Multiple-shot and unambiguous discrimination of von Neumann measurements

Zbigniew Puchała[1,2], Łukasz Pawela[1], Aleksandra Krawiec[1,3], Ryszard Kukulski[1,4], and Michał Oszmaniec[5,6]

[1]Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, ul. Bałtycka 5, 44-100 Gliwice, Poland

[2]Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University, ul. Łojasiewicza 11, 30-348 Kraków, Poland

[3]Institute of Mathematics, Silesian University of Technology, ul. Kaszubska 23, 44-100 Gliwice, Poland

[4]Institute of Mathematics, University of Silesia, ul. Bankowa 14, 40-007 Katowice, Poland

[5]Institute of Theoretical Physics and Astrophysics, National Quantum Information Centre, Faculty of Mathematics, Physics and Informatics, University of Gdańsk, ul. Wita Stwosza 57, 80-308 Gdańsk, Poland

[6]Center for Theoretical Physics, Polish Academy of Sciences
Al. Lotników 32/46, 02-668 Warszawa, Poland

We present an in-depth study of the problem of multiple-shot discrimination of von Neumann measurements in finite-dimensional Hilbert spaces. Specifically, we consider two scenarios: minimum error and unambiguous discrimination. In the case of minimum error discrimination, we focus on discrimination of measurements with the assistance of entanglement. We provide an alternative proof of the fact that all pairs of distinct von Neumann measurements can be distinguished perfectly (i.e. with the unit success probability) using only a finite number of queries. Moreover, we analytically find the minimal number of queries needed for perfect discrimination. We also show that in this scenario querying the measurements *in parallel* gives the optimal strategy, and hence any possible adaptive methods do not offer any advantage over the parallel scheme. In the unambiguous discrimination scenario, we give the general expressions for the optimal discrimination probabilities with and without the assistance of entanglement. Finally, we show that typical pairs of Haar-random von Neumann measurements can be perfectly distinguished with only two queries.

Łukasz Pawela: lpawela@iitis.pl

1 Introduction

With the recent technological progress, quantum information science is not merely a collection of purely theoretical ideas anymore. Indeed, quantum protocols of increasing degree of complexity are currently being implemented on more and more complicated quantum devices [1, 2] and are expected to soon yield practical solutions to some real-world problems [3]. This situation motivates the need for certification and benchmarking of various building-blocks of quantum devices [4–6] (see [7] for a recent review). Discrimination or quantum hypothesis testing constitute one of the paradigms for assessing the quality of parts of quantum protocols [8–12]. In this work, we present a comprehensive study of various scenarios of discrimination of von Neumann measurements on a finite-dimensional Hilbert space. Here, by von Neumann measurements, we understand fine-grained projective measurements. Such measurements are vital for most of the protocols appearing in quantum information and quantum computing. In fact, even the most general quantum measurements are typically implemented using projective measurements performed on enlarged Hilbert space via the so-called Naimark construction [13]. Due to imperfections present in currently available quantum devices [3] and increased complexity associated with Naimark construction, standard projective measurements, such as the computational basis measurement, are the most common measurements implemented currently on quantum processors. It is important to point out that despite their relative simplicity, the actual implementation of projective measurements on quantum hardware is imperfect. For example, in a recent demonstration of quantum computational supremacy (advantage) by the collaboration of Google and UCBS [14] the researchers reported single-qubit measurement errors that are of order of a few percents. This motivates the interest in certification strategies tailored specifically to von-Neumann measurements.

The general problem of quantum channel discrimination has attracted a lot of attention in recent years. One of the first results was the study of discrimination of unitary operators [15, 16]. Later, this has been extended to various settings, such as multipartite unitary operations [17] and the case of discrimination among more than two unitary channels [18]. In the work [19] the authors formulated necessary and sufficient conditions under which quantum channels can be perfectly discriminated. Further works investigated the adaptive [20–23] and parallel [24] schemes for discrimination of channels. Finally, some asymptotic results on discrimination of typical quantum channels in large dimensions were obtained in [25]. Discrimination of quantum measurements, being a subset of quantum channels, is thus of particular interest. Some of the earliest results on this topic involve condition on perfect discrimination of two measurements [26–30]

We are interested in the following problem. Imagine we have an unknown device hidden in a black box. We know it performs one of the two possible von Neumann measurements, either $\mathcal{P}_1$ or $\mathcal{P}_2$. Generally, whenever a quantum state is sent through the box, the box produces, with probabilities predicted by quantum mechanics, classical labels corresponding to the measurement outcomes. Our goal is to find schemes that attain the optimal success probability for discrimination of measurements. The results contained in this work concern the following two scenarios:

Minimum error discrimination— In this setting, we are allowed to use the black box containing von Neumann measurement many times. Furthermore, we can prepare any input state with an arbitrarily large quantum memory (i.e., we can use ancillas of arbitrarily large dimension), and we can perform any channels between usages of the black box. This allows us to implement both parallel (see Fig. 1) as well as adaptive discrimination strategies (see Fig. 2). We focus on the case of entanglement-assisted discrimination. Our main finding is that in the multiple-shot scenario, adaptive strategies do not offer any advantage over parallel queries. Moreover, we derive an explicit dependence of the diamond norm distance in the multiple-shot scenario as a function of the diamond norm in the single-shot setting. As a consequence, we recover a known result [26] stating that given sufficiently many queries, every pair of different von Neumann measurements can be distinguished perfectly (i.e. with zero error probability).

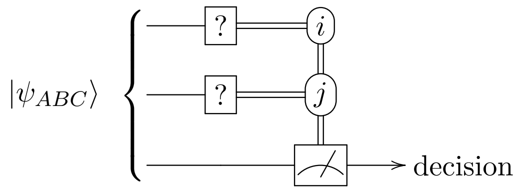

Figure 1: A schematic representation of parallel discrimination scheme of quantum measurements that uses two applications of a measurement.

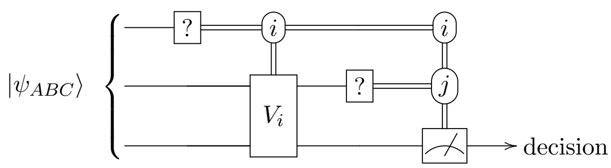

Figure 2: A schematic representation of adaptive discrimination scheme of quantum measurements that uses two applications of a measurement. In this case adaptive scheme amounts to application of the unitary channel Φ_{V_i} which depends on the result i of the first measurement.

Unambiguous discrimination— This scenario is an analogue to the well-known scheme of unambiguous state discrimination [31]. Namely, for every query to the black box, the decision procedure outputs $\mathcal{P}_1$, $\mathcal{P}_2$, or the inconclusive answer. The latter means that the user cannot decide which measurement was contained in the black box. Importantly, we require that the procedure cannot wrongly identify the measurement (see Fig. 3). Our main contribution to this problem is the derivation of the general schemes, which attain the optimal success probability both with and without the assistance of entanglement. Interestingly, we find that optimal success probability p_u for unambiguous discrimination of projective measurements $\mathcal{P}_1, \mathcal{P}_2$ with the assistance of entanglement is functionally related to the diamond norm distance $\|\mathcal{P}_1 - \mathcal{P}_2\|_\diamond$, which quantifies distinguishability of $\mathcal{P}_1, \mathcal{P}_2$ with the assistance of entanglement but without requiring unambiguous results. The specific formula is given by

$$p_u = 1 - \sqrt{1 - \frac{1}{4}\|\mathcal{P}_1 - \mathcal{P}_2\|_\diamond^2} \tag{1}$$

and its geometric interpretation is presented in Fig. 4. Finally, we also present simple formulas for the optimal discrimination probability of von Neumann measurements for qubits.

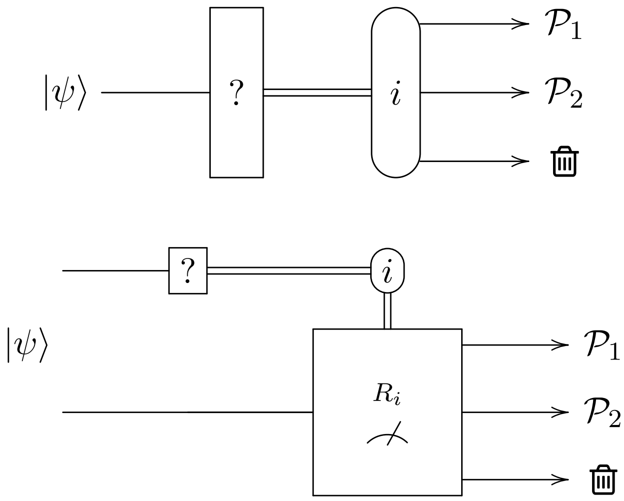

Figure 3: A schematic representation of the setting of unambiguous measurement discrimination scheme. The figure on the left shows an umbiguous dscrimination without the assistance of entanglement while the figure on right shows entanglement-assisted unambiguous discrimination.

