Batch Entanglement Detection in Parameterized Qubit States using Classical Bandit Algorithms

Anonymous authors
Paper under double-blind review

Abstract

Entanglement is a key property of quantum states that acts as a resource for a wide range of tasks in quantum computing. Entanglement detection is a key conceptual and practical challenge. Without adaptive or joint measurements, entanglement detection is constrained by no-go theorems (Lu et al., 2016), necessitating full state tomography. Batch entanglement detection refers to the problem of identifying all entangled states from amongst a set of K unknown states which finds applications in quantum information processing. We devise a method for performing batch entanglement detection by measuring a single-parameter family of entanglement witnesses, as proposed by Zhu et al. (2010), followed by a thresholding bandit algorithm on the measurement data. The proposed method can perform batch entanglement detection conclusively when the unknown states are drawn from a practically well-motivated class of two-qubit states $\mathcal{F}$, which includes Depolarised Bell states, Bell diagonal states, etc. Our key novelty lies in drawing a connection between batch entanglement detection and a Thresholding Bandit problem in classical Multi-Armed Bandits (MAB). The connection to the MAB problem also enables us to derive theoretical guarantees on the measurement/sample complexity of the proposed technique. We demonstrate the performance of the proposed method through numerical simulations and an experimental implementation. More broadly, this paper highlights the potential for employing classical machine learning techniques for quantum entanglement detection.

1 Introduction

Quantum information theory has redefined quantum entanglement from a descriptive property of quantum states to a fundamental non-classical resource. As the basis for applications such as quantum communication, teleportation, and information processing (Bennett et al., 1993; Buhrman et al., 2001; Horodecki et al., 2009), entanglement detection and verification are central problems. Traditionally, this involves performing quantum measurements that yield probabilistic data, enabling techniques like full-state tomography (FST) for state reconstruction. However, this faces two challenges: theoretically, even after FST, determining entanglement remains computationally intractable, and even more so in scenarios involving many qubits; practically, real-world noise and imperfections limit the accuracy of state reconstruction. Modern multi-qubit compute systems may generate a bunch of entangled states across different sets of qubits through quantum gate operations; however, gate noise (e.g., phase flip errors, depolarization) can compromise their entanglement, requiring precise verification to ensure their reliability before use in applications such as quantum computation and communication (Hong et al., 2010). In such contexts, FST can be employed for entanglement detection. However, it comes with a high computational burden which may be unnecessary. We propose an alternative approach for simultaneous real-time verification or detection of entanglement among a given set of quantum states, dubbed *batch entanglement detection*. Instead of relying on FST, learning algorithms utilize statistical patterns to simultaneously analyze measurement data from a batch of quantum states and provide high probability guarantees on what they learn, i.e., whether or not the states are entangled.

Conventional techniques for learning quantum states include extensive research on FST (see Kueng et al. (2017); Wang et al. (2019); O'Donnell & Wright (2015a;b); Banaszek et al. (2013); Flammia et al. (2012)

and references therein and also see Guta et al. (2020); Torlai et al. (2018); Quek et al. (2018); Koutný et al. (2022); Schmale et al. (2022); França et al. (2021) for machine learning-based approaches). Measurements required for FST scale exponentially with the number of qubits. While entangled measurements enable near-optimal copy complexity for FST (O'Donnell & Wright, 2015b; Haah et al., 2017), practical implementations rely on single-copy measurements with reconstruction methods like linear inversion, maximum likelihood estimation, and maximum a posteriori estimation (Teo et al., 2011; Siddhu, 2019). These reconstructed states can be tested for entanglement using well-known criteria (some are outlined in Sec . 2.1). Alternatively, entanglement can be detected by measuring *entanglement witnesses* (Horodecki et al., 1996a; Terhal, 2000; Lewenstein et al., 2000a; Chruciski & Sarbicki, 2014), observables that detect *some* entangled states. No single witness can detect all entangled states, but in the worst case, combining information obtained from measuring different witnesses aids in state reconstruction via FST. This is explored in Zhu et al. (2010), where measurement operators from a family of six witnesses are used for bipartite qubit systems. The proposed approach for entanglement detection involves measuring a witness and formulating a *separability criterion* based on the frequencies of measurement outcomes. A negative value of the criterion indicates entanglement; otherwise, the process is repeated with another witness. If the state remains undetected by all witnesses, a tomographic reconstruction is performed (see Section 2.1 for further details).

Recently, the authors of Lumbreras et al. (2022) proposed using multi-armed bandit (MAB) frameworks for learning quantum states. The MAB algorithm repeatedly chooses from several options ("arms"), with the goal of finding the arm with the best outcome (the "best arm"). The algorithm balances between exploiting the known best options and exploring others to ensure no better option is missed (more details on MAB and policies can be found in Sec. 2.2). In Lumbreras et al. (2022), the inherent linearity in the quantum mechanical description of states is capitalized and a well-known classical learning algorithm that prescribes a sequential order of choosing measurements is employed. The MAB algorithms that are used provide guarantees on the quality of the estimate of the unknown quantum state. This MAB model in Lumbreras et al. (2022) does not directly apply to batch entanglement detection. Instead, it focuses on learning *one* entire quantum state, which may be unnecessary for entanglement detection.

Our first contribution builds on the witness-based separability criterion in Zhu et al. (2010) and using suitable MAB policies for learning the same. We establish a formal connection between batch entanglement detection and the thresholding bandit problem (TBP) (Kano et al., 2018), enabling accurate and quick identification of m entangled states from a batch of K candidate states through adaptive measurement allocation. This formulation, which we refer to as the (m, K)-quantum MAB framework, differs structurally from the setting in Lumbreras et al. (2022) (see Remark 1) and focuses on learning entanglement-specific metrics without requiring full state reconstruction. Our second contribution uses classical MAB policies for adaptive measurement allocation and provides explicit measurement/copy complexity guarantees for batch entanglement detection–guarantees absent in FST and repeated witness testing Zhu et al. (2010). Using statistically guided confidence bounds, these policies are sample-efficient since they prioritise measurement effort only on uncertain states. Finally, we demonstrate the complete MAB-based pipeline for batch entanglement detection across multiple IBM Quantum backends and validate the framework under realistic noise.

The rest of this paper is organized as follows: In Sec. 2, we provide a brief recap of some preliminary concepts in entanglement theory and multi-armed bandits. Readers interested in our connection between the two can move directly to Sec. 3, where we describe the (m, K)-quantum Multi-Armed Bandit framework for entanglement detection. We define a class of parameterized two-qubit states $\mathcal{F}$ and identify measurement operators that conclusively detect entanglement in $\mathcal{F}$, detailed in Sections 4.1, 4.2, and 4.3. In Section 5, we demonstrate two TBP policies for entanglement detection. Section 6 analyzes the MAB policy performance on IBMQ backends and on an ibm-brisbane device for a family of states in $\mathcal{F}$ and details the quantum circuits used for simulation. Section 7 highlights measurement scheme limitations for entanglement detection in arbitrary states through numeric examples. In Section 8, we contextualize the numerical performance gains and discuss the practical advantages of the proposed MAB approach in comparison to existing state-of-the-art methods for entanglement detection like FST and fixed-witness testing (Zhu et al., 2010). Finally, Section 9 conludes the paper. Detailed proofs for the results presented in the paper can be found in Appendix A.

2 Preliminaries

Let $\mathcal{H}$ be a finite dimensional Hilbert space with dimension d. A pure quantum state is represented by a unit norm vector $|\psi\rangle \in \mathcal{H}$. Let $\mathcal{L}(\mathcal{H})$ be the space of linear operators on $\mathcal{H}$, the Frobenius inner product for any $A, B \in \mathcal{L}(\mathcal{H})$, $\langle A, B\rangle \coloneqq \mathrm{Tr}\big(A^\dagger B\big)$ where $\dagger$ represents conjugate transpose. A Hermitian operator satisfies $H = H^\dagger$. A density operator $\rho \in \mathcal{L}(\mathcal{H})$ is Hermitian, positive semi-definite, $\rho \geq 0$, and has unit trace, $\mathrm{Tr}(\rho) = 1$; it can represent both pure and mixed states. A positive operator value measure (POVM) is collection of positive operators $\{E_i \geq 0\}$ that sum to the identity, $\sum_i E_i = I$. A POVM represents a measurement where E_i corresponds to measurement outcome i, but sometimes we compress this and just say E_i is a measurement outcome.

Let $\mathcal{H}_a$ and $\mathcal{H}_b$ be finite-dimensional Hilbert spaces with dimensions d_a and d_b, respectively, and $\mathcal{H}_{ab} \coloneqq \mathcal{H}_a \otimes \mathcal{H}_b$, where $\otimes$ represents tensor product, be a bipartite Hilbert space with dimension $d = d_a d_b$. A density operator $\rho_{ab} \in \mathcal{L}(\mathcal{H}_{ab})$ is called *separable* if it can be written as a convex combination of product states, that is,

$$\rho_{ab} = \sum_i p_i \left|\phi_a^i, \chi_b^i\right\rangle \left\langle \phi_a^i, \chi_b^i\right|, \tag{1}$$

where $p_i \geq 0$ such that $\sum_i p_i = 1$ and $\left|\phi_a^i, \chi_b^i\right\rangle \coloneqq |\phi\rangle_a^i \otimes |\chi\rangle_b^i$ is a product of two pure states. We denote the convex set of all separable states by S_{ab}. Conversely, ρ_{ab} is *entangled* if it can not be written in the form equation 1. We discuss some preliminaries on separability criteria for entanglement detection in Section 2.1 and background on stochastic multi-armed problems in Section 2.2.

2.1 Separability Criteria for Entanglement Detection

2.1.1 Standard Analytical Separability Tests

Using full state tomography (FST), one can reconstruct the bipartite qubit state ρ_{ab} and verify its entanglement through standard separability criteria (Horodecki et al., 2009). For bipartite qubit systems, the Peres-Horodecki criterion (also called the PPT criterion) (Horodecki et al., 1996b; Peres, 1996) establishes that a density operator ρ_{ab} is separable if and only if the eigenvalues of its partial transpose $\rho_{ab}^{\top_b}$ are non-negative. This condition remains necessary and sufficient for (2×3) systems but fails in higher dimensions due to the existence of bound-entangled PPT states. Complementary criteria include the range criterion (Horodecki, 1997), the matrix realignment criterion (Rudolph, 2000), the covariance matrix (CM) criterion (Gühne et al., 2007), and additional methods discussed in Gurvits (2003); Doherty et al. (2004).

2.1.2 Entanglement Witness-based Separability Criterion

Entanglement can be detected by measuring entanglement witnesses and can be defined as follows:

Definition 1 (Entanglement Witness) *An entanglement witness $W \in \mathcal{L}(\mathcal{H}_{ab})$ is a Hermitian operator satisfying,*

$$\langle \rho_{ent}, W\rangle = \mathrm{Tr}(\rho_{ent} W) < 0, \quad \textit{for some entangled } \rho_{\mathrm{ent}}, \tag{2}$$
$$\langle \rho, W\rangle = \mathrm{Tr}(\rho W) \geq 0, \ \forall \rho \in S_{ab}. \tag{3}$$

Geometrically, a witness W defines a hyperplane in the state space, delineating the set of detectable entangled states $D_W = \{\rho \text{ s.t. } \mathrm{Tr}(\rho W) < 0\}$ from all separable states. When comparing two arbitrary witnesses W_1 and W_2, if $D_{W_1} \subseteq D_{W_2}$, then W_2 is said to be *finer* than W_1. Further insights into this topology are detailed in Lewenstein et al. (2000b, Lemma 1). A witness is said to be *optimal* when no strictly finer one exists, implying that it lies tangent to the boundary of the convex set S_{ab} (Bengtsson & Zyczkowski, 2006). We briefly review a witness-based separability criterion from Zhu et al. (2010). The authors propose a single-parameter witness family,

$$\rho_w(\alpha) = \cos^2 \alpha I - \left(|\psi\rangle \langle\psi|\right)^{\top_b}, \tag{4}$$

where $|\psi\rangle = \cos\alpha\,|00\rangle + \sin\alpha\,|11\rangle$ such that $\alpha \in [0, \pi/4]$ and $\top_b$ is the partial transpose with respect to $\mathcal{H}_b$. We denote $\mathcal{E}$ to be the set of projectors onto the eigenstates of

$$\rho(\alpha)^{\top_b} = (|\psi\rangle\langle\psi|)^{\top_b} = \frac{1+\cos 2\alpha}{2}|00\rangle\langle 00| + \frac{1-\cos 2\alpha}{2}|11\rangle\langle 11| + \frac{\sin 2\alpha}{2}\big(|\Psi^+\rangle\langle\Psi^+| - |\Psi^-\rangle\langle\Psi^-|\big).$$

In other words, $\mathcal{E} = \{|00\rangle\langle 00|, |11\rangle\langle 11|, |\Psi^+\rangle\langle\Psi^+|, |\Psi^-\rangle\langle\Psi^-|\}$ forms a POVM, which is referred to as the Witness Basis Measurement (WBM). For the remainder of the paper, we assume that the exact projective forms of the WBM are fixed and known.

