the latter approach we set up full Reinforcement Learning, in which the agent has access to the results of Bayesian Estimation and can choose the estimation strategy, as well as dynamically allocate resources between the estimation and the qubit operation.

## II. MODEL

---

\* j.a.krzywda@liacs.leidenuniv.nl