Relation with classical channels— Let us contrast our results with transformations mapping probability distributions to probability distributions, that is classical channels. In this setting, we know that if such channels cannot be perfectly distinguished in one shot, they cannot be distinguished perfectly in any finite number of uses. What we can do is to study the asymptotic behavior of error probability when the number of applications of the channel tends to infinity. The error probability, formulated in the language of the hypothesis testing of two distributions, decays exponentially, and the optimal exponential error rate, depending on a formulation, is given by the Stein bound, the Chernoff bound, the Hoeffding bound, and the Han-Kobayashi bound, see [32] and references therein. In the case of distinct von Neumann measurements and entanglement-assisted discrimination, we show that one can perform perfect discrimination with the use of a finite number of queries. Therefore, in contrary to the classical channels, we do not have to consider the exponential error rate, as the error probability drops to zero in a finite number of tries.

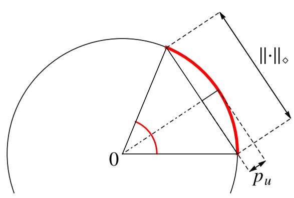

Figure 4: Geometric interpretation of the relationship between the diamond norm $\|\mathcal{P}_1 - \mathcal{P}_2\|_\diamond$ and the probability of unambiguous discrimination $p_u(\mathcal{P}_1, \mathcal{P}_2)$ presented on the unit circle.

This work is organized as follows. In Section 2, we give a survey of the main concepts and notation used throughout this work (including the basic background on discrimination of quantum channels and measurements). Then, in Section 3, we present our results for the scenario of multiple-shot minimum error measurement discrimination. Theorem 1 therein expresses the minimum error in the parallel discrimination scheme as a function of the minimum error in the single-shot discrimination scheme. Theorem 3 gives the upper bound on the probability of correct discrimination of a generic pair of von Neumann measurements coming from the Haar distribution. The following Section 4 contains the results concerning the unambiguous discrimination of quantum measurements. The main result of this section is formulated as Theorem 4, which states the optimal success probability of unambiguous discrimination with the assistance of entanglement. Lastly, in Section 5 we summarize our results and give some directions for future research.

2 Preliminaries and main concepts

By $\mathcal{D}(\mathbb{C}^d)$ we will denote the set of quantum states on $\mathbb{C}^d$. Let $\mathrm{L}(\mathbb{C}^{d_1}, \mathbb{C}^{d_2})$ denote the set of all linear operators acting from $\mathbb{C}^{d_1}$ to $\mathbb{C}^{d_2}$. For brevity we will put $\mathrm{L}(\mathbb{C}^d, \mathbb{C}^d) \equiv \mathrm{L}(\mathbb{C}^d)$. A quantum channel is a linear mapping $\Phi : \mathrm{L}(\mathbb{C}^{d_1}) \to \mathrm{L}(\mathbb{C}^{d_2})$ which is completely positive and tracepreserving. The former means that for every $L(\mathbb{C}^{d_1}) \otimes L(\mathbb{C}^s) \ni \rho \geq 0$ we have $L(C^{d_2}) \otimes L(\mathbb{C}^s) \ni (\Phi \otimes \mathbb{1})(\rho) \geq 0$, while the latter means $\mathrm{Tr}(\Phi(X)) = \mathrm{Tr}\, X$ for every $X \in L(\mathbb{C}^{d_1})$. A set of generalized measurements (POVMs) $\mathbb{C}^d$ will be denoted by $\mathrm{POVM}(\mathbb{C}^d)$. A general quantum measurement $\mathcal{M}$ on $\mathbb{C}^d$ is a tuple of positive semidefinite operators[1] on $\mathbb{C}^d$ that add up to identity on $\mathbb{C}^d$ i.e. $\mathcal{M} = (M_1, \ldots, M_n)$ with $M_i \geq 0$ and $\sum_i M_i = \mathbb{1}$. If a quantum state σ is measured by a measurement $\mathcal{M}$, then the outcome i is obtained with the probability $p(i|\sigma, \mathcal{M}) = \mathrm{tr}(\sigma M_i)$ (Born rule). Therefore, a quantum measurement $\mathcal{M}$ can be uniquely identified with a quantum channel

$$\Psi_{\mathcal{M}}(\sigma) = \sum_{i=1}^{n} \mathrm{tr}(M_i \sigma)|i\rangle\langle i|, \tag{2}$$

where states $|i\rangle\langle i|$ are perfectly distinguishable (orthogonal) pure states that can be regarded as states describing the state of a classical register. In what follows we will abuse the notation and simply treat quantum measurements (denoted by symbols $\mathcal{M}, \mathcal{N}, \mathcal{P}$, ...) as quantum channels having the classical outputs. Using this interpretation one can readily use the results concerning the discrimination of quantum channels for generalized measurements. In particular, for entanglement-assisted discrimination of quantum channels we have a classic result due to Helstrom [33]. It states that the probability of correct discrimination $p_{\mathrm{opt}}(\Phi, \Psi)$ between two quantum channels Φ and Ψ is given by

$$p_{\mathrm{opt}}(\Phi, \Psi) = \frac{1}{2} + \frac{1}{4} \|\Phi - \Psi\|_{\diamond}, \tag{3}$$

where $\|\mathbf{S}\|_{\diamond} = \max_{\|X\|_1 = 1} \|(\mathbf{S} \otimes \mathbb{1})(X)\|_1$ denotes the diamond norm of the linear map $\mathbf{S}$. Any optimal X is called a discriminator. Thus, if the value of the diamond norm of the difference of two channels is strictly smaller than two, then the two channels cannot be distinguished perfectly in a single-shot scenario.

In this work we will be concerned with von Neumann measurements i.e. projective and fine-grained measurements on a Hilbert space of a given dimension d. Von Neumann measurements $\mathcal{P}$ in $\mathbb{C}^d$ are tuples of orthogonal projectors on vectors forming an orthonormal basis $\{|\psi_i\rangle\}_{i=1}^d$ in $\mathbb{C}^d$ i.e.

$$\mathcal{P} = (|\psi_1\rangle\langle\psi_1|, |\psi_2\rangle\langle\psi_2|, \ldots, |\psi_d\rangle\langle\psi_d|). \tag{4}$$

In what follows we will use $\mathcal{P}_{\mathbf{1}}$ to denote the measurement in the standard computational basis. We will also use $\mathcal{P}_U$ to denote the von Neumann measurement in the basis $|\psi_i\rangle = U|i\rangle$ for a unitary $d \times d$ matrix $U \in \mathcal{U}_d$. In other words, vectors $|\psi_i\rangle$ from Eq.(4) are columns of the matrix U. We also specify the subset $\mathcal{DU}_d \subset \mathcal{U}_d$ of diagonal, unitary matrices. Consider now a general task of discriminating between two projective measurements $\mathcal{P}_{U_1}$, $\mathcal{P}_{U_2}$, and let $p_{\mathrm{opt}}(\mathcal{P}_{U_1}, \mathcal{P}_{U_2})$ be the optimal probability for discriminating between these measurements (both for minimum error and for unambiguous discrimination). Then, due to the unitary invariance of the discrimination problem and identity $\mathcal{P}_U(\cdot) = \mathcal{P}_{\mathbf{1}} \circ (U^\dagger \cdot U)$ we obtain that $p_{\mathrm{opt}}(\mathcal{P}_{U_1}, \mathcal{P}_{U_2}) = p_{\mathrm{opt}}(\mathcal{P}_{\mathbf{1}}, \mathcal{P}_{U_1^\dagger U_2})$. Therefore, for any discrimination of projective measurements, without loss of generality, we can limit ourselves to considering the problem of distinguishing between the measurement in the standard basis $\mathcal{P}_{\mathbf{1}}$ and another projective measurement $\mathcal{P}_U$. It is important to note that definition of measurement $\mathcal{P}_U$ distinguishes projective measurements differing only by ordering of elements of the basis. Moreover, a set of unitary matrices $\{UE |\ E \in \mathcal{DU}_d\}$ specifies the same projective measurement, ie. $\mathcal{P}_U = \mathcal{P}_{UE}$ for all $E \in \mathcal{DU}_d$.