The witness expectation value serves as a detection statistic, that is, $\mathrm{Tr}(\rho W) < 0$ certifies entanglement, while non-negativity renders the test inconclusive. If the test is inconclusive for the base witness in equation 4, that is, $\mathrm{Tr}(\rho_w(\alpha)\rho) \geq 0$, then subsequent witnesses are generated via local unitary transformations U_1 and U_2 as,

$$\rho_w(\alpha) \longrightarrow (U_1 \otimes U_2)^\dagger \rho_w(\alpha)(U_1 \otimes U_2). \tag{5}$$

with $(U_1, U_2) \in \big\{(I,I),(I,X),(C^\dagger,C),(C^\dagger,XC),(C,C^\dagger),(C,XC^\dagger)\big\}$. Here, the operator C cyclically permutes the Pauli operators X, Y and Z, satisfying that $CX{=}YC$, $CY{=}ZC$, $CZ{=}XC$. We note that measuring each witness provides estimates for a distinct set of local observables. For instance, the first witness in equation 4 yields estimates for three observables: $ZI+IZ$, ZZ, and $XX+YY$. The six-witness ensemble in total provides 15 independent expectation values, which provide sufficient information about the two-qubit state. Thus, the measurement effort is reduced from 16 tomographic settings to 6 structured witness configurations offers significant practical benefits. Instead of perfoming a negativity test, the authors in Zhu et al. (2010) adopt a more stringent criterion:

$$\min_{\alpha} \mathrm{Tr}\big\{\rho_{\mathrm{sep}}\left(\cos^2\alpha I - \rho_w(\alpha)\right)\big\} \geq 0, \quad \forall \rho_{\mathrm{sep}} \in S_{ab}. \tag{6}$$

The criterion is violated by the set of entangled states that *can* be detected by this witness family. The above optimisation leads to the following quadratic WBM criterion,

$$S = 4f_1 f_2 - (f_3 - f_4)^2 \geq 0, \quad \forall \rho_{\mathrm{sep}} \in S_{ab}. \tag{7}$$

where $f_i \coloneqq \mathrm{Tr}\{E_i\rho\}$ are probabilities obtained from WBM $\mathcal{E}$. In essence, the process of measuring the family entanglement witnesses $\rho_w(\alpha)$ corresponds to measuring the projectors onto the eigenstate basis. The value of S in equation 7 depends on the underlying WBM. Thus, for a WBM $\mathcal{E}$ and state ρ, we denote equation 7 as $S_{\mathcal{E}}(\rho)$.

2.2 Fixed-Confidence Multi-Armed Bandit Policies

In this section, we briefly review some fixed-confidence policies for Best Arm Identification (BAI) and Good Arm Identification (GAI) in the stochastic Multi-Armed Bandit (MAB) setting, a canonical framework for sequential decision-making problems under uncertainty. A bandit instance (problem instance) consists of K arms, each described by a reward distribution ν_i supported on $\mathbb{R}$ with unknown mean μ_i. In each round t, the learner chooses an arm X_t, receives an independent reward $Z_t \sim \nu_{X_t}$, and chooses the subsequent action based on a specified policy. Below, we detail MAB policies that operate under fixed-confidence guarantees, where the objective is to make statistically reliable recommendations while minimising the number of samples.

2.2.1 Fixed-Confidence Best Arm Identification

In the BAI problem, the learner's objective is to identify the arm $i^\star = \arg\max_{i\in[K]} \mu_i$ with the largest expected reward. Without loss of generality, we enumerate the arms based on their expected reward $\mu_1 > \mu_2 \geq \cdots \geq \mu_K$ and denote the sub-optimality gap of arm i by $\Delta_i = \mu_{i^\star} - \mu_i$. We consider the class of (ϵ,δ)-*PAC* (Probably Approximately Correct) policies. Given a fixed probability of error δ and a tolerance in sub-optimality ϵ, the policies ensure that the learners recommendation is ϵ-suboptimal in expected reward with probability at least $1-\delta$. Throughout this paper, we operate in the *exact-correctness* regime by setting $\epsilon = 0$, requiring the learner to provide an exactly correct recommendation with confidence at least $1-\delta$. For brevity, we refer to these as δ-PAC policies. Formally, a BAI policy is said to be δ-PAC if it satisfies,

$$\mathbb{P}_{\boldsymbol{\mu}}(\hat{i}_\tau \neq i^\star) \leq \delta, \quad \mathbb{P}_{\boldsymbol{\mu}}(\tau < \infty) = 1. \tag{8}$$

where $\hat{i}_\tau$ is the best arm recommended at stoppage τ. The performance of these policies is primarily characterised by the expected stopping time $\mathbb{E}_{\boldsymbol{\mu}}[\tau]$, which represents the number of samples required to recommend a best arm with confidence $1-\delta$. Classical results show that the sample complexity improves progressively across algorithms: $\mathcal{O}(\Delta^{-2}\log(n\Delta^{-2}))$ for Successive Elimination (Even-Dar et al., 2002), $\mathcal{O}(\Delta^{-2}\log\Delta^{-2})$ for LUCB (Kalyanakrishnan et al., 2012), and $\mathcal{O}(\Delta^{-2}\log\log\Delta^{-2})$ for Exponential-Gap Elimination (Karnin et al., 2013). These results approach the theoretical lower bound of $\mathcal{O}(\Delta^{-2})$ (Mannor & Tsitsiklis, 2004), differing only by logarithmic factors. Notably, Farrell (1964) bridges this gap, proving that $\mathcal{O}(\Delta^{-2}\log\log\Delta^{-2})$ samples suffice to identify the best arm with error probability δ. Building upon this, lil'UCB Jamieson et al. (2014) uses finite-sample LIL-based concentration bounds to achieve near-optimal sample complexity.

2.2.2 Fixed-Confidence Good Arm Identification

The GAI problem generalises BAI by introducing a threshold ζ and defining the set $\mathcal{G} = \{\, i \in [K] : \mu_i \geq \zeta \,\}$ of "good" arms. The learner is unaware of the number of good arms $|\mathcal{G}| = m$, leading to the (m, K)-GAI formulation. Without loss of generality, assume $\mu_1 > \mu_2 \geq \ldots \geq \mu_m \geq \zeta \geq \mu_{m+1} \ldots \geq \mu_K$ and the learner is unaware of this indexing. Unlike the BAI setting, the GAI problem admits no notion of approximate correctness, since each arms mean reward is either above or below the fixed threshold ζ. Accordingly, fixed-confidence guarantees are expressed through (λ, δ)-*PAC* policies (Kano et al., 2018). A GAI policy is said to be (λ, δ)-PAC if, with probability at least $1-\delta$, it correctly identifies at least λ true good arms and does not misclassify any arm with $\mu_i < \zeta$. Here, λ specifies the number of correctly identified good arms. For each arm $i \in [K]$, the sub-optimality gaps are denoted by $\Delta_i \coloneqq |\mu_i - \zeta|$ and $\Delta_{i,j} = \mu_i - \mu_j$ and the sample complexity is expressed in terms of $\Delta = \min(\min_{i\in[K]} \Delta_i, \min_{j\in[K-1]} \Delta_{j,j+1}/2)$.

The goal, as in BAI, is to minimise the expected stopping time $\mathbb{E}_{\boldsymbol{\mu}}[\tau]$. However, a key difficulty in GAI is the exploration-exploitation dilemma of confidence, where the learner explores arms other than the empirical best arm to identify potentially 'good' arms with fewer pulls, while simultaneously exploiting the empirical best arm to increase confidence in its classification as a good arm. The Hybrid Dilemma-of-Confidence (HDoC) algorithm (Kano et al., 2018) combines UCB-based exploration (Auer et al., 2002) with LUCB-based elimination (Kalyanakrishnan et al., 2012), achieving sample complexity $\mathcal{O}\left(\Delta^{-2}\left(K\log\frac{1}{\delta} + K\log K + K\log\frac{1}{\Delta}\right)\right)$. The LIL-based refinement Tsai et al. (2024) lil'HDoC, employs tighter confidence widths to achieve $\mathcal{O}\left(\Delta^{-2}\left(K\log\frac{1}{\delta} + K\log K + K\log\log\frac{1}{\Delta}\right)\right)$ samples, the best-known order for fixed-confidence GAI policies. The specific connections between BAI/GAI and entanglement detection are elaborated in Section 3 and 5.

3 The Quantum MAB Framework For Entanglement Detection

In this section, we introduce the quantum Multi-Armed Bandit (MAB) framework for batch entanglement detection. We formalise the structural similarity between the stochastic MAB model and its quantum analogue, where the learner interacts with a batch of quantum states by performing structured measurements.

3.1 Problem Setting and Objective

In the stochastic MAB setting, pulling an arm i corresponds to sampling from a probability distribution $p_i(\cdot)$ with known support and unknown mean μ_i. Each pull yields a reward j with probability (w.p.) $p_i(j)$ and rewards across arm pulls are independent and identically distributed (i.i.d.). Analogously, in the quantum setting, each arm represents an unknown quantum state ρ. When ρ is measured, the reward distribution is determined by the fixed WBM. Specifically, if a WBM $\mathcal{E}$ is chosen, measuring a state ρ with $\mathcal{E}$ will result in a reward $j \in \{1,2,3,4\}$ with probability $\mathrm{Tr}(\rho E_j)$. Once the measurement is fixed, repeated measurements of ρ yield i.i.d. rewards. The key distinction lies in the source of the rewards: in the stochastic MAB model, rewards are sampled from classical distributions, whereas in the quantum MAB model, the rewards depend on the chosen WBM $\mathcal{E}$.

Given a batch of K unknown quantum states $\{\rho_1, \ldots, \rho_K\}$, of which an unknown subset $m < K$ states are entangled, the learner's objective is to correctly identify all entangled states while minimizing the total

number of measurements performed. Given a fixed WBM $\mathcal{E}$, the goal is to estimate the quadratic WBM criterion $S_{\mathcal{E}}(\rho_i)$ which indicates that ρ_i is entangled if $S_{\mathcal{E}}(\rho_i) < 0$ and is inconclusive otherwise. The learner applies the MAB routine to this (m, K) instance of quantum states under the chosen WBM $\mathcal{E}$ with the objective of accurately identifying $\mathcal{A}_{\text{ent}} = \{i \in [K] \text{ such that } S_{\mathcal{E}}(\rho_i) < 0\}$, using the fewest possible number of measurements. Since a single WBM may not detect all m entangled states, and the value of m itself is unknown, the MAB routine must be repeated for the six WBMs. Importantly, the measurement data collected under one WBM is *not* used to decide the next WBM; each WBM configuration should be treated as an independent instance.

3.2 The (m, K)-quantum MAB model

We summarise the stochastic-quantum MAB correspondence concisely in Table 1.

Table 1: Stochastic-Quantum MAB

Attributes	**Stochastic MAB**	**Quantum MAB**
Arms	Probability distributions $(p_1, p_2, \dots p_K)$	Density operators $\{\rho_1, \rho_2, \dots, \rho_K\}$
Measurement	–	WBM $\mathcal{E}$
Measurement Data	j w.p. $p_i(j), \forall i \in [K]$	j w.p. $\text{Tr}(E_j\ \rho_i),\ \forall j \in [4], \forall i \in [K]$
Parameters to estimate	$\boldsymbol{\mu} = (\mu_1, \mu_2, \dots \mu_K)$	$\boldsymbol{S}_{\mathcal{E}} = (S_{\mathcal{E}}(\rho_1), S_{\mathcal{E}}(\rho_2), \dots, S_{\mathcal{E}}(\rho_K))$
Objective	Identify $\mathcal{G}^C = \{i \in [K] \text{ such that } \mu_i \leq \zeta\}$	Identify $\mathcal{A}_{\text{ent}} = \{i \in [K] \text{ such that } S_{\mathcal{E}}(\rho_i) < 0\}$

The objective of the (m, K)-quantum MAB problem for entanglement detection aligns with the classical (m, K)-Bad Arm identification where the goal is to find the set of "bad" arms $\mathcal{G}^C = \{i \in [K] \text{ such that } \mu_i \leq \zeta\}$ whose mean rewards fall below a threshold ζ. Analogously, the (m, K)-quantum MAB problem seeks to identify the set of entangled states $\mathcal{A}_{\text{ent}}$ whose quadratic WBM scores $\boldsymbol{S}_{\mathcal{E}}$ violate the separability threshold. In essence, the (m, K)-Bad Arm identification setting and the (m, K)-quantum MAB problem for entanglement detection share a unified statistical structure, differing only in the interpretation of the reward model. To the best of our knowledge, this work is the first to establish a direct connection between stochastic MAB and quantum entanglement detection. This correspondence enables existing MAB algorithms to be directly applied in quantum settings, where the reward is encoded in the outcomes of witness-based measurements. We now formalise this correspondence by defining the (m, K)-quantum MAB setting.

Definition 2 *The (m, K)-quantum Multi-Armed Bandit (MAB) setting for entanglement detection is fully characterized by the tuple $(\mathcal{A}, \mathcal{E})$. Here, $\mathcal{A}$ denotes a finite action set with $|\mathcal{A}| = K$, consisting of $(K - m)$ two-qubit separable states and m two-qubit entangled states. The term $\mathcal{E}$ corresponds to a suitable Witness Basis Measurement (WBM).*

Remark 1 *The d-dimensional discrete multi-armed quantum bandit model (Lumbreras et al., 2022) is different from our formulation. The authors consider arms to be a finite set of observables and the environment, an unknown quantum state ρ. The objective is to learn the unknown quantum state ρ through an exploration-exploitation tradeoff. Given sequential oracle access to copies of ρ, each round involves selecting an observable to maximize its expectation value (reward). The information from previous rounds (history) aids in refining the action choice, thereby minimizing the regret, which is the difference between the obtained and maximal rewards. The authors also exploit the inherent linear structure in measurement outcomes and map it to the linear bandit setting. Specifically, let $\{\sigma\}_{i=1}^{d^2}$ be a set of orthogonal Hermitian matrices. The unknown environment $\rho = \sum_{i=1}^{d^2} \text{Tr}(\rho\sigma_i)\sigma_i = \sum_{i=1}^{d^2} \theta_i\sigma_i$ and arm $\mathcal{O}_t = \sum_{i=1}^{d^2} \text{Tr}(\mathcal{O}_t\sigma_i)\sigma_i = \sum_{i=1}^{d^2} A_{t,i}\sigma_i$. Then, $\text{Tr}(\rho\mathcal{O}_t) = \boldsymbol{\theta}^\top \mathbf{A}_t$ where $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots \theta_{d^2})$ and $\mathbf{A}_t = (A_{t,1}, A_{t,2}, \dots A_{t,d^2})$. In round t, pulling arm $\mathcal{O}_t$ provides a reward $X_t = \boldsymbol{\theta}^\top \mathbf{A}_t + \eta_t$, where η_t is 1-subgaussian.*

4 Parameterized Qubit States $\mathcal{F}$

To demonstrate the applicability of the stochastic MAB policies for entanglement detection, we consider *three* parameterized two-qubit states denoted by $\mathcal{F}$. We identify suitable WBMs from the wit-

ness family in equation 4 that are able to detect the same. We denote the first two witnesses in the witness family by $\mathcal{E}_1$ (base witness) and $\mathcal{E}_2$ (adapted using $(U_1, U_2) = (I, X)$), respectively. Here, $\mathcal{E}_1 := \{|00\rangle\langle 00|, |11\rangle\langle 11|, |\Psi^+\rangle\langle\Psi^+|, |\Psi^-\rangle\langle\Psi^-|\}$ and $\mathcal{E}_2 := \{|01\rangle\langle 01|, |10\rangle\langle 10|, |\Phi^+\rangle\langle\Phi^+|, |\Phi^-\rangle\langle\Phi^-|\}$.

4.1 Two-qubit Depolarized Bell States

For $w \in \mathbb{R}, -1/3 \leq w \leq 1$, a two-qubit **Depolarized Bell** state $\rho(w)$ is given by,

$$\rho(w) = w\,|\Upsilon\rangle\langle\Upsilon| + (1-w)\frac{I}{4}. \tag{9}$$

Here, $|\Upsilon\rangle$ represents any one of the four Bell states $|\Psi^\pm\rangle = (|01\rangle \pm |10\rangle)/\sqrt{2}$, $|\Phi^\pm\rangle = (|00\rangle \pm |11\rangle)/\sqrt{2}$. When $\Upsilon = |\Psi^-\rangle$, equation 9 is called a Werner state, and when $\Upsilon = |\Phi^+\rangle$, equation 9 is called an Isotropic state. The Peres-Horodecki criterion guarantees that $\rho(w)$ is separable when $-1/3 \leq w \leq 1/3$ and is entangled when $1/3 < w \leq 1$. Table 2 outlines the specific choices of WBM for the combination of the maximally mixed state with each of the four Bell states. When measured with these corresponding WBMs, the entangled depolarized Bell states are conclusively detected, determined by the value of $S = (w-1)^2/4 - w^2$ which is strictly positive for $-1 \leq w \leq 1/3$ and negative for $w > 1/3$.

Table 2: WBM for Depolarized Bell States

Depolarized State	**Pauli Basis**	**WBM**
$w\,\|\Phi^+\rangle\langle\Phi^+| + (1-w)I/4$	$\left[I + \alpha(XX - YY + ZZ)\right]/4$	$\mathcal{E}_2$
$w\,\|\Psi^+\rangle\langle\Psi^+| + (1-w)I/4$	$\left[I + \alpha(XX + YY - ZZ)\right]/4$	$\mathcal{E}_1$
$w\,\|\Psi^-\rangle\langle\Psi^-| + (1-w)I/4$	$\left[I + \alpha(-XX - YY - ZZ)\right]/4$	$\mathcal{E}_1$
$w\,\|\Phi^-\rangle\langle\Phi^-| + (1-w)I/4$	$\left[I + \alpha(-XX + YY + ZZ)\right]/4$	$\mathcal{E}_2$

4.2 Two-qubit Bell diagonal States

Bell diagonal states are a probabilistic mixture of the four Bell states. These states are more general than the ones in equation 9. Given parameters p_1, p_2, p_3 and p_4 such that $p_i \geq 0, \sum_i p_i = 1$, the Bell diagonal state is defined,

$$\rho_{\text{Bell}} = p_1\,|\Phi^+\rangle\langle\Phi^+| + p_2\,|\Psi^+\rangle\langle\Psi^+| + p_3\,|\Psi^-\rangle\langle\Psi^-| + p_4\,|\Phi^-\rangle\langle\Phi^-|. \tag{10}$$

The eigenvalues of $\rho_{\text{Bell}}^{\top_b}$ are calculated to be $1/2 - p_1$, $1/2 - p_2$, $1/2 - p_3$ and $1/2 - p_4$. Consequently, a Bell diagonal state is entangled if any one of these probabilities exceeds $1/2$, while the sum of the other three probabilities is less than $1/2$. Conversely, a Bell diagonal state is separable if all probabilities are less than or equal to $1/2$. Expressing equation 10 in the Pauli basis yields,

$$\rho_{\text{Bell}} = \frac{1}{4}\left[I + aXX + bYY + cZZ\right],$$

where $a = p_1 + p_2 - p_3 - p_4$, $b = -p_1 + p_2 - p_3 + p_4$ and $c = p_1 - p_2 - p_3 + p_4$.

Table 3: WBM for Bell Diagonal States

Probabilistic mixture	**a**	**b**	**c**	**WBM**
$p_1 > 0.5,\ p_2 + p_3 + p_4 < 0.5$	$+$	$-$	$+$	$\mathcal{E}_2$
$p_2 > 0.5,\ p_1 + p_3 + p_4 < 0.5$	$+$	$+$	$-$	$\mathcal{E}_1$
$p_3 > 0.5,\ p_1 + p_2 + p_4 < 0.5$	$-$	$-$	$-$	$\mathcal{E}_1$
$p_4 > 0.5,\ p_1 + p_2 + p_3 < 0.5$	$-$	$+$	$-$	$\mathcal{E}_2$

When ρ_{Bell} is entangled, the index for which $p_i > 1/2$ determines the sign of a, b, and c, see Table 3. It is notable that the signs of a, b and c follow a similar pattern to the Pauli basis expansion of various Depolarized

Bell states listed in Table 2. We observe that, for suitable combinations of a, b, and $c \in \{+1, -1\}$, the Bell diagonal state reduces to one of the Depolarized Bell states and states can be detected using the same WBMs, as in Table 2. Specifically, the value of S under the two WBMs in Table 3 is equal to $(1-p_1-p_4)^2-4(p_1-p_4)^2$ and $(1-p_2-p_3)^2-4(p_2-p_3)^2$, respectively. Depending on the probabilistic mixture, one of the two WBMs will conclusively result in $S < 0$.

4.3 Two-qubit Amplitude Damping on Depolarized Bell States

A qubit amplitude damping channel is a source of noise in superconducting circuit-based quantum computing and thus, serves as a realistic channel model for simulating lossy processes in these systems. Mathematically, it can be obtained from an isometry J,

$$J : \mathcal{H}_a \mapsto \mathcal{H}_b \otimes \mathcal{H}_c; \quad J^\dagger J = I_a \tag{11}$$

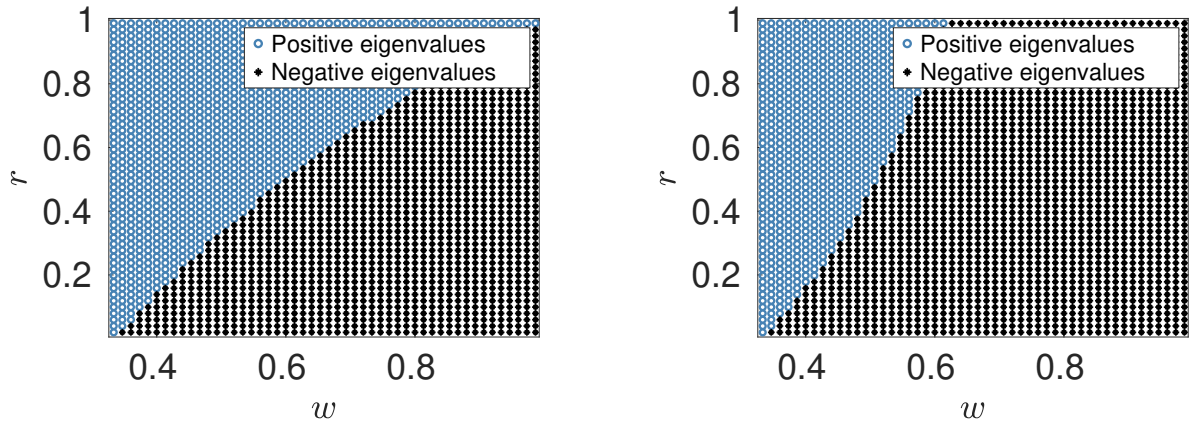
where $\mathcal{H}_a$ denotes the Hilbert space for the channel's input, and $\mathcal{H}_b$ and $\mathcal{H}_c$ represent the Hilbert spaces for the direct and complementary channel outputs, respectively. An isometry of the form,

$$\begin{aligned} J_1 \left|0\right\rangle_a &= \left|0\right\rangle_b \left|1\right\rangle_c, \\ J_1 \left|1\right\rangle_a &= \sqrt{1-r}\left|1\right\rangle_b \left|1\right\rangle_c + \sqrt{r}\left|0\right\rangle_b \left|0\right\rangle_c, \end{aligned} \tag{12}$$

where $0 \leq r \leq 1$ defines a pair of channels, $\mathcal{B}(A) = \mathrm{Tr}_c(JAJ^\dagger)$ and $\mathcal{C}(A) = \mathrm{Tr}_b(JAJ^\dagger)$. Here, $\mathcal{B}$ is an amplitude damping channel with damping probability r for the state $\left|1\right\rangle_a$ to decay to output state $\left|0\right\rangle_b$. The isometry $J_1 = K_0 \otimes \left|0\right\rangle + K_1 \otimes \left|1\right\rangle$ where K_0 and K_1 (Kraus) damping operators such that $K_0 = [0, \sqrt{r}; 0, 0]$ and $K_1 = [1, 0; 0, \sqrt{1-r}]$. For a single qubit represented by state ρ, the amplitude damped output is given by,

$$\mathcal{B}(\rho) = K_0 \rho K_0^\dagger + K_1 \rho K_1^\dagger. \tag{13}$$

We can extend equation 13 for two qubit states with damping probabilities r and q for the first and second qubit respectively. Assuming that $r = q$, we consider Depolarized Bell states equation 9 with amplitude damping.



(a) Range of r for eigenvalue corresponding to $\left|\Phi^\pm\right\rangle$ (b) Range of r for eigenvalue corresponding to $\left|\Psi^\pm\right\rangle$

Figure 1: A phase diagram representing the region of damping and depolarizing parameters, r and w, respectively, where the dampeddepolarized Bell state has negative or positive partial transpose.

Proposition 3 *For any damping probability $r > 0$, a Depolarized Bell state with amplitude damping can not be expressed as a Bell diagonal state equation 10.*

This fact can be readily demonstrated through a straightforward calculation. Consider the Isotropic state $\rho(w) = w\left|\Phi^+\right\rangle\left\langle\Phi^+\right| + (1-w)\frac{I}{4}$, which can be represented by the Bell diagonal state formed with probability

Table 4: The four eigenvalues of amplitude damped Depolarized Bell states

State with $\|\Phi\rangle^{\pm}$	State with $\|\Psi\rangle^{\pm}$	**Sign of eigenvalue**
$\frac{(w+1)(1-r^2)}{4}$	$\frac{(1-r)(1+r+w-wr)}{4}$	Always positive
$\frac{(w+1)(1-r)^2}{4}$	$\frac{(1-r)(1+r+w-wr)}{4}$	Always positive
$\frac{w(r-1)^2+(r+1)^2}{4}$	$\frac{r^2+1-w(1-r)^2+2\sqrt{w^2(1-r)^2+r^2}}{4}$	Always positive
$\frac{-r^2(w-1)+wr+(1-3w)}{4}$	$\frac{r^2+1-w(1-r)^2-2\sqrt{w^2(1-r)^2+r^2}}{4}$	Positive and Negative

distribution $(p_1, p_2, p_3, p_4) = ((3w+1)/4, (1-w)/4, (1-w)/4, (1-w)/4)$. In a Bell diagonal state, the diagonal elements corresponding to $|00\rangle\langle 00|$ and $|11\rangle\langle 11|$ are identical. In the case of an amplitude damped Isotropic state, we observe that,

$$p_2 = p_3 = \frac{1-r}{4}\left(w - wr - r - 1\right).$$

However, obtaining closed-form expressions for p_1 and p_4 when $r > 0$ is cumbersome. Specifically, the values on the diagonal corresponding to $|00\rangle\langle 00|$ and $|11\rangle\langle 11|$ is given by $w(r^2+1)/2 - (w-1)4(r+1)^2/4$ and $w(r-1)^2/2 - (w-1)(r-1)^2/4$, respectively. These expressions are equal only when $r = 0$.

Proposition 4 *For every $w \in [\frac{1}{3}, 1]$, there exists $\tilde{r} \subset [0,1]$ such that an amplitude damped Depolarized Bell state becomes separable.*

The PPT criterion asserts that a two-qubit state is entangled if and only if its partial transpose contains atleast one negative eigenvalue. For Bell states that are both amplitude damped and depolarized, we evaluate the eigenvalues and observe that one of them can exhibit either positive or negative values contingent upon the range of r. Detailed findings are presented in Table 4 and depicted graphically in Fig. 1a and Fig. 1b. Furthermore, the WBM for amplitude damped and Depolarized Bell states aligns with that of depolarized Bell states, as outlined in Table 2.

5 Stochastic MAB policies for Entanglement Detection

We apply stochastic MAB algorithms for entanglement detection in the parameterized states $\mathcal{F}$ from Section 3. The terminology follows the alignment with classical counterparts, as outlined in Table 1. Consider a set of K unknown states, denoted by $\mathcal{A} = \{\rho_1, \rho_2, \ldots, \rho_K\} \in \mathcal{F}$. To perform measurements on the arms, the learner must know the underlying WBM. Thus, we assume knowledge of the specific forms of the arms in $\mathcal{A}$. For instance, $\mathcal{A}$ could represent the set of isotropic states detectable under the WBM $\mathcal{E}_2$, where each state is of the form $\rho_i = w_i |\Phi^+\rangle\langle\Phi^+| + (1-w_i)\frac{I}{4}$, with w_i being unknown for all $i \in [K]$. With this assumption, we describe the MAB routine as follows: In each round $t \in \mathbb{N}$,

- The learner selects a state $\rho_i \in \mathcal{A}$.
- The learner performs a measurement $\mathcal{E}$ and obtains outcome j with probability $f_j = \mathrm{Tr}\{\rho_i E_j\}$, where $j \in \{1,2,3,4\}$.
- The learner updates the values of $\hat{\boldsymbol{S}}_{\mathcal{E}}$ and identifies the entangled arm(s) or continues.

For a given WBM $\mathcal{E}$, the values of $S_{\mathcal{E}}$ are bounded in $[-1,1]$. We use concentration inequalities applicable to 1-subgaussian[1] random variables—specifically, the law of iterated logarithm (Jamieson et al., 2014) for a finite sum of 1-subgaussian random variables:

[1]A 1-subgaussian random variable is a real, centered random variable X that satisfies $\mathbb{E}[e^{sX}] \leq e^{s^2/2}$ for any $s \in \mathbb{R}$.