Distinguishability of quantum measurements is strictly related to the distinguishability of unitary channels. The prominent result [34, 35] gives an expression which makes calculating the diamond norm of the difference of unitary channels Φ_U, $\Phi_{\mathbf{1}}$ substantially easier. It says that the for a unitary matrix U we have

$$\|\Phi_U - \Phi_{\mathbf{1}}\|_{\diamond} = 2\sqrt{1 - \nu^2}, \tag{5}$$

where $\nu = \min_{x \in W(U)} |x|$ and $W(X) := \{\langle\psi|X|\psi\rangle : \langle\psi|\psi\rangle = 1\}$ denotes the numerical range of the operator X. Building on this result, in [29] the following characterization of the diamond norm of the distance between von Neumann measurements was obtained

$$\|\mathcal{P}_U - \mathcal{P}_{\mathbf{1}}\|_{\diamond} = \min_{E \in \mathcal{DU}_d} \|\Phi_{UE} - \Phi_{\mathbf{1}}\|_{\diamond}. \tag{6}$$

Therefore the distance between two von Neumann measurements is the minimal value of the

[1] In this paper we restrict our attention to measurements with a finite number of outcomes.

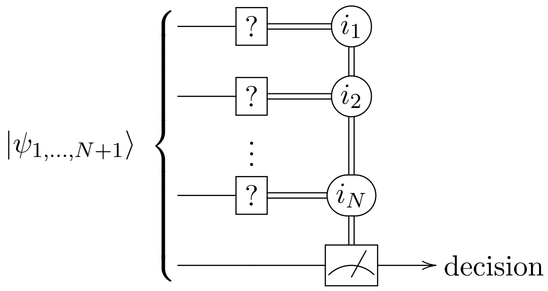

Figure 5: A schematic representation of the parallel discrimination scheme of quantum measurements.

diamond norm of the difference between optimally coherified channels [36]. Moreover, there is a simple condition which lets us check whether two measurements are perfectly distinguishable [24, 26, 29]. It holds that $\|\mathcal{P}_U - \mathcal{P}_{\mathbf{1}}\|_\diamond = 2$ if and only if there exists a state ρ such that

$$\operatorname{diag}(\rho U) = 0. \tag{7}$$

Equation (6) will be of significant importance throughout this work. We will also make use of the dephasing channel denoted $\Delta(\rho) = \sum_i |i\rangle\langle i|\rho|i\rangle\langle i|$. Finally, when talking about the eigenvalues of a unitary matrix $U \in \mathcal{U}_d$, we will follow convention that $\lambda_1 = \mathrm{e}^{\mathrm{i}\alpha_1}, \ldots, \lambda_d = \mathrm{e}^{\mathrm{i}\alpha_d}$ are ordered by their phases, if $\alpha_1 \leq \ldots \leq \alpha_d$ for $\alpha_1, \ldots, \alpha_d \in [0, 2\pi)$. The angle of the shortest arc containing all eigenvalues will be denoted by $\Theta(U)$.

3 Minimum error discrimination

Multiple-shot minimum error discrimination is a natural generalization of the single-shot scheme studied in [29]. In this scenario we have access to multiple queries to quantum measurements which is mathematically equivalent to the problem of single-shot discrimination of channels $\mathcal{P}_{U^{\otimes N}}, \mathcal{P}_{\mathbf{1}^{\otimes N}}$ acting on states defined on multiple-component Hilbert space $(\mathbb{C}^d)^{\otimes N}$ (see Fig. 5). In the following sections will write shortly $\mathcal{P}_{\mathbf{1}}$ instead of $\mathcal{P}_{\mathbf{1}^{\otimes N}}$.

As we described in the preceding section, the problem of distinguishing quantum measurements is intimately related to distinguishing unitary channels [29]. In what follows we leverage this result to prove a number of results regarding multiple-shot discrimination of von Neumann measurements.

3.1 Optimality of the parallel scheme

The authors of [37] showed that parallel discrimination scheme is optimal among all possible architectures for the case of discrimination of unitary channels. In this subsection we will prove a theorem which thesis is rendered in the spirit of the results obtained in [37], nevertheless, our theorem concerns the discrimination of von Neumann measurements. It is worth mentioning here that the optimality of the parallel scheme is no longer true when studying the discrimination of general (non-projective) POVMs. In the case of discrimination of POVMs with rank-one effects one may need to use adaptive discrimination scheme to obtain perfect distinguishability [23]. While considering the discrimination of von Neumann measurements, one could expect that we can improve the discrimination by performing some processing based on the obtained measurement's labels. It appears that such processing will not improve the discrimination.

We will focus on the minimum error discrimination in the parallel scenario and our goal will be to characterize the probability of correct discrimination after N queries to the black box. The first step is to extend Eq. (6) to the parallel setting. We study the form of the optimal matrix E in the parallel scheme. The following theorem, which proof is presented in Appendix A, states that it has a tensor product form.

Theorem 1. *Let $N \in \mathbb{N}$, $U \in U_d$ and let $\mathcal{P}_U$ be the corresponding von Neumann measurement on $\mathbb{C}^d$. Then we have the following equality*

$$\|\mathcal{P}_{U^{\otimes N}} - \mathcal{P}_{\mathbf{1}}\|_\diamond = \min_{E \in \mathcal{D}\mathcal{U}_d} \|\Phi_{U^{\otimes N} E^{\otimes N}} - \Phi_{\mathbf{1}}\|_\diamond . \tag{8}$$

Now we are interested in calculating the number of usages of the black box required for perfect discrimination. Let us recall here that in the case of distinguishing unitary operations this can always be achieved in a finite number of steps $N = \lceil \frac{\pi}{\Theta(U)} \rceil$ [15]. A similar result is achievable in the case of distinguishing von Neumann measurements. Let UE_0 be an *optimal* unitary matrix, that is a matrix for which

$$\|\mathcal{P}_U - \mathcal{P}_{\mathbf{1}}\|_\diamond = \|\Phi_{UE_0} - \Phi_{\mathbf{1}}\|_\diamond . \tag{9}$$

Now we compute the number of queries needed for perfect discrimination. We have already proven in Theorem 1 that if UE_0 is the optimal matrix, then the matrix $(UE_0)^{\otimes N}$ will also be optimal. Therefore, the value of diamond norm $\|\mathcal{P}_{U^{\otimes N}} - \mathcal{P}_{\mathbf{1}}\|_\diamond$ can be expressed as a function of $\Theta\left((UE_0)^{\otimes N}\right)$. As long as $0 \notin W\left((UE_0)^{\otimes N}\right)$, it holds that $\Theta((UE_0)^{\otimes N}) = N\Theta(UE_0)$. The first time zero enters the numerical range $W\left((UE_0)^{\otimes N}\right)$ is therefore equal to $N = \lceil \frac{\pi}{\Upsilon(U)} \rceil$, where $\Upsilon(U)$ is an optimized version of $\Theta(U)$ *i.e.*

$$\Upsilon(U) := \min_{E \in \mathcal{DU}_d} \Theta(UE). \tag{10}$$

We summarize the above discussion as the Corollary below providing the value of $\|\mathcal{P}_{U^{\otimes N}} - \mathcal{P}_{\mathbf{1}}\|_\diamond$ in terms of $\Theta(U)$.

Corollary 1. *Let $N \in \mathbb{N}$, $U \in \mathcal{U}_d$. The following holds*

(i) if $N\Upsilon(U) \geq \pi$, then $\|\mathcal{P}_{U^{\otimes N}} - \mathcal{P}_{\mathbf{1}}\|_\diamond = 2$;

(ii) if $N\Upsilon(U) < \pi$, then $\|\mathcal{P}_{U^{\otimes N}} - \mathcal{P}_{\mathbf{1}}\|_\diamond = 2\sin\left(\frac{N}{2}\Upsilon(U)\right)$.

Another interesting property resulting from Theorem 1 and its proof is the amount of the discriminator's entanglement with environment. The minimal dimension of an auxiliary system needed for optimal discrimination is equal to the rank of the input state. We found out that for the majority of von Neumann measurements it is sufficient when the dimension of the auxiliary system is either two or three. A more detailed discussion is presented in Appendix A after the proof of Theorem 1.

Eventually, we are in the position to prove the optimality of parallel discrimination scheme, which we present as the following theorem.

Theorem 2. *Let $U \in \mathcal{U}_d$. Consider the distinguishability of general quantum network with N uses of the black box in which there is one of two measurements - either $\mathcal{P}_U$ or $\mathcal{P}_{\mathbf{1}}$. Then the probability of correct distinction cannot be better then in the parallel scenario.*

Proof. Without loss of generality we may assume that the processing is performed using only unitary operations. Indeed, using Stinespring dilation theorem, any channel might be represented via a unitary channel on a larger system followed by the partial trace operation. What is left to observe is that $\|\mathrm{tr}_B(X_{AB})\|_1 \leq \|X_{AB}\|_1$ for arbitrary bipartite matrix X_{AB}.

The sequential scheme of discrimination of von Neumann measurements is shown in Fig. 6 and can be expressed as a channel

$$\Psi_U = (\Delta_{1,\ldots,N} \otimes \mathbb{1})\,\Phi_{A_U}, \tag{11}$$

associated with a matrix A_U. Here $\Delta_{1,\ldots,N}$ is the dephasing channel on subsystems $1,\ldots,N$. The channel Φ_{A_U} is shown in Fig. 7 and the exact form of this transformation can be found in Appendix B.

Assuming that matrix U is optimal, that is $\Upsilon(U) = \Theta(U)$, we have $\|\mathcal{P}_{U^{\otimes N}} - \mathcal{P}_{\mathbf{1}}\|_\diamond = \|\Phi_{U^{\otimes N}} - \Phi_{\mathbf{1}}\|_\diamond$. Hence, we may calculate the distance between Ψ_U and $\Psi_{\mathbf{1}}$ as

$$\begin{aligned}
&\max_\rho \|(\Psi_U - \Psi_{\mathbf{1}})(\rho)\|_1 \\
&= \max_\rho \|[(\Delta_{1,\ldots,N} \otimes \mathbb{1})(\Phi_{A_U} - \Phi_{A_{\mathbf{1}}})](\rho)\|_1 \\
&\leq \max_\rho \|(\Phi_{A_U} - \Phi_{A_{\mathbf{1}}})(\rho)\|_1 \\
&\leq \max_\rho \|(\Phi_{U^{\otimes N} \otimes \mathbf{1}} - \Phi_{\mathbf{1}})(\rho)\|_1 \\
&= \|\Phi_{U^{\otimes N}} - \Phi_{\mathbf{1}}\|_\diamond = \|\mathcal{P}_{U^{\otimes N}} - \mathcal{P}_{\mathbf{1}}\|_\diamond,
\end{aligned} \tag{12}$$

where we maximize over states ρ of appropriate dimensions. The induced trace norm is monotonically decreasing under the action of channels and this gives us first inequality. The second one follows from the optimality of the parallel scheme of distinguishing unitary channels [37]. Therefore, the adaptive scenario does not give any advantage over the parallel scheme. □

3.2 Discrimination of random measurements

The structural characterization of multiple-shot discrimination of von Neumann measurements given above allows us to draw strong conclusions about distinguishability of generic pairs of von Neumann measurements. In this work we restrict our attention to pairs of measurements distributed independently according to the natural distribution coming from the Haar measure $\mu(\mathcal{U}_d)$ [38].

Theorem 3. *Consider two independently distributed Haar-random von Naumann measurements on $\mathbb{C}^d$, i.e. $\mathcal{P}_U, \mathcal{P}_V$, where $U \sim \mu(\mathcal{U}_d)$, $V \sim \mu(\mathcal{U}_d)$. Let $p_{\mathrm{opt}}(\mathcal{P}_{U^{\otimes 2}}, \mathcal{P}_{V^{\otimes 2}})$ be the optimal probability of discrimination measurements $\mathcal{P}_U$ and $\mathcal{P}_V$ using two queries and as-*

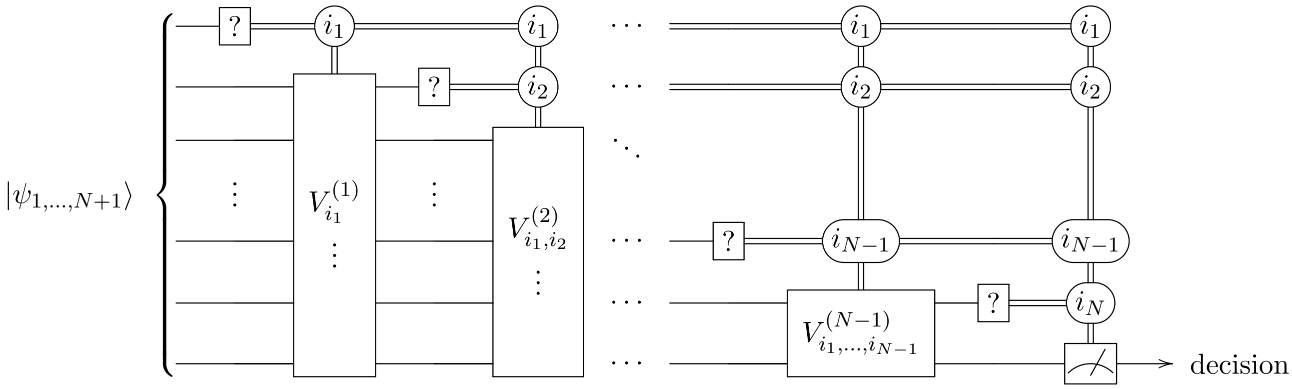

Figure 6: Schematic depiction of the sequential scheme. In the k^{th} step, after obtaining the label i_k, we utilize all labels $i_1, \dots, i_k$ to modify the remaining parts of the state in the hope of improving distinguishability.

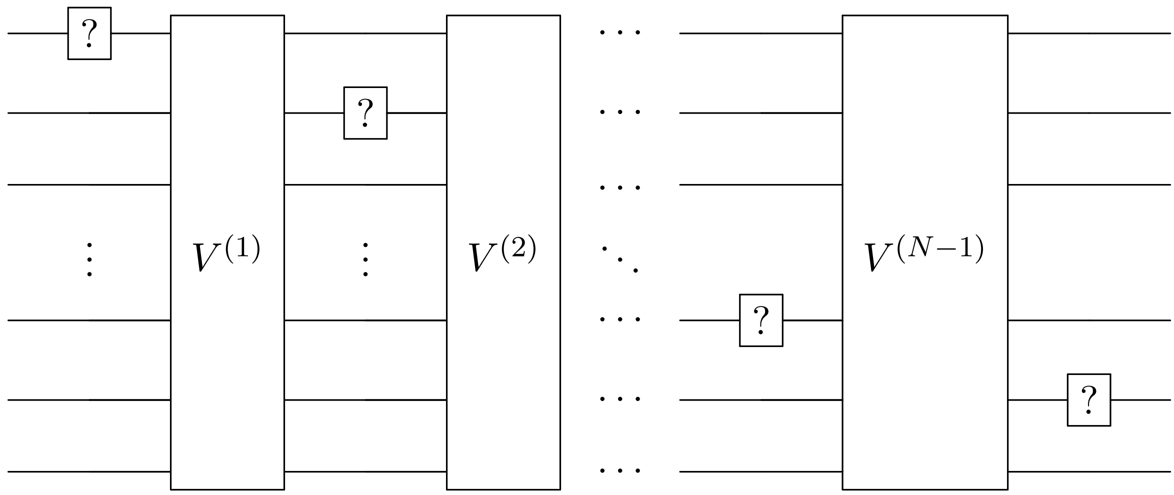

Figure 7: Visualization of the channel Φ_{A_U}. Here $V^{(k)} = \sum_{i_1,\dots,i_k} |i_1,\dots,i_k\rangle\langle i_1,\dots,i_k| \otimes V^{(k)}_{i_1,\dots,i_k}$.

sistance of entanglement. Then, we have the following bound

$$\Pr_{U,V\sim\mu(\mathcal{U}_d)} \left(p_{\text{opt}}(\mathcal{P}_{U^{\otimes 2}}, \mathcal{P}_{V^{\otimes 2}}) < 1\right) \leq \frac{1}{2^{d-1}}. \quad (13)$$

In other words, in the limit of large dimensions d, typical Haar-random von Neumann measurements are perfectly distinguishable with the usage of two queries and assistance of entanglement (the probability that they cannot be perfectly distinguished is exponentially suppressed as a function of d).

Proof. From the unitary invariance of the Haar measure and the symmetry of the problem of measurement discrimination it follows that the distribution of the random variable $p_{\text{opt}}(\mathcal{P}_{\mathbf{1}^{\otimes 2}}, \mathcal{P}_{U^{\otimes 2}})$ is identical to that of $p_{\text{opt}}(\mathcal{P}_{U^{\otimes 2}}, \mathcal{P}_{V^{\otimes 2}})$. Consequently, we have

$$\begin{aligned}&\Pr_{U,V\sim\mu(\mathcal{U}_d)} \left(p_{\text{opt}}(\mathcal{P}_{U^{\otimes 2}}, \mathcal{P}_{V^{\otimes 2}}) < 1\right) \\ &= \Pr_{U\sim\mu(\mathcal{U}_d)} \left(p_{\text{opt}}(\mathcal{P}_{\mathbf{1}^{\otimes 2}}, \mathcal{P}_{U^{\otimes 2}}) < 1\right).\end{aligned} \quad (14)$$

From Corollary 1 it follows that the condition $\|\mathcal{P}_U - P_{\mathbf{1}}\|_\diamond \geq \sqrt{2}$ implies $\|\mathcal{P}_{U^{\otimes 2}} - P_{\mathbf{1}^{\otimes 2}}\|_\diamond = 2$ and consequently we have also perfect discrimination of two copies of measurements: $p_{\text{opt}}(\mathcal{P}_{U^{\otimes 2}}, \mathcal{P}_{\mathbf{1}^{\otimes 2}}) = 1$. Therefore we have

$$\begin{aligned}&\Pr_{U\sim\mu(\mathcal{U}_d)} \left(p_{\text{opt}}(\mathcal{P}_{U^{\otimes 2}}, \mathcal{P}_{\mathbf{1}^{\otimes 2}}) < 1\right) \\ &\leq \Pr_{U\sim\mu(\mathcal{U}_d)} \left(\|\mathcal{P}_U - \mathcal{P}_{\mathbf{1}}\|_\diamond \leq \sqrt{2}\right).\end{aligned} \quad (15)$$