Algorithm 1 Successive Elimination Algorithm

Input: $\zeta = 0$, δ, $\mathcal{A}$, WBM $\mathcal{E}$
Output: Ω
  Initialize active set $\Omega \leftarrow \mathcal{A}$
  Set initial estimates: $\hat{S}_{i,N_i(t)} = 0,\ \forall i \in \Omega$
  for $t = 1, 2, 3, \ldots$ **do**
    for $\rho_i \in \Omega$ **do**
      Perform measurement $\mathcal{E}$ on ρ_i
      Update $\hat{S}_{i,N_i(t)}$ based on outcome $j \in \{1, 2, 3, 4\}$
      Update confidence width $U\left(N_i(t), \frac{\delta}{c_\varepsilon K}\right)$ (see Lemma 5)
      Compute lower confidence bound: $\mathrm{LCB}_i(t) \leftarrow \hat{S}_{i,N_i(t)} - U\left(N_i(t), \frac{\delta}{c_\varepsilon K}\right)$
    end for
    if $\mathrm{LCB}_i(t) > 0$ for $i \in \Omega$ **then**
      Update active set: $\Omega \leftarrow \Omega - \{i\}$
    end
    if $|\Omega| = 1$ **then**
      Return Ω
    end
  end for

Lemma 5 *Let $X_1, X_2, \ldots X_t$ be* i.i.d. *sub-gaussian random variables with scale parameter $\sigma = 1$. For any $\varepsilon \in (0, 1)$, $\delta \in \left(0, \frac{\log(1+\varepsilon)}{e}\right)$, one has with probability at least $1 - c_\varepsilon \delta^{(1+\varepsilon)}$ for all $t \geq 1$,*

$$\frac{1}{t}\sum_{s=1}^{t} X_s \leq U(t, \delta), \tag{14}$$

where $U(t,\delta) = (1+\sqrt{\varepsilon})\sqrt{\frac{2(1+\varepsilon)}{t}\log\left(\frac{\log((1+\varepsilon)t)}{\delta}\right)}$ is the confidence width and $c_\varepsilon = \frac{2+\varepsilon}{\varepsilon}\left(\frac{1}{\log(1+\varepsilon)}\right)^{1+\varepsilon}$.

Proof: Readers can refer in Jamieson et al. (2014, Lemma 1). □

In the subsequent sections, we discuss two MAB policies: successive elimination, which is applicable when there is a guarantee of one entangled arm among K arms, and the HDoC policy, designed for scenarios where there are m entangled arms among K, with m being unknown.

5.1 Successive Elimination Algorithm

Consider the set of states $\mathcal{A} = \{\rho_1, \rho_2, \ldots, \rho_K\}$ detectable under WBM $\mathcal{E}$, with the guarantee that exactly one arm in the set is entangled. The underlying problem instance $\boldsymbol{S}_\mathcal{E}$ satisfies the condition $S_\mathcal{E}(\rho_1) \geq S_\mathcal{E}(\rho_2) \geq \cdots > S_\mathcal{E}(\rho_{K-1}) > 0 > S_\mathcal{E}(\rho_K)$. To address this, we adapt the Successive Elimination algorithm (Even-Dar et al., 2002), as outlined in Algorithm 1. This modified algorithm takes as input the set $\mathcal{A}$, the threshold $\zeta = 0$, WBM $\mathcal{E}$ and the error probability δ, and it outputs the entangled state $i^\star = \arg\min_{i\in[K]} S_\mathcal{E}(\rho_i)$. Let $N_i(t)$ denote the number of times ρ_i has been measured in t rounds and $\hat{S}_{i,N_i(t)}$ is the estimate of $S_\mathcal{E}(\rho_i)$ obtained on measuring ρ_i until time t. The algorithm maintains an active set Ω and measures every state in it. In order to identify $i^\star$, the policy eliminates states states Lower Confidence Bound (LCB) exceeds the threshold and halts when only one state remains in the active set.

Lemma 6 *Algorithm 1 is δ-PC.*

Proof: The proof is presented in Appendix A.1.1. □

The correctness of Algorithm 1 and the copy complexity of identifying the entangled arm is presented below.

Theorem 7 *With probability at least $1-\delta$, the entangled state $i^\star = K = \arg\min_{i\in[K]} S_{\mathcal{E}}(\rho_i)$ remains in the active set Ω till termination.*

Proof: The proof is presented in Appendix A.1.2. □

Theorem 8 *With probability at least $1-\delta$, Algorithm 1 identifies the entangled state $i^\star$, requiring $\sum_{i\in[K]} \mathcal{O}\left(\Delta_i^{-2}\log\left(\frac{K\log\Delta_i^{-2}}{\delta}\right)\right)$ copies. Here, $\Delta_i = |S_{\mathcal{E}}(\rho_i)-\zeta|$ denotes the sub-optimality gap with respect to the threshold ζ.*

Proof: The proof is presented in Appendix A.1.3. □ We observe that the sample complexity derived in Theorem 8 is within a $\log(K)$ factor of the optimal bound, as demonstrated in Theorem 1 of Jamieson et al. (2014). This result follows from the concentration bound established in Lemma 5, which forms the basis for the MAB policy described in the following section.

5.2 lil'HDoC Algorithm

The lil'HDoC algorithm (Tsai et al., 2024) builds on the HDoC algorithm (Kano et al., 2018) by leveraging finite LIL concentration bounds (Lemma 5) instead of the LCB-based identification rule (Kalyanakrishnan et al., 2012). To explore among promising arms, lil'HDoC adopts the sampling rule from Kano et al. (2018), derived from the UCB algorithm for regret minimization (Auer et al., 2002). It improves sample complexity over HDoC by utilizing the LIL bound, where the $\sqrt{\log\log t/t}$ factor has a higher growth rate than the $\sqrt{\log t/t}$ factor in the LCB bound. In other words, there exists a value T such that for all $t > T$, $c_1, c_2 \in \mathbb{R}^+$,

$$c_1\sqrt{\frac{\log t}{t}} > c_2\sqrt{\frac{\log\log t}{t}}.$$

The confidence bound for HDoC grows as $\alpha(t) = \sqrt{\ln\left(\frac{4Kt^2}{\delta}\right)/2t}$. Through straightforward calculations, the smallest integer T such that the confidence bound $U\left(T, \delta/c_\varepsilon K\right)$ is greater than $\alpha(T)$ is,

$$T \geq \frac{1}{4}\log(K+1)\log\left(\max\left(\frac{1}{\delta}, 2\right)\right)c_\varepsilon^{3/2}. \tag{15}$$

Thus, if each state is measured T times initially, lil'HDoC achieves comparable identification capabilities to HDoC with $\mathcal{O}\left(\log(K+1)\log\left(\max\left(1/\delta, 2\right)\right)\right)$ copies of each state. We note that small values of ϵ tighten the confidence radius and therefore incur more samples before elimination, while large ϵ values reduce the number of samples with a higher chance of premature arm elimination. The asymptotic growth of $U(t,\delta)$ with ϵ is sublinear and the policy's correctness remains unaffected for $\epsilon > 0$. The 'warm-start' phase parameter T controls the number of measurements collected before adaptive allocation begins. Once the threshold in equation 15 is crossed, the δ-PAC guarantees and copy complexity depend primarily on (Δ_i, K, δ) and not on T itself.

Consider K states such that $S_{\mathcal{E}}(\rho_1) \geq S_{\mathcal{E}}(\rho_2) \ldots > S_{\mathcal{E}}(\rho_{K-m}) > 0 > S_{\mathcal{E}}(\rho_{K-m+1}) \ldots > S_{\mathcal{E}}(\rho_K)$, with m being unknown. The algorithm takes as input, the set of states $\mathcal{A}$, threshold $\zeta = 0$, WBM $\mathcal{E}$ and the error probability δ and outputs $\mathcal{A}_{\text{ent}} = \{i \in [K] \text{ such that } S_{\mathcal{E}}(\rho_i) < 0\}$. The algorithm maintains an active set Ω and terminates when the set $\Omega = \emptyset$. To demonstrate the correctness of Algorithm 2, we first show that the algorithm is (λ, δ)-PAC for all $\lambda \in [K]$ and then characterize the copy complexity of identifying m entangled states.

Lemma 9 *Algorithm 2 is δ-PAC.*

Proof: *The proof is presented in Appendix A.2.1.* □

Theorem 10 *With probability at least $1-\delta$, the algorithm identifies all the states in $\mathcal{A}_{ent}$.*

Algorithm 2 LIL'HDOC ALGORITHM

Input: $\zeta = 0$, δ, $\mathcal{A}$, WBM $\mathcal{E}$
Output: $\mathcal{A}_{\text{ent}}$
  Initialize active set $\Omega \leftarrow \mathcal{A}$, $\mathcal{A}_{\text{ent}} \leftarrow \emptyset$
  Set initial estimates: $\hat{S}_{i,N_i(t)} = 0, \ \forall i \in \Omega$
  for $\rho_i \in \Omega$ **do**
    Perform measurement $\mathcal{E}$ on ρ_i for T times
    $N_i(t) \leftarrow T$
    Update $\hat{S}_{i,T}$ based on outcome $j \in \{1,2,3,4\}$
  end for
  while $\Omega \neq \emptyset$ **do**
    Find $h_t = \arg\max_{i \in \mathcal{A}} \ \hat{S}_{i,N_i(t)} + \sqrt{\frac{\log t}{2N_i(t)}}$
    Perform measurement $\mathcal{E}$ on ρ_{h_t}
    $t \leftarrow t+1$
    Update $\hat{S}_{i,N_i(t)}$ based on outcome $j \in \{1,2,3,4\}$
    Update confidence width $U\left(N_i(t), \frac{\delta}{c_\varepsilon K}\right)$
    if $\hat{S}_{h_t,N_{h_t}(t)} - U\left(N_{h_t}(t), \frac{\delta}{c_\varepsilon K}\right) \geq \zeta$ **then**
      Remove h_t from Ω
    else if $\hat{S}_{h_t,N_{h_t}(t)} + U\left(N_{h_t}(t), \frac{\delta}{c_\varepsilon K}\right) < \zeta$ **then**
      Add h_t to $\mathcal{A}_{\text{ent}}$
      Remove h_t from Ω
    end
  end while

Proof: The proof is presented in Appendix A.2.2. □

With $T = 1$ in equation 15, it can be seen from Theorem 8 that the number of samples required to identify an entangled state $\rho_i \in \mathcal{A}$ is $\mathcal{O}\left(\Delta_i^{-2} \log\left(\frac{K \log \Delta_i^{-2}}{\delta}\right)\right)$. However, in practice, T is chosen to be larger than 1, and the total sample complexity is expressed in terms of $\Delta = \min_{i \in [K]} \Delta_i$.

Theorem 11 *With probability $1-\delta$ and an initialization of T measurements, Algorithm 2 identifies the entangled states using $\mathcal{O}\left(\Delta^{-2}\left(K \log \frac{1}{\delta} + K \log K + K \log\log \frac{1}{\Delta}\right)\right) + \mathcal{O}\left(K \log(K+1) \log\left(\max\left(\frac{1}{\delta}, e\right)\right)\right)$ copies.*

Proof: The first term in the sample complexity is derived in Appendix A.1.3 and the second term follows from equation 15. □

6 Implementation and Simulations on IBMQ Cloud

This section presents an experimental workflow for detecting entangled states from an ensemble of Bell Diagonal states. Sections 6.1 and 6.2 describe the procedures for generating Bell Diagonal states (BDS) and their corresponding WBMs, respectively. The performance of the MAB policies (see Section 5) are presented through numerical findings in Sections 6.4.

6.1 Generating Bell Diagonal States

Bell Diagonal States (BDS) are constructed as convex combinations of the four Bell states equation 10, forming a geometric tetrahedron $\mathcal{T}$ and are represented by:

$$\rho_{\text{Bell}} = \sum_{j=1}^{4} p_j \left|\Upsilon\right\rangle \left\langle\Upsilon\right| = \frac{1}{4}\left[I + \sum_{j=1}^{3} t_j \sigma_j^A \otimes \sigma_j^B\right]. \tag{16}$$

Here, σ_j's are the Pauli operators and (t_1, t_2, t_3) are the coordinates within the tetrahedron $\mathcal{T}$.

The mapping $\{p_j\}_{j=1}^4 \to (t_1, t_2, t_3)$ equation 17 is implemented through the quantum circuit proposed by Pozzobom & Maziero (2019); Riedel Gårding et al. (2021) and is shown in Fig. 2.

$$\begin{aligned}
\sqrt{p_1} &= \cos(\psi) \\
\sqrt{p_2} &= \sin(\psi)\cos(\theta) \\
\sqrt{p_3} &= \sin(\psi)\sin(\theta)\cos(\varphi) \\
\sqrt{p_4} &= \sin(\psi)\sin(\theta)\sin(\varphi)
\end{aligned} \tag{17}$$

The sub-circuit G encodes the probabilities $\{p_j\}_{j=1}^4$ into canonical coordinates (ψ, θ, φ) on the unit 3-sphere, and sub-circuit B entangles the states in the Bell basis. Finally, BDS $\rho_{\text{Bell}} = \rho_{cd}$ is obtained by taking a partial trace on qubits a and b.

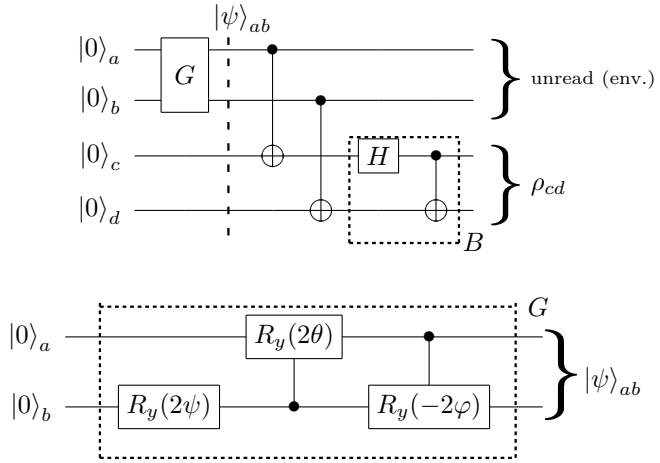


Figure 2: Four-qubit circuit for generating BDS with canonical encoder G shown below.