Using now the characterization given in Eq.(6) in conjunction with the formula in Eq.(5) we obtain $\|\mathcal{P}_U - P_{\mathbf{1}}\|_\diamond \geq 2\sqrt{1 - |U_{11}|^2}$ (note that

$U_{11} = \operatorname{tr}(|1\rangle\langle 1|U) \in W(U)$. Using this and simple algebra we get

$$\Pr_{U\sim\mu(\mathcal{U}_d)} (p_{\mathrm{opt}}(\mathcal{P}_{U^{\otimes 2}}, \mathcal{P}_{\mathbf{1}^{\otimes 2}}) < 1) \leq \Pr_{U\sim\mu(\mathcal{U}_d)} \left(|U_{11}|^2 \geq \frac{1}{2}\right). \tag{16}$$

The right-hand side of the above inequality can be computed exactly using the property that for Haar-distributed U the random variable $X = |U_{11}|^2$ is distributed according to the beta distribution $p(X) = (d-1)(1-X)^{d-2}$ (see for instance Eq. (9) in [39]). The simple integration gives $(1/2)^{d-1}$, which together with Eq.(14) gives the claimed result. □

4 Unambiguous discrimination

The unambiguous discrimination of measurements $\mathcal{P}_{\mathbf{1}}$ and $\mathcal{P}_U$ can be understood as unambiguous discrimination [31] of states generated by the corresponding channels. Specifically, for a fixed input state σ, the output states $(\mathcal{P}_{\mathbf{1}} \otimes \mathbb{1})(\sigma), (\mathcal{P}_U \otimes \mathbb{1})(\sigma)$ can be unambiguously discriminated using the measurement strategy $\mathcal{M} = (M_{\mathbf{1}}, M_U, M_?)$, where the first two effects represent conclusive answers and the last one corresponds to the inconclusive output of the procedure. For equal a priori probabilities of occurrence of $\mathcal{P}_{\mathbf{1}}$ and $\mathcal{P}_U$, as well as fixed σ (possibly entangled) and $\mathcal{M}$, the success probability is given by

$$p_{\mathrm{u}}(\mathcal{P}_{\mathbf{1}}, \mathcal{P}_U; \sigma, \mathcal{M}) = \frac{1}{2}\operatorname{tr}(M_{\mathbf{1}}(\mathcal{P}_{\mathbf{1}} \otimes \mathbb{1})(\sigma)) + \frac{1}{2}\operatorname{tr}(M_U(\mathcal{P}_U \otimes \mathbb{1})(\sigma)), \tag{17}$$

where additionally the unambiguity condition has to be satisfied:

$$\operatorname{tr}(M_U(\mathcal{P}_{\mathbf{1}} \otimes \mathbb{1})(\sigma)) = \operatorname{tr}(M_{\mathbf{1}}(\mathcal{P}_U \otimes \mathbb{1})(\sigma)) = 0. \tag{18}$$

The optimal success probability of unambiguous discrimination of measurements $\mathcal{P}_{\mathbf{1}}, \mathcal{P}_U$ can be now defined as the maximum of (17) over all strategies. Formally, we have

$$p_{\mathrm{u}}(\mathcal{P}_{\mathbf{1}}, \mathcal{P}_U) := \max_{\sigma\in\mathcal{D}(\mathbb{C}^d\otimes\mathbb{C}^{d'})} \max_{\mathcal{M}\in\mathrm{POVM}(\mathbb{C}^d\otimes\mathbb{C}^{d'})} p_{\mathrm{u}}(\mathcal{P}_{\mathbf{1}}, \mathcal{P}_U; \sigma, \mathcal{M}), \tag{19}$$

where $\mathcal{M} \in \mathrm{POVM}(\mathbb{C}^d \otimes \mathbb{C}^{d'})$ is a three-outcome measurement on $\mathbb{C}^d \otimes \mathbb{C}^{d'}$ that satisfies constrains (18) end σ is a state on the extended Hilbert space $\mathbb{C}^d \otimes \mathbb{C}^{d'}$.

4.1 Unambiguous discrimination with assistance of entanglement

In this subsection we present our main result which gives the probability of unambiguous discrimination with the use of entanglement This is presented as Theorem 4 while its proof is postponed to Appendix C. Aside from giving a simple expression for this probability, this result reduces the problem of unambiguous measurement discrimination to a convex optimization task and gives a simple relationship between the diamond norm and the probability of unambiguous discrimination.

Theorem 4. *The optimal success probability of unambiguous discrimination between von Neumann measurements $\mathcal{P}_{\mathbf{1}}$ and $\mathcal{P}_U$ is given by*

$$p_{\mathrm{u}}(\mathcal{P}_{\mathbf{1}}, \mathcal{P}_U) = 1 - \min_{\rho\in\mathcal{D}(\mathbb{C}^d)} \sum_i |\langle i|\rho U|i\rangle|. \tag{20}$$

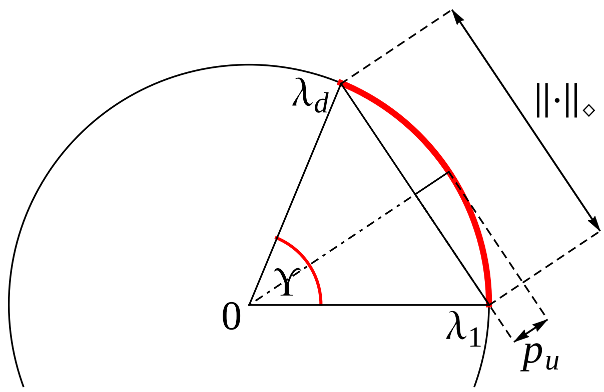

Figure 8: Schematic depiction of the relationship between the diamond norm $\|\mathcal{P}_U - \mathcal{P}_{\mathbf{1}}\|_\diamond$, the probability of unambiguous discrimination $p_{\mathrm{u}}(\mathcal{P}_{\mathbf{1}}, \mathcal{P}_U)$ and $\Upsilon(U)$.

The results coming from Theorem 1 and Theorem 4 give a nice geometric interpretation for the relationship between the diamond norm and the probability of unambiguous discrimination. This is depicted in Fig. 8. We start with a von Neumann measurement in a basis given by some unitary matrix U and try to distinguish it from the measurement in the computational basis. We assume that U is optimal and denote U's eigenvalues as $\lambda_1, \ldots, \lambda_d$ ordered according to their phases and use the symbol $\Upsilon(U)$ to denote the angle between two most distant eigenvalues λ_1 and λ_d. The dependence of the diamond norm and probability of unambiguous discrimination is clearly shown.

Remark 1. *The above calculations can be easily extended to the case of parallel discrimination scheme. It suffices to substitute U with $U^{\otimes N}$ and then we obtain that*

$$p_{\mathrm{u}}(\mathcal{P}_{U^{\otimes N}}, \mathcal{P}_{\mathbf{1}}) = 1 - \min_{\rho \in \mathcal{D}\left(\mathbb{C}^{d^N}\right)} \sum_i |\langle i|\rho U^{\otimes N}|i\rangle|. \tag{21}$$

Basing on Remark 1, we note that the angle $\Upsilon(U)$ increases in the multiple-shot case with the number of queries. This is depicted in Fig. 9 for two- and three-shots scenarios.

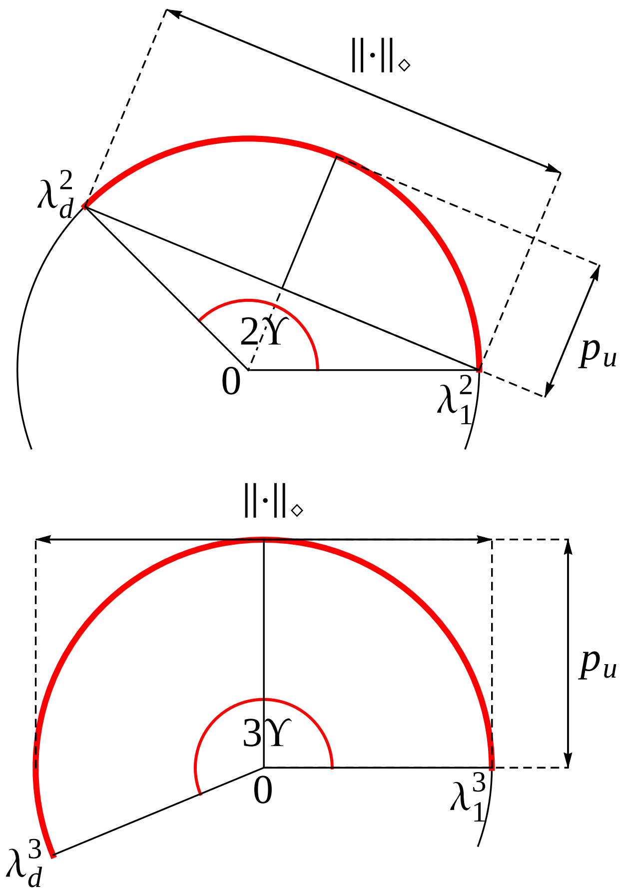

Figure 9: Figure similar to Fig. 8 which represents two- (left) and three- (right) shots scenario.

In the general scheme we are allowed to use conditional unitary transformations $\{V^{(k)}\}$ after each measurement, thus our setting for discrimination is the same as presented in Fig. 6 and Fig. 7. Similarly to the multiple-shot minimum error discrimination, we will show that adaptive discrimination scheme does not give any advantage over the parallel one for unambiguous discrimination. We state this formally in the following theorem, which proof is moved to Appendix D.

Theorem 5. *Let $U \in \mathcal{U}_d$. Consider the unambiguous discrimination of general quantum network with N uses of the black box in which there is one of two measurements - either $\mathcal{P}_U$ or $\mathcal{P}_{\mathbf{1}}$. Then the probability of correct distinction cannot be better then in the parallel scenario.*

4.2 Unambiguous discrimination without assistance of entanglement

In this subsection we provide a brief discussion on a special case of unambiguous discrimination without the utilization of entangled states. We will use the following notation. Let $\Gamma, \Lambda \subset \{1, \ldots, d\}$. For given unitary matrix U, we define $P_\Gamma := \sum_{i\in\Gamma} |i\rangle\langle i|$ and $Q_\Lambda := U P_\Lambda U^\dagger$. We set $\mathbb{P}_{\Gamma,\Lambda}$ to be the orthogonal projector onto $\mathrm{Span}\left(\{U|i\rangle\}_{i\in\Gamma^c}\right) \cap \mathrm{Span}\left(\{|j\rangle\}_{j\in\Lambda^c}\right)$, where Γ^c, Λ^c denote the complements of Γ and Λ respectively. The following theorem states the optimal success probability of unambiguous discrimination without the use of entanglement.

Theorem 6. *The optimal success probability of unambiguous discrimination, without the use of entanglement, between von Neumann measurements $\mathcal{P}_{\mathbf{1}}$ and $\mathcal{P}_U$ is given by*

$$\begin{aligned}&\tilde{p}_{\mathrm{u}}\left(\mathcal{P}_{\mathbf{1}}, \mathcal{P}_U\right)\\ &= \frac{1}{2} \max_{\Gamma,\Lambda \subset \{1,\ldots,d\}: \Gamma\cap\Lambda=\emptyset} \left\| \mathbb{P}_{\Gamma,\Lambda}(P_\Gamma + Q_\Lambda)\mathbb{P}_{\Gamma,\Lambda} \right\|\end{aligned} \tag{22}$$

with $P_\Gamma, Q_\Lambda, \mathbb{P}_{\Gamma,\Lambda}$ defined as above.

The proof of this theorem is presented in Appendix E.

Remark. *The projector $\mathbb{P}_{\Gamma,\Lambda}$ projects onto the intersection of supports of P_{Λ^c} and Q_{Γ^c}. By the use of Theorem 4 from [40], we can express the optimal probability of unambiguous discrimination as*

$$\tilde{p}_{\mathrm{u}}(\mathcal{P}_{\mathbf{1}},\mathcal{P}_U) = 2 \max_{\Gamma,\Lambda\subset\{1,\ldots,d\}:\Gamma\cap\Lambda=\emptyset} \Big\| P_{\Lambda^c}(P_{\Lambda^c}+Q_{\Gamma^c})^{-1}Q_\Lambda(P_{\Lambda^c}+Q_{\Gamma^c})^{-1}\ P_{\Lambda^c} + Q_{\Gamma^c}(P_{\Lambda^c}+Q_{\Gamma^c})^{-1}P_\Gamma(P_{\Lambda^c}+Q_{\Gamma^c})^{-1}\ Q_{\Gamma^c}\Big\|, \tag{23}$$

where $(\cdot)^{-1}$ denotes Moore-Penrose pseudo inverse [13]. Moreover, the optimal input state is the one which gives the above norm.

Corollary 2. *In the case of qubit measurements, the optimal probability of unambiguous discrimination of $\mathcal{P}_{\mathbf{1}}$ and $\mathcal{P}_U$ is given by the following (discontinuous) function*

$$\tilde{p}_{\mathrm{u}}(\mathcal{P}_{\mathbf{1}},\mathcal{P}_U) = \begin{cases} 1 & \text{if} \quad |U_{1,2}|^2 = 1 \\ \frac{1}{2}|U_{1,2}|^2 & \text{if} \quad |U_{1,2}|^2 < 1 \end{cases}. \tag{24}$$

In both cases the optimal input state can be chosen to be $|1\rangle\langle 1|$.

The following corollary states that in the qubit case the unambiguous discrimination with the assistance of entanglement always outperforms the unambiguous discrimination without the use of entanglement. On top of that, the special cases for which the use of entanglement does not give any advantage are described.

Corollary 3. *Let $\mathcal{P}_{\mathbf{1}}$ and $\mathcal{P}_U$ be two von Neumann measurements on a qubit. If $|U_{1,1}| \notin \{0,1\}$, then the probability of entanglement-assisted unambiguous discrimination is given by*

$$p_{\mathrm{u}} = 1 - |U_{1,1}| \tag{25}$$

and it is always greater then the probability without assistance of entanglement

$$\tilde{p}_{\mathrm{u}} = \begin{cases} 1 & \text{if} \quad |U_{1,2}|^2 = 1 \\ \frac{1}{2}|U_{1,2}|^2 & \text{if} \quad |U_{1,2}|^2 < 1. \end{cases} \tag{26}$$

Moreover, if $|U_{1,1}| \in \{0,1\}$, then $p_{\mathrm{u}} = \tilde{p}_{\mathrm{u}}$.

Remark. *The above considerations can be extended to unambiguous discrimination of multiple copies of von Neumann measurements applied in parallel. To this end, if we have access to N parallel queries to a black box measurement, it suffices to replace unitaries $\mathbb{1}$ by $\mathbb{1}^{\otimes N}$ and U by $U^{\otimes N}$ in the above computations. Interestingly, in the contrast to unambiguous discrimination of quantum states [31], having access to two copies of black box measurement, sometimes allows attaining perfect discrimination. Specifically, consider the problem of discriminating between $\mathcal{P}_{\mathbf{1}}$ and $\mathcal{P}_H$, where $H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. Explicit computation shows that by taking the input state as $|\Psi\rangle = \frac{1}{\sqrt{2}}(|1\rangle|1\rangle - |2\rangle|2\rangle)$ allows us to perfectly distinguish between $\mathcal{P}_{\mathbf{1}}^{\otimes 2}$ and $\mathcal{P}_H^{\otimes 2}$.*

5 Conclusions

We have presented a comprehensive treatment of the problem of discrimination of von Neumann measurements. First of all, we showed an alternative proof of the fact that for any pair of measurements $\mathcal{P}_1$ and $\mathcal{P}_2$, $\mathcal{P}_1 \neq \mathcal{P}_2$, there exists a finite number N of uses of the black box which allows us to achieve perfect discrimination. Moreover, we calculated the exact value of the diamond norm for given N. This is formally stated in Corollary 1. We also proved that the parallel discrimination scheme is optimal in the scenario of multiple-shot minimum error discrimination of von Neumann measurements (see Theorem 2).

Moreover, we studied unambiguous discrimination of von Neumann measurements. Our main contribution to this problem was the derivation of the general schemes that attain the optimal success probability both with (see Theorem 4) and without (see Theorem 6) the assistance of entanglement. Interestingly, for entanglement-assisted unambiguous discrimination the optimal success probability is functionally related to the corresponding success probability for minimum error discrimination. Finally, we show that the parallel scheme is also optimal for unambiguous discrimination.

There are many interesting directions for further study that still remain to be explored. Below we list the most important (in our opinion) open research problems:

- Generalization of our results from projective measurements to other classes of measurements such as projective-simulable measurements [41], measurements with limited number of outcomes [42] or general quantum measurements (POVMs).