6.2 Implementing Witness Basis Measurements

As outlined in Table 3, BDS are detectable under WBMs $\mathcal{E}_1$ and $\mathcal{E}_2$. To measure in the Pauli-Z basis, we apply appropriate unitary transformations to $\mathcal{E}_1$ and $\mathcal{E}_2$. The corresponding transformations are realized through circuits $\text{CIRC}_{\mathcal{E}_1}$ and $\text{CIRC}_{\mathcal{E}_2}$ shown in Fig. 3 and applied to qubits c and d (see Fig. 2) before measurement.

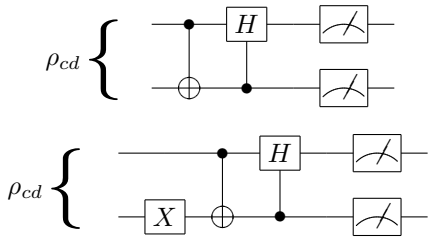


Figure 3: Circuits $\text{CIRC}_{\mathcal{E}_1}$ (top) and $\text{CIRC}_{\mathcal{E}_2}$ (bottom) perform the unitary transformations required to map $\mathcal{E}_1$ and $\mathcal{E}_2$ into the Pauli-Z basis.

6.3 Workflow for entanglement detection

We propose a workflow for detecting entanglement in BDS without assuming prior knowledge of the specific WBM. Instead, WBMs are sequentially adapted using suitable unitary transformations, as detailed in Section

2.1.2. To generate set $\mathcal{A} = \{\rho_1, \rho_2, \ldots, \rho_K\}$ of BDS, we construct K sets of probabilities for combining the four Bell states. Specifically, m states are generated with $\max_j p_j > \frac{1}{2}$, while the remaining $K - m$ states satisfy $\max_j p_j \leq \frac{1}{2}$. These probabilities are encoded following the procedure outlined in Fig. 2, where the BDS circuit for state ρ_i is denoted as BDSi. Subsequently, one of two WBM circuits, CIRC$_{\mathcal{E}_1}$ or CIRC$_{\mathcal{E}_2}$, is appended to the respective BDS circuit. Algorithm 3 outlines this workflow for BDS and takes the following inputs: threshold $\zeta = 0$, error δ, BDS circuits $\{\text{BDS}_i\}$, WBM circuits CIRC$_{\mathcal{E}_1}$ and CIRC$_{\mathcal{E}_2}$ and the initial choice of WBM. Notably, the initial WBM selection is arbitrary, as the sequence of WBM adaptations does not rely on prior state estimation.

Algorithm 3 Workflow for Entanglement Detection in BDS

Input: $\zeta = 0$, δ, $\{\text{BDS}_i\}$, CIRC$_{\mathcal{E}}$, WBM choice = 1
Output: A_{ent}, Stopping time τ
  Run Algorithm 2 on $\{\text{BDS}_i\}$ with circuit CIRC$_{\mathcal{E}_1}$ on K states
  Return entangled states $|A_{\text{ent},1}| = \tilde{m}$ and stopping time τ_1.
  if $(\tilde{m} = K)$ **then**
    $A_{\text{ent},2} \leftarrow \emptyset$, $\tau_2 \leftarrow 0$.
  else if $\tilde{m} < K$ **then**
    Run Algorithm 2 on $\{\text{BDS}_i\}$ with circuit CIRC$_{\mathcal{E}_2}$ on $K - \tilde{m}$ states
    Return entangled states $A_{\text{ent},2}$ and stopping time τ_2.
  end
  $A_{\text{ent}} \leftarrow A_{\text{ent},1} + A_{\text{ent},2}$, $\tau \leftarrow \tau_1 + \tau_2$

The learner does not initially know under which WBM the BDS are detectable. Consequently, at least one iteration of Algorithm 2 must be executed. In the first iteration, the algorithm processes circuits corresponding to K states with WBM $\mathcal{E}_1$ (or $\mathcal{E}_2$) and identifies a subset of entangled states, $\tilde{m}$, where $0 \leq \tilde{m} \leq K$. In the second iteration, Algorithm 2 is applied to the $K - \tilde{m}$ states that remain undetected by using circuits with WBM $\mathcal{E}_2$ (or $\mathcal{E}_1$) as inputs. Let us define $\Delta_1 := \min |\boldsymbol{S}_{\mathcal{E}_1}|$, $\Delta_2 := \min |\boldsymbol{S}_{\mathcal{E}_2}|$ and $\Delta_{\min} = \min\{\Delta_1, \Delta_2\}$, then

Corollary 12 *With probability* $1 - \delta$ *and* $T = 1$*, Algorithm 3 identifies entangled BDS using* $2\mathcal{O}\left(\Delta_{min}^{-2}\left(K \log \frac{1}{\delta} + K \log K + K \log\log \frac{1}{\Delta_{min}}\right)\right)$ *copies.*

6.4 Qiskit Experiment

The workflow presented in Algorithm 3 is simulated on IBM's Qiskit. The implementation is available in Bharati (2025). We present numerical results on the achievable copy complexity for entanglement detection in BDS. The experimental setup is given as follows:

- *Simulation Environments:* The workflow is executed across three computational setups: (i) **AerSimulator** for idealized, noiseless simulations, (ii) **FakeBrisbane** backend to simulate noisy quantum environments, and (iii) **ibm-brisbane** for real quantum processing unit (QPU) computations.
- *Problem Instance*: We consider $K = 5$ states of which $m = 3$ are entangled. The probabilities are suitably generated and the true corresponding parameters under $\mathcal{E}_1$ and $\mathcal{E}_2$ are,

$$\boldsymbol{S}_{\mathcal{E}_1} = [0.6306, -0.2688, 0.5232, 0.1796, 0.0695]\,,$$
$$\boldsymbol{S}_{\mathcal{E}_2} = [-0.0749, 0.5963, -0.1735, 0.2801, 0.3768]\,.$$

- Each state was measured 10^6 times on backends (i, ii) and 10^5 times on (iii). Algorithm 3 was run 20 times on (i, ii) and 5 times on (iii) for $\delta \in (0, 1)$. We plot the average number of copies measured until stoppage on the y-axis and $\log(1/\delta)$ on the x-axis, as shown in Fig.4. Here, we note that the large standard deviation for the trend in backend (iii) arises due to the limited number of experiment iterations, constrained by available compute resources.

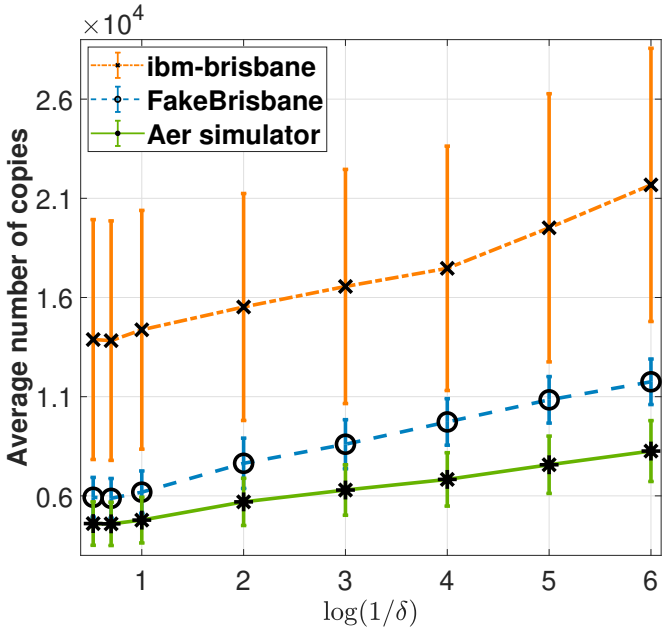


Figure 4: Copy complexity for entanglement detection in BDS v/s $\log(1/\delta)$

From Corollary 12, we observe that the factor $\log(1/\delta)$ has a multiplicative effect on the sample complexity, while the average copy complexity is primarily determined by $\Delta_{\min}$. The values of $S_{\mathcal{E}}$ are governed by the four frequencies f_1, f_2, f_3, and f_4, as defined in equation 7. While the true values of the f_i's are calculated using $\mathrm{Tr}\{\rho_{\mathrm{Bell}} E_i\}$, the values of f_i obtained from register countsbased on simulations performed on different backendsdiffer from the true values upto $\mathcal{O}(10^{-2})$. Due to measurement noise and decoherence, the goalpost for $S_{\mathcal{E}}$ varies across different backends and these differences influence $\Delta_{\min}$. One option is to mitigate the measurement noise (see details in the appendix, Sec. A.3)

7 Entanglement Detection in Arbitrary Quantum States

This section outlines a routine for detecting entanglement in arbitrary two-qubit quantum states. Specifically, we consider K arbitrary states, one of which is entangled, and describe an MAB routine along with numerical results.

7.1 Numerical Experiment

The workflow outlined in Algorithm 4 is implemented in MATLAB. The algorithm takes the following inputs: a threshold ζ, an error parameter δ, a set of K states $\mathcal{A}$ (with the promise that one state is entangled), and a permutation of $\{1, 2, 3, 4, 5, 6\}$ that specifies the order in which the WBMs should be adapted. As this is a promise problem, the algorithm terminates as soon as it identifies an entangled state, without needing to measure with all six WBMs. The different modules in the software are described below:

- *Generating arbitrary quantum states*: To generate random density matrices, we follow the method described in Zyczkowski & Sommers (2001). Specifically, we start by generating a complex matrix $A \in \mathbb{C}^{4\times 4}$, where the real and imaginary parts of each element are independently sampled from a normal distribution. We then compute the density matrix ρ by normalizing $AA^\dagger$, resulting in $\rho = AA^\dagger/\mathrm{Tr}(AA^\dagger)$. This procedure ensures that ρ is a valid density matrix.
- *Experiment Setup*: In this experiment, we generate 1000 distinct instances of $K = 5$ full rank arbitrary states, ensuring that each instance contains exactly one entangled state. These instances are validated using the PPT criterion to confirm their validity.

Algorithm 4 Entanglement detection for arbitrary states

Input: $\zeta = 0$, δ, $\mathcal{A} \leftarrow \{\rho_1, \rho_2 \dots \rho_K\}$, WBM Order P
Output: A_{ent}
  flag $\leftarrow 1$, $I \leftarrow 1$
  while flag **do**
    With $\mathcal{E} \leftarrow \mathcal{E}_{P(I)}$, run Algorithm 2 for K arms
    Return entangled arm $A_{\text{ent}}^{(I)}$
    if $|A_{\text{ent}}^{(I)}| = 1$ **then**
      flag $\leftarrow 0$
    else
      $I \leftarrow I + 1$
    end
  end while
  $A_{\text{ent}} \leftarrow A_{\text{ent}}^{(I)}$

- We test the efficacy of using the single parameter family of witnesses equation 4 to detect entanglement in arbitrary states. For $\delta \in (0, 1)$, we report the detection ratio which is the fraction of times the entangled state is accurately identified by the MAB policy. This result is shown in Fig. 5. We observe that the detection ratio diminishes with larger error margins.

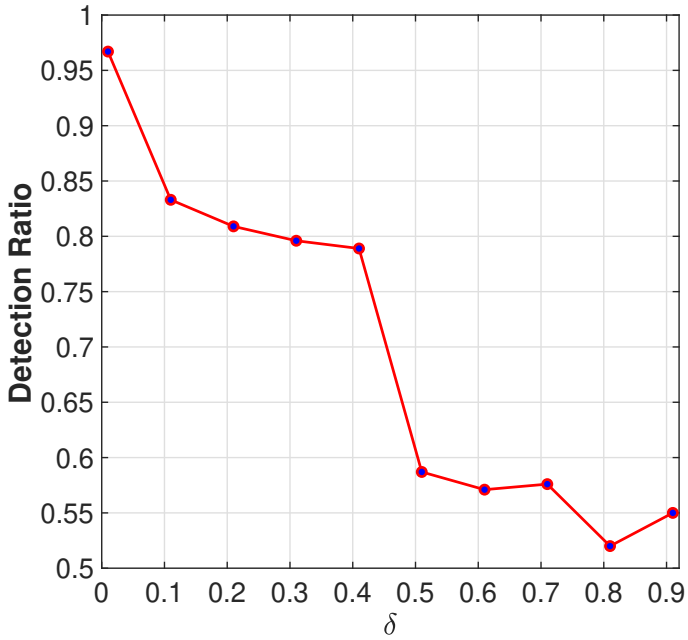


Figure 5: Entanglement Detection ratio v/s δ for arbitrary quantum states

- For a random order of WBM, we analyze how many measurements from the witness family are required to detect a single valid entangled state among a set of K states. For $\delta \in (0, 1)$, we present the frequency distribution of the number of WBMs used, displayed as a cumulative histogram in Fig. 6. For significantly larger values of δ, the lower detection ratios indicate that the algorithm terminates upon identifying the wrong state, preventing further adaptation and primarily (around 85%) relying on up to three witnesses.

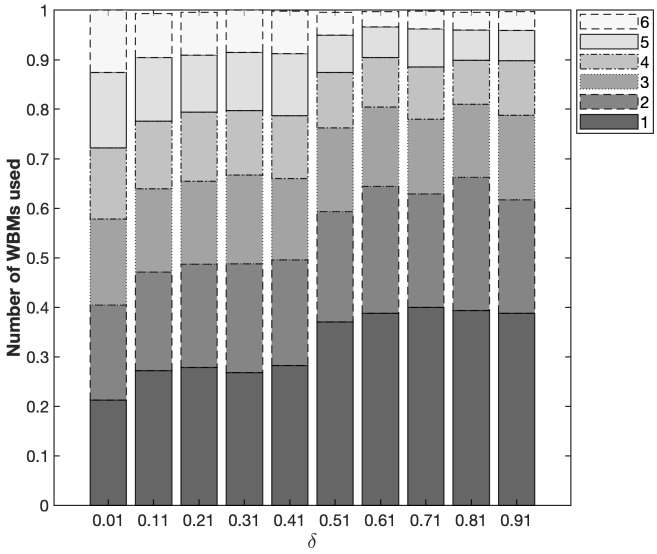


Figure 6: The cumulative histograms compare between the number of WBMs used to detect one valid entangled state across different values of δ.

Table 5: Examples of arbitrary pure entangled states detected by the family of witnesses equation 4

Pure entangled states $\|\psi_1\rangle, \|\psi_2\rangle$ and $\|\psi_3\rangle$	Values under $(S_{\mathcal{E}_i})_{i=1}^{6}$
$[0.2687 + 0.0375i; 0.2406 + 0.4090i; 0.0502 + 0.6162i; 0.2413 + 0.5107i]$	$(-0.1851, 0.3160, 0.1598, -0.0058, 0.2177, -0.1947)$
$[0.0565 + 0.3355i; 0.0508 + 0.0686i; 0.4885 + 0.5191i; 0.5689 + 0.2125i]$	$(0.1562, -0.0280, -0.1135, 0.1832, -0.0779, 0.1373)$
$[0.1953 + 0.4438i; 0.4958 + 0.4009i; 0.0069 + 0.3495i; 0.0322 + 0.4848i]$	$(-0.1851, 0.3160, 0.1598, -0.0058, 0.2177, -0.1947)$

In this experiment, we encountered edge cases, i.e., instances of pure states ρ with value of $S_{\mathcal{E}}(\rho) = 0$. For such edge cases, the algorithm took a significantly long time to converge and, despite this, incorrectly estimated the value of $S_{\mathcal{E}}(\rho)$. Consequently, we adjusted the threshold to -1×10^{-3} and imposed a cutoff on the sample complexity at 1×10^{12} to better reflect the real-time performance of this policy. This experiment can be extended to the scenario where there are m entangled states. However, since m is unknown and the states may be detectable under any of the WBMs, the routine would necessitate measuring under all WBMs to reliably identify the entangled states.

7.2 Numerical Outliers

We present an example of a PPT-verified entangled state that yields positive values for $S_{\mathcal{E}}(\rho)$ under all six WBMs. Consider the pure entangled states and their corresponding $\boldsymbol{S}_{\mathcal{E}}$ values, as shown in Table 5. The state $\rho = \sum_{i=1}^{3} p_i \left|\psi_i\right\rangle \left\langle\psi_i\right|$, where $\left|\psi_i\right\rangle$ are defined in Table 5, and $(p_i)_{i=1}^{3} = (0.2936, 0.0655, 0.6409)$, has a negative eigenvalue of -0.029 after applying the partial transpose, thus confirming it as a PPT-verified entangled state. However, the values of $(S_{\mathcal{E}}) = (0.0732, 0.1727, 0.1257, 0.1139, 0.0736, 0.0296)$ under the six WBMs are all non-negative. This indicates that the state cannot be detected by the witness family described in equation 4.

We derive an observation on the nature of such states, focusing specifically on the eigenstate $\left|\lambda\right\rangle_{\max} = [0.3773 - 0.1445i, 0.4768 - 0.3244i, 0.4598 + 0.0809i, 0.5351]$, which corresponds to the largest eigenvalue of ρ. This eigenstate has a Schmidt coefficient close to, but not exactly equal to 1, suggesting that it lies near the boundary of separable states while still remaining entangled. The pure state $\left|\lambda\right\rangle_{\max} \left\langle\lambda\right|_{\max}$ produces the following values for $(S_{\mathcal{E}}) = (0.0380, 0.1269, 0.0401, 0.1054, 0.0221, 0.0074)$. This highlights that both pure and mixed entangled states can yield inconclusive results when measured using this specific witness family.

In these cases, it is crucial to measure all six witnesses a sufficient number of times to accurately obtain the expected values of the corresponding observables. Additionally, performing FST can help determine the entanglement of these states using other separability criteria.

8 Discussions

8.1 The MAB Advantage

While traditional FST and fixed witness testing methods (Zhu et al., 2010) are direct and computationally feasible for the state space of two qubits, the proposed MAB-based pipeline provides a practical and theoretical advantage for entanglement detection in parameterised two-qubit states considered in this work. First, the MAB policies operate under fixed-confidence guarantees, namely δ-PAC in the BAI setting and δ-correct in the GAI setting, providing statistically certified recommendations with a probability of at least $1-\delta$. In the BAI case, this corresponds to identifying the best arm (or entangled state) with exact correctness ($\epsilon = 0$ regime), whereas in the GAI case, it ensures that all arms declared as good indeed exceed the specified threshold ζ with the same level of confidence. Thus, the recommended set of m entangled states is statistically certified without requiring any state reconstruction. In contrast, FST reconstructs the state to a specified accuracy and does not provide any explicit confidence guarantee on the entanglement aspect.

Second, the key advantage in the MAB formulation lies in its adaptive allocation of measurement resources across the K candidate states. In contrast, both FST and repeated witness-testing approaches (Zhu et al., 2010) treat each state independently and sequentially. FST expends copies toward exhaustive state reconstruction, while repeated witness measurements reduce statistical noise in the $S_{\mathcal{E}}$ estimates through non-adaptive measurement reuse, without a principled stopping rule. This results in a significant redundancy: sample-optimal FST with collective measurements requires $\mathcal{O}(16/\epsilon^2)$ copies per state Haah et al. (2017), and when applied to all K states, scales linearly as $\mathcal{O}(16K/\epsilon^2)$. On the contrary, the MAB-based approach estimates $S_{\mathcal{E}}(\cdot)$ with the detection difficulty controlled by Δ_i, incurring $O\left(\sum_{i=1}^{K} \Delta_i^{-2} \log \frac{K \log \Delta_i^{-2}}{\delta}\right)$ copies. Unlike the uniform ϵ^{-2} scaling in FST, the MAB complexity scaling achieves polynomial dependence on the separability gaps $\{\Delta_i\}$ and only polylogarithmic dependence on K and $(1/\delta)$. This adaptivity in allocating measurements allows the MAB-based strategy to concentrate the sampling effort on the most uncertain states and achieve early stopping when confidence thresholds are met.

Finally, we note a pronounced empirical advantage using the proposed MAB approach. As illustrated in Fig. 4, the lil-HDoC policy achieves δ-correctness with high confidence using only $\mathcal{O}(10^4)$ copies for the scale of $K = 5$ parameterised states across IBMQ backends (Aer, FakeBrisbane, ibm-brisbane) while FST requires $\mathcal{O}(10^6)$ total copies to achieve $\epsilon \approx 10^{-2}$ for the same instance. We emphasise that this two-order-of-magnitude reduction in copy complexity is enabled through adaptive sampling and early stopping. Overall, the MAB-based approach offers a quantitatively demonstrated reduction in copy complexity, explicit confidence guarantees, adaptivity in distributing measurement effort and scalability for batch detection tasks within the tested family of parameterised states.

8.2 Does the MAB routine optimize WBM ordering?

As outlined in Dai et al. (2014), there are WBM optimization strategies that prescribe an optimal WBM ordering for efficiently detecting whether a *single* arbitrary two-qubit quantum state is entangled. One such adaptive strategy uses the maximum-likelihood maximum-entropy (MLME) estimate of the unknown state, based on causal measurement data. Using this estimate, the subsequent WBM is identified to be the one minimizing the quadratic separability criterion. This leads to partial estimation of the quantum state.

In the context of batch entanglement detection, where an unknown number m of entangled states out of a set of K states may be detectable under different witnesses, implementing the WBM adaptive strategies from Dai et al. (2014) would be both time-consuming and complex. This is because each of the K states may require a unique permutation of the WBM ordering. Furthermore, the goal of the proposed MAB framework is to *minimize* the number of measurements needed for detecting entanglement in a given set of quantum

states under a specific WBM. Notably, this framework *does not* optimize the WBM ordering across multiple MAB runs.

The closest comparison is with Fig. 6, which depicts the cumulative frequency of WBMs used. This aligns with the cumulative percentage of states identified under the WBM family, as seen in schemes 1A and 4A of the recently reported incomplete state estimation techniques (see (Dai et al., 2014, Fig. 1)). However, this approach does not address the batch entanglement detection problem. The WBM adaptation scheme A in Dai et al. (2014) successfully detects 98% of random pure states but only 33% of full-rank mixed states. We specifically analyze the latter category, generating multiple instances of K states to quantify the number of WBMs required to detect a single entangled state, presenting results for varying δ. Notably, Dai et al. (2014) lacks numerical insights into the sample complexity and convergence rate of its proposed schemes.

9 Future Work And Conclusion

Batch entanglement detection, as discussed in this paper, is particularly useful for verifying the integrity of a batch of practically relevant entangled states, before use in applications like secure multi-channel quantum communication. We established a novel correspondence between the problem of batch entanglement detection and the Thresholding Bandit problem in stochastic Multi-Armed Bandits. We proposed the (m, K)-quantum Multi-Armed Bandit framework for entanglement detection. Focus of this framework is on identifying m entangled states out of K states, where m is potentially unknown. We apply this framework to two-qubit states using two key ingredients: a specialized set of six measurements for two-qubit states called Witness Basis Measurements (WBM) $\mathcal{E}$ and a separability criterion $S_{\mathcal{E}}$, which is based on the data obtained from these measurements and serves as the parameter that needs to be estimated. We present theoretical guarantees and numerical simulations to demonstrate how this parameter can be estimated quickly and accurately using MAB policies. First, we show that entangled states belonging to a class of parameterised two-qubit states $\mathcal{F}$ can be detected by measuring a subset of the six WBMs. With the knowledge of the WBM, we show that we can directly apply some suitable MAB policies. Second, for the same parameterised states, we present a routine for entanglement detection when the WBM is not known by enabling arbitrary sequential adaptation of the WBMs. We extend this to arbitrary two qubit quantum states and provide numerical results on the efficacy of using these measurements for detecting entanglement.

An exciting avenue for future research lies in identifying WBMs for higher-dimensional bipartite systems. The minimalistic tomographic scheme proposed in Zhu et al. (2010) significantly reduces the number of required witnesses for two-qutrits from 81 to just 11, demonstrating the potential for more efficient entanglement detection. Meanwhile, recent advancements in data-driven machine learning, particularly the use of SVMs to construct linear entanglement witnesses from local measurements (Greenwood et al., 2023), open new possibilities for tackling the (m, K)-quantum MAB problem. By leveraging these techniques, one could optimize the number of witnesses needed to reliably detect all m states.

Entanglement detection can be reframed as a membership problem, where a state belongs to a set if it exhibits a specific propertysuch as entanglement. This perspective aligns with the partition identification problem (Juneja & Krishnasamy, 2019), where the objective is to determine the partition to which a data point belongs based on a hyperplane structure. Extending this framework to the (m, K)-quantum MAB problem could pave the way for groundbreaking approaches to adaptive entanglement detection.

References

Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47:235–256, 05 2002. doi: 10.1023/A:1013689704352.

K Banaszek, M Cramer, and D Gross. Focus on quantum tomography. *New Journal of Physics*, 15 (12):125020, dec 2013. doi: 10.1088/1367-2630/15/12/125020. URL `https://dx.doi.org/10.1088/1367-2630/15/12/125020`.

Ingemar Bengtsson and Karol Zyczkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press, 2006.

Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters. Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels. *Phys. Rev. Lett.*, 70:1895–1899, Mar 1993. doi: 10.1103/PhysRevLett.70.1895.

K. Bharati. Qiskit workflow for entanglement detection, 2025. URL `https://github.com/borate267/TQE-Codes.git`.

Sergey Bravyi, Sarah Sheldon, Abhinav Kandala, David C. Mckay, and Jay M. Gambetta. Mitigating measurement errors in multiqubit experiments. *Physical Review A*, 103(4), April 2021. ISSN 2469-9934. doi: 10.1103/physreva.103.042605. URL `http://dx.doi.org/10.1103/PhysRevA.103.042605`.

Harry Buhrman, Richard Cleve, and Wim van Dam. Quantum entanglement and communication complexity. *SIAM Journal on Computing*, 30(6):1829–1841, 2001. doi: 10.1137/S0097539797324886.

Dariusz Chruciski and Gniewomir Sarbicki. Entanglement witnesses: construction, analysis and classification. *Journal of Physics A: Mathematical and Theoretical*, 47(48):483001, November 2014. ISSN 1751-8121. doi: 10.1088/1751-8113/47/48/483001. URL `http://dx.doi.org/10.1088/1751-8113/47/48/483001`.

Jibo Dai, Yink Loong Len, Yong Siah Teo, Berthold-Georg Englert, and Leonid A. Krivitsky. Experimental detection of entanglement with optimal-witness families. *Phys. Rev. Lett.*, 113:170402, Oct 2014. doi: 10.1103/PhysRevLett.113.170402. URL `https://link.aps.org/doi/10.1103/PhysRevLett.113.170402`.

Andrew C. Doherty, Pablo A. Parrilo, and Federico M. Spedalieri. Complete family of separability criteria. *Phys. Rev. A*, 69:022308, Feb 2004. doi: 10.1103/PhysRevA.69.022308. URL `https://link.aps.org/doi/10.1103/PhysRevA.69.022308`.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Pac bounds for multi-armed bandit and markov decision processes. In *International Conference on Computational Learning Theory*, pp. 255–270. Springer, 2002.

R. H. Farrell. Asymptotic Behavior of Expected Sample Size in Certain One Sided Tests. *The Annals of Mathematical Statistics*, 35(1):36 – 72, 1964. doi: 10.1214/aoms/1177703731. URL `https://doi.org/10.1214/aoms/1177703731`.

Steven T Flammia, David Gross, Yi-Kai Liu, and Jens Eisert. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New Journal of Physics*, 14(9):095022, sep 2012. doi: 10.1088/1367-2630/14/9/095022. URL `https://dx.doi.org/10.1088/1367-2630/14/9/095022`.

Daniel Stilck França, Fernando G.S L. Brandão, and Richard Kueng. Fast and Robust Quantum State Tomography from Few Basis Measurements. In Min-Hsiu Hsieh (ed.), *16th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2021)*, volume 197 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 7:1–7:13, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-198-6. doi: 10.4230/LIPIcs.TQC.2021.7. URL `https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.TQC.2021.7`.

Alexander C.B. Greenwood, Larry T.H. Wu, Eric Y. Zhu, Brian T. Kirby, and Li Qian. Machine-learning-derived entanglement witnesses. *Phys. Rev. Appl.*, 19:034058, Mar 2023. doi: 10.1103/PhysRevApplied.19.034058. URL `https://link.aps.org/doi/10.1103/PhysRevApplied.19.034058`.