- Systematical study of the problem of unambiguous discrimination of projective measurements in the multiple-shot regime.
- Can typical pairs of Haar-random projective measurements on $\mathbb{C}^d$ be discriminated perfectly using only one query and the assistance of entanglement as $d \to \infty$?
- How much entanglement is needed to attain the optimal success probability of multiple-shot discrimination of generic projective measurements on $\mathbb{C}^d$? In the same scenario, is it necessary to adopt the final measurement to the pair of measurements to be discriminated?

Acknowledgements

We would like to thank Alessandro Bisio and Giulio Chiribella for fruitful discussions.

This work was supported by the Polish National Science Centre under project numbers 2016/22/E/ST6/00062 (ZP, AK, RK) and 2015/18/E/ST2/ 00327 (ŁP). MO acknowledges the support of Homing programme of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund.

References

[1] J. Carolan, C. Harrold, C. Sparrow, E. Martín-López, N. J. Russell, J. W. Silverstone, P. J. Shadbolt, N. Matsuda, M. Oguma, M. Itoh, G. D. Marshall, M. G. Thompson, J. C. F. Matthews, T. Hashimoto, J. L. O'Brien, and A. Laing, "Universal linear optics," *Science*, vol. 349, no. 6249, pp. 711–716, 2015.

[2] S. Boixo, S. V. Isakov, V. N. Smelyanskiy, R. Babbush, N. Ding, Z. Jiang, M. J. Bremner, J. M. Martinis, and H. Neven, "Characterizing quantum supremacy in near-term devices," *Nature Physics*, vol. 14, no. 6, pp. 595–600, 2018.

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

[4] E. Magesan, J. M. Gambetta, and J. Emerson, "Scalable and Robust Randomized Benchmarking of Quantum Processes," *Physical Review Letters*, vol. 106, p. 180504, 2011.

[5] L. Aolita, C. Gogolin, M. Kliesch, and J. Eisert, "Reliable quantum certification of photonic state preparations," *Nature Communications*, vol. 6, p. 8498, 2015.

[6] J. R. Wootton, "Benchmarking of quantum processors with random circuits," *arxiv:1806.02736*, 2018.

[7] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, "Quantum certification and benchmarking," *Nature Reviews Physics*, pp. 1–9, 2020.

[8] A. Chefles, "Quantum state discrimination," *Contemporary Physics*, vol. 41, no. 6, pp. 401–424, 2000.

[9] S. M. Barnett and S. Croke, "Quantum state discrimination," *Advances in Optics and Photonics*, vol. 1, no. 2, pp. 238–278, 2009.

[10] J. A. Bergou, "Discrimination of quantum states," *Journal of Modern Optics*, vol. 57, no. 3, pp. 160–180, 2010.

[11] J. Bae and L.-C. Kwek, "Quantum state discrimination and its applications," *Journal of Physics A: Mathematical and General*, vol. 48, no. 8, p. 083001, 2015.

[12] S. Pirandola, R. Laurenza, C. Lupo, and J. L. Pereira, "Fundamental limits to quantum channel discrimination," *npj Quantum Information*, vol. 5, no. 1, pp. 1–8, 2019.

[13] J. Watrous, *The Theory of Quantum Information*. Cambridge University Press, 2018.

[14] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, *et al.*, "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, no. 7779, pp. 505–510, 2019.

[15] A. Acín, "Statistical distinguishability between unitary operations," *Physical Review Letters*, vol. 87, no. 17, p. 177901, 2001.

[16] R. Duan, Y. Feng, and M. Ying, "Entanglement is not necessary for perfect discrimination between unitary operations," *Physical Review Letters*, vol. 98, no. 10, p. 100503, 2007.

[17] R. Duan, Y. Feng, and M. Ying, "Local distinguishability of multipartite unitary operations," *Physical Review Letters*, vol. 100, no. 2, p. 020503, 2008.

[18] G. Chiribella, G. M. D'Ariano, and M. Roetteler, "Identification of a reversible quantum gate: assessing the resources," *New Journal of Physics*, vol. 15, no. 10, p. 103019, 2013.

[19] R. Duan, Y. Feng, and M. Ying, "Perfect distinguishability of quantum operations," *Physical Review Letters*, vol. 103, no. 21, p. 210501, 2009.

[20] A. W. Harrow, A. Hassidim, D. W. Leung, and J. Watrous, "Adaptive versus nonadaptive strategies for quantum channel discrimination," *Physical Review A*, vol. 81, no. 3, p. 032339, 2010.

[21] T. P. Cope and S. Pirandola, "Adaptive estimation and discrimination of Holevo-Werner channels," *Quantum Measurements and Quantum Metrology*, vol. 4, no. 1, pp. 44–52, 2017.

[22] S. Pirandola and C. Lupo, "Ultimate precision of adaptive noise estimation," *Physical Review Letters*, vol. 118, no. 10, p. 100502, 2017.

[23] A. Krawiec, Ł. Pawela, and Z. Puchała, "Discrimination of POVMs with rank-one effects," *Quantum Information Processing*, vol. 19, no. 12, pp. 1–12, 2020.

[24] R. Duan, C. Guo, C.-K. Li, and Y. Li, "Parallel distinguishability of quantum operations," in *2016 IEEE International Symposium on Information Theory (ISIT)*, pp. 2259–2263, IEEE, 2016.

[25] I. Nechita, Z. Puchała, Ł. Pawela, and K. Życzkowski, "Almost all quantum channels are equidistant," *Journal of Mathematical Physics*, vol. 59, no. 5, p. 052201, 2018.

[26] Z. Ji, Y. Feng, R. Duan, and M. Ying, "Identification and distance measures of measurement apparatus," *Physical Review Letters*, vol. 96, no. 20, p. 200401, 2006.

[27] G. M. D'Ariano, P. L. Presti, and M. G. Paris, "Using entanglement improves the precision of quantum measurements," *Physical Review Letters*, vol. 87, no. 27, p. 270404, 2001.

[28] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Memory Effects in Quantum Channel Discrimination," *Physical Review Letters*, vol. 101, p. 180501, oct 2008.

[29] Z. Puchała, Ł. Pawela, A. Krawiec, and R. Kukulski, "Strategies for optimal single-shot discrimination of quantum measurements," *Physical Review A*, vol. 98, p. 042103, 2018.

[30] M. Sedlák and M. Ziman, "Optimal single-shot strategies for discrimination of quantum measurements," *Physical Review A*, vol. 90, no. 5, p. 052312, 2014.

[31] D. Dieks, "Overlap and distinguishability of quantum states," *Physics Letters A*, vol. 126, no. 5, pp. 303 – 306, 1988.

[32] M. Hayashi, "Discrimination of two channels by adaptive methods and its application to quantum system," *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3807–3820, 2009.

[33] C. W. Helstrom, *Quantum Detection and Estimation Theory*. Elsevier, 1976.

[34] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Theoretical framework for quantum networks," *Physical Review A*, vol. 80, no. 2, p. 022339, 2009.

[35] A. Bisio, G. Chiribella, G. D'Ariano, and P. Perinotti, "Quantum networks: general theory and applications," *Acta Physica Slovaca. Reviews and Tutorials*, vol. 61, no. 3, pp. 273–390, 2011.

[36] K. Korzekwa, S. Czachórski, Z. Puchała, and K. Życzkowski, "Coherifying quantum channels," *New Journal of Physics*, vol. 20, no. 4, p. 043028, 2018.

[37] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Memory effects in quantum channel discrimination," *Physical Review Letters*, vol. 101, no. 18, p. 180501, 2008.

[38] I. Bengtsson and K. Życzkowski, *Geometry of quantum states: an introduction to quantum entanglement*. Cambridge University Press, 2017.

[39] K. Życzkowski and H.-J. Sommers, "Truncations of random unitary matrices," *Journal of Physics A: Mathematical and General*, vol. 33, no. 10, p. 2045, 2000.

[40] R. Piziak, P. Odell, and R. Hahn, "Constructing projections on sums and intersections," *Computers & Mathematics with Applications*, vol. 37, no. 1, pp. 67–74, 1999.

[41] M. Oszmaniec, L. Guerini, P. Wittek, and A. Acín, "Simulating Positive-Operator-Valued Measures with Projective Measurements," *Physical Review Letters*, vol. 119, p. 190501, Nov 2017.

[42] L. Guerini, J. Bavaresco, M. Terra Cunha, and A. Acín, "Operational framework for quantum measurement simulability," *Journal of Mathematical Physics*, vol. 58, no. 9, p. 092102, 2017.

[43] F. Hausdorff, "Der Wertvorrat einer Bilinearform," *Mathematische Zeitschrift*, vol. 3, no. 1, pp. 314–316, 1919.

[44] O. Toeplitz, "Das algebraische Analogon zu einem Satze von Fejér," *Mathematische Zeitschrift*, vol. 2, no. 1-2, pp. 187–197, 1918.

[45] G. Jaeger and A. Shimony, "Optimal distinction between two non-orthogonal quantum states," *Physics Letters A*, vol. 197, no. 2, pp. 83–87, 1995.