Leonid Gurvits. Classical deterministic complexity of edmonds' problem and quantum entanglement, 2003. URL `https://arxiv.org/abs/quant-ph/0303055`.

M Guta, J Kahn, R Kueng, and J A Tropp. Fast state tomography with optimal error bounds. *Journal of Physics A: Mathematical and Theoretical*, 53(20):204001, apr 2020. doi: 10.1088/1751-8121/ab8111. URL `https://dx.doi.org/10.1088/1751-8121/ab8111`.

O. Gühne, P. Hyllus, O. Gittsovich, and J. Eisert. Covariance matrices and the separability problem. *Physical Review Letters*, 99(13), September 2007. ISSN 1079-7114. doi: 10.1103/physrevlett.99.130504. URL `http://dx.doi.org/10.1103/PhysRevLett.99.130504`.

Jeongwan Haah, Aram W. Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, pp. 11, 2017. ISSN 1557-9654. doi: 10.1109/tit.2017.2719044. URL http://dx.doi.org/10.1109/TIT.2017.2719044.

Chang Ho Hong, Jin O Heo, Gyong Luck Khym, Jongin Lim, Suc-Kyung Hong, and Hyung Jin Yang. N quantum channels are sufficient for multi-user quantum key distribution protocol between N users. *Optics Communications*, 283(12):2644–2646, 2010. ISSN 0030-4018. doi: https://doi.org/10.1016/j.optcom.2010.02.037. URL https://www.sciencedirect.com/science/article/pii/S0030401810001628.

Micha Horodecki, Pawe Horodecki, and Ryszard Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A*, 223(1):1–8, 1996a. ISSN 0375-9601. doi: https://doi.org/10.1016/S0375-9601(96)00706-2. URL https://www.sciencedirect.com/science/article/pii/S0375960196007062.

Micha Horodecki, Pawe Horodecki, and Ryszard Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A*, 223(1):1–8, 1996b. ISSN 0375-9601. doi: https://doi.org/10.1016/S0375-9601(96)00706-2. URL https://www.sciencedirect.com/science/article/pii/S0375960196007062.

Pawel Horodecki. Separability criterion and inseparable mixed states with positive partial transposition. *Physics Letters A*, 232(5):333–339, 1997. ISSN 0375-9601. doi: https://doi.org/10.1016/S0375-9601(97)00416-7. URL https://www.sciencedirect.com/science/article/pii/S0375960197004167.

Ryszard Horodecki, Pawe Horodecki, Micha Horodecki, and Karol Horodecki. Quantum entanglement. *Reviews of Modern Physics*, 81(2):865–942, June 2009. ISSN 0034-6861, 1539-0756. doi: 10.1103/RevModPhys.81.865.

Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil' ucb : An optimal exploration algorithm for multi-armed bandits. In Maria Florina Balcan, Vitaly Feldman, and Csaba Szepesvári (eds.), *Proceedings of The 27th Conference on Learning Theory*, volume 35 of *Proceedings of Machine Learning Research*, pp. 423–439, Barcelona, Spain, 13–15 Jun 2014. PMLR. URL https://proceedings.mlr.press/v35/jamieson14.html.

Sandeep Juneja and Subhashini Krishnasamy. Sample complexity of partition identification using multi-armed bandits, 2019. URL https://arxiv.org/abs/1811.05654.

Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. Pac subset selection in stochastic multi-armed bandits. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, ICML'12, pp. 227234, Madison, WI, USA, 2012. Omnipress. ISBN 9781450312851.

Hideaki Kano, Junya Honda, Kentaro Sakamaki, Kentaro Matsuura, Atsuyoshi Nakamura, and Masashi Sugiyama. Good arm identification via bandit feedback, 2018.

Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In *Proceedings of the 30th International Conference on Machine Learning*, pp. 1238–1246, 2013.

Dominik Koutný, Libor Motka, Zdenk Hradil, Jaroslav eháek, and Luis L. Sánchez-Soto. Neural-network quantum state tomography. *Physical Review A*, 106(1), July 2022. ISSN 2469-9934. doi: 10.1103/physreva.106.012409. URL http://dx.doi.org/10.1103/PhysRevA.106.012409.

Richard Kueng, Holger Rauhut, and Ulrich Terstiege. Low rank matrix recovery from rank one measurements. *Applied and Computational Harmonic Analysis*, 42(1):88–116, 2017. ISSN 1063-5203. doi: https://doi.org/10.1016/j.acha.2015.07.007. URL https://www.sciencedirect.com/science/article/pii/S1063520315001037.

M. Lewenstein, B. Kraus, J. I. Cirac, and P. Horodecki. Optimization of entanglement witnesses. *Phys. Rev. A*, 62:052310, Oct 2000a. doi: 10.1103/PhysRevA.62.052310. URL https://link.aps.org/doi/10.1103/PhysRevA.62.052310.

M. Lewenstein, B. Kraus, J. I. Cirac, and P. Horodecki. Optimization of entanglement witnesses. *Phys. Rev. A*, 62:052310, Oct 2000b. doi: 10.1103/PhysRevA.62.052310. URL https://link.aps.org/doi/10.1103/PhysRevA.62.052310.

Dawei Lu, Tao Xin, Nengkun Yu, Zhengfeng Ji, Jianxin Chen, Guilu Long, Jonathan Baugh, Xinhua Peng, Bei Zeng, and Raymond Laflamme. Tomography is necessary for universal entanglement detection with single-copy observables. *Phys. Rev. Lett.*, 116:230501, Jun 2016. doi: 10.1103/PhysRevLett.116.230501. URL https://link.aps.org/doi/10.1103/PhysRevLett.116.230501.

Josep Lumbreras, Erkka Haapasalo, and Marco Tomamichel. Multi-armed quantum bandits: Exploration versus exploitation when learning properties of quantum states. *Quantum*, 6:749, June 2022. ISSN 2521-327X. doi: 10.22331/q-2022-06-29-749. URL http://dx.doi.org/10.22331/q-2022-06-29-749.

Shie Mannor and John N. Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. *J. Mach. Learn. Res.*, 5:623648, dec 2004. ISSN 1532-4435.

Ryan O'Donnell and John Wright. Efficient quantum tomography. *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, 2015a. URL https://api.semanticscholar.org/CorpusID:769062.

Ryan O'Donnell and John Wright. Efficient quantum tomography ii. *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, 2015b. URL https://api.semanticscholar.org/CorpusID:5245926.

Asher Peres. Separability criterion for density matrices. *Phys. Rev. Lett.*, 77:1413–1415, Aug 1996. doi: 10.1103/PhysRevLett.77.1413. URL https://link.aps.org/doi/10.1103/PhysRevLett.77.1413.

Mauro B Pozzobom and Jonas Maziero. Preparing tunable bell-diagonal states on a quantum computer. *Quantum Information Processing*, 18(5):142, 2019.

Qiskit Community. Readout error mitigation, 2024. URL qiskit-community.github.io/qiskit-experiments.

Yihui Quek, Stanislav Fort, and Hui Khoon Ng. Adaptive quantum state tomography with neural networks, 2018. URL https://arxiv.org/abs/1812.06693.

Elias Riedel Gårding, Nicolas Schwaller, Chun Lam Chan, Su Yeon Chang, Samuel Bosch, Frederic Gessler, Willy Robert Laborde, Javier Naya Hernandez, Xinyu Si, Marc-André Dupertuis, et al. Bell diagonal and werner state generation: Entanglement, non-locality, steering and discord on the ibm quantum computer. *Entropy*, 23(7):797, 2021.

Oliver Rudolph. A separability criterion for density operators. *Journal of Physics A: Mathematical and General*, 33(21):39513955, May 2000. ISSN 1361-6447. doi: 10.1088/0305-4470/33/21/308. URL http://dx.doi.org/10.1088/0305-4470/33/21/308.

Tobias Schmale, Moritz Reh, and Martin Gärttner. Efficient quantum state tomography with convolutional neural networks. *NPJ Quantum Information*, 8(1), September 2022. ISSN 2056-6387. doi: 10.1038/s41534-022-00621-4. URL http://dx.doi.org/10.1038/s41534-022-00621-4.

Vikesh Siddhu. Maximum a posteriori probability estimates for quantum tomography. *Physical Review A*, 99(1), January 2019. ISSN 2469-9934. doi: 10.1103/physreva.99.012342. URL http://dx.doi.org/10.1103/PhysRevA.99.012342.

Yong Siah Teo, Huangjun Zhu, Berthold-Georg Englert, Jaroslav Řeháček, and Zden ěk Hradil. Quantum-state reconstruction by maximizing likelihood and entropy. *Phys. Rev. Lett.*, 107:020404, Jul 2011. doi: 10.1103/PhysRevLett.107.020404. URL https://link.aps.org/doi/10.1103/PhysRevLett.107.020404.

Barbara M. Terhal. Bell inequalities and the separability criterion. *Physics Letters A*, 271(5):319–326, 2000. ISSN 0375-9601. doi: https://doi.org/10.1016/S0375-9601(00)00401-1. URL https://www.sciencedirect.com/science/article/pii/S0375960100004011.

Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko, and Giuseppe Carleo. Neural-network quantum state tomography. *Nature Physics*, 14(5):447450, February 2018. ISSN 1745-2481. doi: 10.1038/s41567-018-0048-5. URL http://dx.doi.org/10.1038/s41567-018-0048-5.

Tzu-Hsien Tsai, Yun-Da Tsai, and Shou-De Lin. lil'hdoc: An algorithm for good arm identification under small threshold gap, 2024. URL https://arxiv.org/abs/2401.15879.

Jinzhao Wang, Volkher B. Scholz, and Renato Renner. Confidence polytopes in quantum state tomography. *Physical Review Letters*, 122(19), May 2019. ISSN 1079-7114. doi: 10.1103/physrevlett.122.190401. URL http://dx.doi.org/10.1103/PhysRevLett.122.190401.

Huangjun Zhu, Yong Siah Teo, and Berthold-Georg Englert. Minimal tomography with entanglement witnesses. *Phys. Rev. A*, 81:052339, May 2010. doi: 10.1103/PhysRevA.81.052339. URL https://link.aps.org/doi/10.1103/PhysRevA.81.052339.

Karol Zyczkowski and Hans-Jürgen Sommers. Induced measures in the space of mixed quantum states. *Journal of Physics A: Mathematical and General*, 34(35):71117125, August 2001. ISSN 1361-6447. doi: 10.1088/0305-4470/34/35/335. URL http://dx.doi.org/10.1088/0305-4470/34/35/335.

A Appendix

The following lemma is useful for some calculations.

Lemma 13 *For $t \geq 1, c > 0, \varepsilon \in (0,1), 0 < w \leq 1$,*

$$\frac{1}{t}\log\left(\frac{\log\left((1+\varepsilon)t\right)}{w}\right) \geq c \implies t \leq \frac{1}{c}\log\left(\frac{2\log\left(\frac{(1+\varepsilon)}{cw}\right)}{w}\right). \tag{18}$$

A.1 Proof for Section 5.1

A.1.1 Proof of Lemma 6

Proof: Let $\mathcal{B}$ denote the "good" event that at any time $t > 0$ and for all arms $i \in [K]$, the true value $S_{\mathcal{E}}(\rho_i)$ is well concentrated around its estimate $\hat{S}_{i,N_i(t)}$.

$$\mathcal{B} \coloneqq \bigcup_{i=1}^{K}\bigcup_{t=1}^{\infty}\left\{|\hat{S}_{i,N_i(t)} - S_i| \leq U\left(N_i(t), \frac{\delta}{c_\varepsilon K}\right)\right\}$$

From Lemma 5 and by applying the union bound, we get that

$$\mathbb{P}\left[\mathcal{B}\right] \geq 1 - c_\varepsilon K\left(\frac{\delta}{c_\varepsilon K}\right)^{1+\varepsilon} \geq 1 - \delta \tag{19}$$

where Eq. 19 holds because $\varepsilon \in (0,1)$ and $c_\varepsilon \geq 1$. □

A.1.2 Proof of Theorem 7

Proof: Recall that the threshold $\zeta = 0$ and problem instance $\boldsymbol{S}_{\mathcal{E}}$ is such that $S_{\mathcal{E}}(\rho_1) \geq S_{\mathcal{E}}(\rho_2) \geq S_{\mathcal{E}}(\rho_3) \ldots > S_{\mathcal{E}}(\rho_{K-1}) > 0 > S_{\mathcal{E}}(\rho_K)$. Let us consider the case that the event $\mathcal{B}$ described in Lemma 6 holds. As outlined in Algorithm 1, the arm $i^\star$ will be dropped from the active set Ω if $\mathrm{LCB}_{i^\star}(t) > 0$. That is,

$$\begin{aligned}
&\hat{S}_{i^\star,N_{i^\star}(t)} - U\left(N_{i^\star}(t), \frac{\delta}{c_\varepsilon K}\right) > 0 \\
&\hat{S}_{i^\star,N_{i^\star}(t)} - |\hat{S}_{i^\star,N^\star_i(t)} - S_{i^\star}| > 0 \\
&\implies S_{i^\star} > 0
\end{aligned}$$

This contradicts the assumption about the problem instance $\boldsymbol{S}$ because $S_{i^\star} = S_{\mathcal{E}}(\rho_K) < 0$ and so, the arm $i^\star$ will not be dropped from the active set Ω as long as event $\mathcal{B}$ holds. □

A.1.3 Proof of Theorem 8

Proof: Let us consider the case where $\mathcal{B}$ holds. By the elimination rule of Algorithm 1, an arm i is removed from the active set Ω if $\text{LCB}_i(t) > 0$. We have that,

$$\begin{aligned}
&\hat{S}_{i,N_i(t)} - U\left(N_i(t), \frac{\delta}{c_\varepsilon K}\right) \geq \zeta \\
&\hat{S}_{i,N_i(t)} - S_i + \Delta_i \geq U\left(N_i(t), \frac{\delta}{c_\varepsilon K}\right) \\
&\implies \Delta_i \geq 2U\left(N_i(t), \frac{\delta}{c_\varepsilon K}\right)
\end{aligned} \tag{20}$$

Let us denote N_i to be the number of samples of arm i, that is, $N_i = \inf\{t : U\left(N_i(t), \frac{\delta}{c_\varepsilon K}\right) \leq \frac{\Delta_i}{2}\}$. The minimum value of N_i can be obtained by solving,

$$\begin{aligned}
&U\left(N_i, \frac{\delta}{c_\varepsilon K}\right) = \frac{\Delta_i}{2} \\
&(1+\sqrt{\varepsilon})\sqrt{\frac{2(1+\varepsilon)}{N_i}\log\left(\frac{\log\left((1+\varepsilon)N_i\right)}{\delta/c_\varepsilon K}\right)} = \frac{\Delta_i}{2} \\
&\frac{1}{N_i}\log\left(\frac{\log\left((1+\varepsilon)N_i\right)}{\delta/c_\varepsilon K}\right) = \frac{\Delta_i^2}{8(1+\varepsilon)(1+\sqrt{\varepsilon})^2}
\end{aligned} \tag{21}$$

From Lemma 13, we get that,

$$N_i = \frac{8(1+\varepsilon)(1+\sqrt{\varepsilon})^2}{\Delta_i^2}\log\left(\frac{2c_\varepsilon K \log\left(\frac{8c_\varepsilon(1+\varepsilon)^2(1+\sqrt{\varepsilon})^2}{\delta}\frac{K}{\Delta_i^2}\right)}{\delta}\right) \tag{22}$$

Thus, the total number of samples required to identify the arm $i^\star$ with a probability of at least $1-\delta$ is $N \leq \sum_{i=1}^{K} N_i$. □

A.2 Proof for Section 5.2

A.2.1 Proof of Lemma 9

Proof: Firstly, we show that Algorithm 2 is (λ, δ)-PAC for arbitrary $\lambda \in [K]$. In the case where there are arms greater than or equal to λ, we show that $\mathbb{P}\left[\{\hat{m} < \lambda\} \cup \bigcup_{i \in \mathcal{A}_{\text{ent}}}\{S_i < \zeta\}\right] \leq \delta$ where $\hat{m}$ is the number of good arms identified by the agent. Since we are now considering the case when $m \geq \lambda$, the event $\{\hat{m} < \lambda\}$ implies that at least one good arm $j \in [m]$ is identified as a bad arm by the agent. That is, for some $j \in [m]$ and $t \in \mathbb{N}$, the upper confidence bound $\hat{S}_{j,N_j(t)} + U\left(N_j(t), \frac{\delta}{c_\varepsilon K}\right) < \zeta$. Thus, we have that,

$$\begin{aligned}
\mathbb{P}\left[\hat{m} < \lambda\right] &\leq \sum_{j\in[m]} \mathbb{P}\left[\bigcup_{t\in\mathbb{N}}\{\hat{S}_{j,N_j(t)} + U\left(N_j(t), \frac{\delta}{c_\varepsilon K}\right) < \zeta\}\right] \\
&\leq \sum_{j\in[m]} c_\varepsilon\left(\frac{\delta}{c_\varepsilon K}\right)^{1+\varepsilon} \qquad \text{(By Lemma 5)} \\
&\leq m c_\varepsilon\left(\frac{\delta}{c_\varepsilon K}\right)
\end{aligned} \tag{23}$$

The event $\bigcup_{i\in\{\hat{X}_1,\hat{X}_2,\ldots\hat{X}_\lambda\}}\{\mu_i < \zeta\}$ considers all those outcomes where a bad arm is identified to be a good one. Thus, for some bad arm $j \in \{\hat{X}_1, \hat{X}_2, \ldots \hat{X}_{\hat{m}}\}$ such that $j \in [K] \setminus [m]$, we have,

$$\begin{aligned}
&\mathbb{P}\left[\bigcup_{i\in\{\hat{X}_1,\hat{X}_2,\dots\hat{X}_\lambda\}}\{S_i<\zeta\}\right]\\
&\leq \sum_{j\in[K]\setminus[m]}\mathbb{P}\left[\bigcup_{t\in\mathbb{N}}\{\hat{S}_{j,N_j(t)}-U\left(N_j(t),\frac{\delta}{c_\varepsilon K}\right)>\zeta\}\right]\\
&\leq (K-m)c_\varepsilon\left(\frac{\delta}{c_\varepsilon K}\right) \qquad (24)
\end{aligned}$$

Thus, putting Eq. 23 and Eq. 24 together, we get that $\mathbb{P}\left[\{\hat{m}<\lambda\}\cup\bigcup_{i\in\{\hat{X}_1,\hat{X}_2,\dots\hat{X}_{\hat{m}}\}}\{\mu_i<\zeta\}\right]\leq\delta$. Next, we consider the case when the number of good arms m is less than λ and show that $\mathbb{P}\left[\hat{m}\geq\lambda\right]\leq\delta$. Since there are at most λ good arms, the event $\{\hat{m}>\lambda\}$ implies that one of the output arms $j\in\{\hat{X}_1,\hat{X}_2,\dots\hat{X}_\lambda\}$ is such that there exists some index j such that $\hat{X}_j$ is a bad arm. Thus, we have that,

$$\begin{aligned}
\mathbb{P}\left[\hat{m}\geq\lambda\right]&\leq\sum_{j\in[K]\setminus[m]}\mathbb{P}[j\in\{\hat{X}_1,\hat{X}_2,\dots\hat{X}_\lambda\}]\\
&\leq(K-m)c_\varepsilon\left(\frac{\delta}{c_\varepsilon K}\right)^{1+\varepsilon}\\
&\leq\frac{K-m}{K}c_\varepsilon\left(\frac{\delta}{c_\varepsilon}\right)\\
&\leq\delta \qquad (25)
\end{aligned}$$

We see that the algorithm is (λ,δ)-PAC for all such $\lambda\in[K]$, thereby giving us that the algorithm is δ-PAC. □

A.2.2 Proof of Theorem 10

Proof: Recall that the threshold $\zeta=0$ and problem instance $\boldsymbol{S}_\mathcal{E}$ is such that $S_\mathcal{E}(\rho_1)\geq S_\mathcal{E}(\rho_2)\dots>S_\mathcal{E}(\rho_{K-m})>0>S_\mathcal{E}(\rho_{K-m+1})\dots>S_\mathcal{E}(\rho_K)$, with m being unknown. Let us consider the case that the event $\mathcal{B}$ described in Lemma 6 holds. As outlined in Algorithm 2, an arm i will be dropped if $\mathrm{LCB}_i(t)>0$. That is,

$$\begin{aligned}
&\hat{S}_{i,N_i(t)}-U\left(N_i(t),\frac{\delta}{c_\varepsilon K}\right)>0\\
&\hat{S}_{i,N_i(t)}-|\hat{S}_{i,N_i(t)}-S_i|>0\\
&\implies S_i>0
\end{aligned}$$

Thus, as long as event $\mathcal{B}$ holds, all the arms that have $S_\mathcal{E}<0$ will not dropped. Thus the lil'HDoC algorithm identifies all the arms correctly. □

A.3 Integrating Error Mitigation in MAB Algorithms for Batch Entanglement Detection

In the MAB-based workflow for entanglement detection described in Section 6, one state is measured at every time instant as dictated by the sampling rule, and the statisticsnamely, the estimates of f_1, f_2, f_3, and $f_4$are updated as new measurement outcomes are obtained. These estimates are susceptible to measurement errors, particularly readout errors, which induce inaccuracies in the measurement counts. To improve the accuracy of the estimates, we characterize such errors and wish to mitigate them (Qiskit Community, 2024). To this end, we carry out a preliminary investigation by incorporating a procedure for (a) error mitigation and (b) including error mitigation in the MAB routine, and study the impact of error mitigation on the overall copy complexity of batch entanglement detection.

A.3.1 Procedure for Error Mitigation

In Fig. 3, we apply a unitary transformation to WBM $\mathcal{E}_1$ and $\mathcal{E}_2$ to measure the state of the system ρ in the computational (Pauli Z) basis. Consequently, we obtain expectation values of the diagonal Pauli operators ZZ, ZI, and IZ. The estimates of f_1, f_2, f_3, and f_4 are linear combinations of these expectation values.

$$\begin{aligned} f_1 &= 0.25\,(1 + \langle IZ\rangle_\rho + \langle ZI\rangle_\rho + \langle ZZ\rangle_\rho) \\ f_2 &= 0.25\,(1 - \langle IZ\rangle_\rho + \langle ZI\rangle_\rho - \langle ZZ\rangle_\rho) \\ f_3 &= 0.25\,(1 + \langle IZ\rangle_\rho - \langle ZI\rangle_\rho - \langle ZZ\rangle_\rho) \\ f_4 &= 0.25\,(1 - \langle IZ\rangle_\rho - \langle ZI\rangle_\rho + \langle ZZ\rangle_\rho)\,. \end{aligned} \tag{26}$$

Thus, it is essential to obtain precise expectation values for the diagonal Pauli operators to improve the accuracy of our estimates. To do this, we use a LocalReadOut scheme from IBM's Qiskit Experiments library (Bravyi et al., 2021). In this scheme we characterize the readout errors of physical qubits on the **FakeBrisbane** backend. These errors are assumed to be local in the sense they are independent across qubits. Readout error mitigation uses a mitigator object (matrix) computed from an assignment matrix A, where each element $A_{i,j}$ represents the probability of observing outcome i when the true outcome is j. By applying this mitigator to unmitigated measurement counts, we refine our estimates by obtaining more accurate expectation values for ZZ, ZI, and IZ.

A.3.2 How and where does it fit in the MAB Routine?

In each round of the MAB policy, based on an Upper Confidence Bound (UCB) score, the sampling rule selects a quantum state to measure. Since only a single-shot measurement is performed per round, the error mitigation procedure described in Section A.3.1 is applied after a state has been measured several times. To illustrate this process, consider a specific round $t = F$, where state ρ_1 has previously been measured $T^\star$ times. The unmitigated measurement *counts* for the four possibrle outcomes are denoted as $F_1^{\text{um}}, F_2^{\text{um}}, F_3^{\text{um}}$, and F_4^{um}. The empirical frequencies of these outcomes are given by,

$$\hat{f}_i^{\text{um}}(F) = \frac{F_i^{\text{um}}}{T^\star}, \quad i \in [4]. \tag{27}$$

At this point, we invoke the error mitigation routine, supplying it with the unmitigated counts $\{F_i^{\text{um}}\}$ as input. The mitigation routine corrects for readout errors and returns mitigated expectation values of the diagonal Pauli observables, yielding mitigated estimates $\hat{f}_i^{\text{m}}(F)$. With post-processing adjustments to correct for decimal rounding errors, the corresponding mitigated measurement counts,

$$F_i^{\text{m}} = \hat{f}_i^{\text{m}}(F) \times T^\star, \quad i \in [4]. \tag{28}$$

We propose a **nested mitigative process** where the MAB algorithm invokes the error mitigation routine once every F measurement shots per state and uses the mitigated values in subsequent shots. For instance, at $t = F$, the routine produces mitigated estimates $\hat{f}_i^{\text{m}}(F)$ from which we obtain mitigated counts. Future measurement outcomes update on these mitigated counts. At $t = 2F$, the routine takes input these new counts and outputs a new set of mitigated estimates $\hat{f}_i^{\text{m}}(2F)$. This creates a nested-mitigation cycle, where each round of mitigation refines the previous one.

We conduct an empirical study to assess the impact of error mitigation on the average copy complexity of the MAB algorithm. Mitigation is invoked once every F rounds, where F ranges from 50 to 10,000 in steps of 50. Here, smaller F values correspond to high-frequency mitigation and larger values indicate lower-frequency mitigation. For the problem instance described in Section 6, with $\delta \in (0, 1)$ and range of F, we execute Algorithm 3 on FakeBrisbane, averaging the copy complexity at stoppage over 20 runs. The percentage of error mitigation is quantified as the relative reduction in copy complexity compared to the case without mitigation. To ensure the algorithm correctly identifies the entangled states, we employ an error indicator that verifies whether its error remains within the prescribed threshold δ. Using this framework, we generate the heatmap in Fig. 7, which visualizes the percentage reduction in copy complexity due to error mitigation. Notably, the white regions indicate cases where the algorithm converged in finite time but failed to correctly

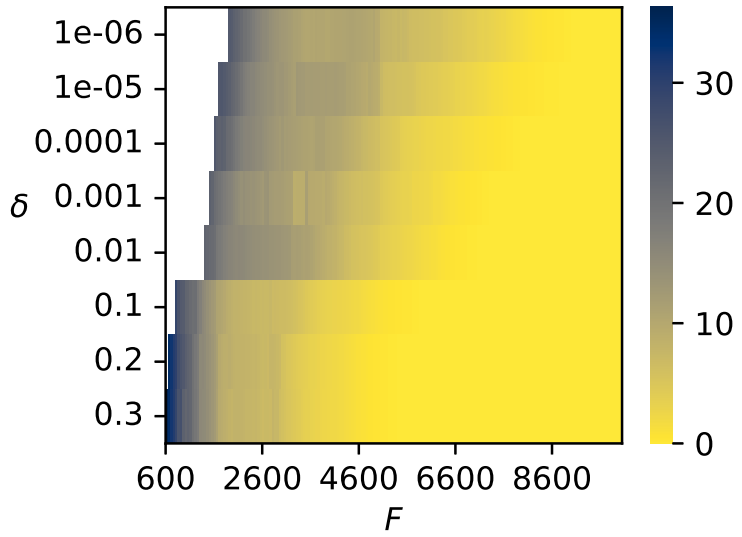


Figure 7: Heatmap of percentage error mitigation on FakeBrisbane backend for $\delta \in (0, 1)$ and various mitigation frequencies

identify the entangled states. We observe and report the following inferences from Fig. 7. First, the effect of mitigation is δ-dependent. For larger values of δ, the mitigation effect starts only as early as ($F = 600$) and stabilizes faster ($F \sim 4000$). In contrast, for smaller values of δ, the effect of mitigation is prominent only mid-range and stabilizes at $F \sim 7000$. Second, for $F < 600$ and smaller values of δ, the algorithm fails to detect the correct set of states under the prescribed δ. This can be attributed to over-mitigation which could potentially lead to random fluctuations in the estimates. Third, the observed stabilization zone (yellow) across values of δ suggests a critical threshold for F beyond which reducing mitigation frequency (increasing the value of F) no longer reduces errors. It remains an open question to fully understand and optimize for the use of error-mitigation and integrate them with MAB strategies.