A Proof of Theorem 1

In this appendix we will begin with quoting two lemmas from [29] which contribute the main part of the proof of Theorem 1. The first lemma states that a function $|\operatorname{Tr}(\rho U E)|$ has a saddle point. The second lemma gives an equivalent condition to the existence of the saddle point and presents the form of the optimal state. Then, we present the proof of Theorem 1 and a short discussion about the amount of entanglement needed for the discrimination.

Lemma 1 (Lemma 4 from [29])**.** *Let us fix a unitary matrix $U \in \mathcal{U}_d$. Then,*

$$\max_{E\in\mathcal{DU}_d} \min_{\rho\in\mathcal{D}(\mathbb{C}^d)} |\operatorname{Tr}(\rho U E)| = \min_{\rho\in\mathcal{D}(\mathbb{C}^d)} \max_{E\in\mathcal{DU}_d} |\operatorname{Tr}(\rho U E)|. \tag{27}$$

Lemma 2 (Lemma 5 from [29])**.** *Let us fix a unitary matrix $U \in \mathcal{U}_d$ and*

- *$E_0 \in \mathcal{DU}_d$ and $D(E) = \min_{\rho\in\mathcal{D}(\mathbb{C}^d)} |\operatorname{Tr}\rho U E|$,*
- *$D(E_0) > 0$,*
- *λ_1, λ_d denote the most distant pair of eigenvalues of UE_0*
- *P_1, P_d denote the projectors on the subspaces spanned by the eigenvectors corresponding to λ_1, λ_d.*

Then, the function $\mathcal{D}(\mathbb{C}^d) \times \mathcal{DU}_d \ni (\rho, E) \mapsto |\operatorname{Tr}(\rho U E)|$ has a saddle point in (ρ_0, E_0) i.e.

$$\max_{E\in\mathcal{DU}_d} \min_{\rho\in\mathcal{D}(\mathbb{C}^d)} |\operatorname{Tr}(\rho U E)| = |\operatorname{Tr}(\rho_0 U E_0)| = \min_{\rho\in\mathcal{D}(\mathbb{C}^d)} \max_{E\in\mathcal{DU}_d} |\operatorname{Tr}(\rho U E)| \tag{28}$$

if and only if there exist states ρ_1, ρ_d such that

- *$\rho_1 = P_1 \rho_1 P_1$,*
- *$\rho_d = P_d \rho_d P_d$,*
- *$\operatorname{diag}(\rho_1) = \operatorname{diag}(\rho_d)$.*

Moreover, if the above holds, then the state ρ_0 satisfying Eq. (28) can be chosen as $\frac{1}{2}\rho_1 + \frac{1}{2}\rho_d$.

Proof of Theorem 1. In the first step of the proof assume that $\mathcal{P}_U$ and $\mathcal{P}_{\mathbf{1}}$ are perfectly distinguishable in a single-shot scenario i.e.

$$\|\mathcal{P}_U - \mathcal{P}_{\mathbf{1}}\|_\diamond = \min_{E\in\mathcal{DU}_d} \|\Phi_{UE} - \Phi_{\mathbf{1}}\|_\diamond = 2. \tag{29}$$

This trivially implies that for each $N \in \mathbb{N}$ the measurements $\mathcal{P}_{U^{\otimes N}}$ and $\mathcal{P}_{\mathbf{1}}$ are perfectly distinguishable and it holds that

$$\begin{aligned} 2 = \|\mathcal{P}_{U^{\otimes N}} - \mathcal{P}_{\mathbf{1}}\|_\diamond &= \min_{F\in\mathcal{DU}_{d^N}} \left\|\Phi_{(U^{\otimes N})F} - \Phi_{\mathbf{1}}\right\|_\diamond \\ &\le \min_{E\in\mathcal{DU}_d} \|\Phi_{U^{\otimes N} E^{\otimes N}} - \Phi_{\mathbf{1}}\|_\diamond \le 2, \end{aligned} \tag{30}$$

which proves the thesis of the theorem in this case.

Now, consider the second case when $\mathcal{P}_U$ and $\mathcal{P}_{\mathbf{1}}$ are not perfectly distinguishable using a single query. Then, according to equality

$$\|\mathcal{P}_U - \mathcal{P}_{\mathbf{1}}\|_\diamond = \min_{E\in\mathcal{DU}_d} \|\Phi_{UE} - \Phi_{\mathbf{1}}\|_\diamond \tag{31}$$

there exists an optimal matrix $E_0 \in \mathcal{DU}_d$ such that $0 \notin W(UE_0)$ and $\|\mathcal{P}_U - \mathcal{P}_{\mathbf{1}}\|_\diamond = \|\Phi_{UE_0} - \Phi_{\mathbf{1}}\|_\diamond$.

The general proof strategy is to utilize the Lemma 2 in order to construct optimal discriminator ρ_0 for measurements $\mathcal{P}_{U^{\otimes N}}$ and $\mathcal{P}_{\mathbf{1}}$. Hence we start by establishing that the assumptions of this lemma are fulfilled. This will follow from Lemma 1. According to it the function $(\rho, E) \mapsto |\operatorname{Tr}(\rho U E)|$ has a saddle point. Let us remind that $\|\Phi_U - \Phi_{\mathbf{1}}\|_\diamond = 2\sqrt{1-\nu^2}$, where $\nu = \min_{x\in W(U)} |x|$. Due to equality $\min_{E\in\mathcal{DU}_d} \|\Phi_{UE} - \Phi_{\mathbf{1}}\|_\diamond = \|\Phi_{UE_0} - \Phi_{\mathbf{1}}\|_\diamond$ we have

$$\max_{E\in\mathcal{DU}_d} \min_{x\in W(UE)} |x| = \min_{x\in W(UE_0)} |x|. \tag{32}$$

Hence using the property that the numerical range is a convex set [43, 44] we obtain

$$\max_{E\in\mathcal{DU}_d} \min_{\rho\in\mathcal{D}(\mathbb{C}^d)} |\operatorname{Tr}(\rho U E)| = \min_{\rho\in\mathcal{D}(\mathbb{C}^d)} |\operatorname{Tr}(\rho U E_0)|. \tag{33}$$

Therefore the former assumptions of Lemma 2 are satisfied for the matrix E_0 and hence it is possible to take states ρ_1, ρ_d which fulfill the latter conditions of this Lemma.

To complete the proof we will separately study two cases. In the first one, we assume that POVMs $\mathcal{P}_U$ and $\mathcal{P}_{\mathbf{1}}$ are not perfectly distinguishable using N queries i.e. $0 \notin W\left(U^{\otimes N} E_0^{\otimes N}\right)$. Hence, as $\operatorname{diag}(\rho_1) = \operatorname{diag}(\rho_d)$, then

$$\operatorname{diag}(\rho_1^{\otimes N}) = \operatorname{diag}(\rho_d^{\otimes N}) \tag{34}$$

and $\rho_1^{\otimes N}, \rho_d^{\otimes N}$ lie on the subspaces spanned by the eigenvectors of the matrix $U^{\otimes N} E_0^{\otimes N}$ eigenvalues λ_1^N and λ_d^N, respectively. Consequently, we fulfilled the latter assumptions of Lemma 2 and the reverse implication of this Lemma states that the unitary matrix $E_0^{\otimes N}$ is optimal and for $\rho_0 = \frac{1}{2}\rho_1^{\otimes N} + \frac{1}{2}\rho_d^{\otimes N}$ it holds that

$$\begin{aligned} \min_{\rho\in\mathcal{D}(\mathbb{C}^d)} \left|\operatorname{tr}\left(\rho (UE_0)^{\otimes N}\right)\right| &= \left|\operatorname{tr}\left(\rho_0 (UE_0)^{\otimes N}\right)\right| \\ &= \max_{F\in\mathcal{DU}_{d^N}} \min_{\rho\in\mathcal{D}(\mathbb{C}^{d^N})} \left|\operatorname{tr}\left(\rho U^{\otimes N} F\right)\right|. \end{aligned} \tag{35}$$

Hence

$$\left\|\Phi_{U^{\otimes N} E_0^{\otimes N}} - \Phi_{\mathbf{1}}\right\|_\diamond = \min_{F\in\mathcal{DU}_{d^N}} \|\Phi_{U^{\otimes N} F} - \Phi_{\mathbf{1}}\|_\diamond \le \min_{E\in\mathcal{DU}_d} \|\Phi_{U^{\otimes N} E^{\otimes N}} - \Phi_{\mathbf{1}}\|_\diamond \tag{36}$$

and eventually

$$\|\mathcal{P}_{U^{\otimes N}} - \mathcal{P}_{\mathbf{1}}\|_\diamond = \min_{E\in\mathcal{DU}_d} \|\Phi_{U^{\otimes N} E^{\otimes N}} - \Phi_{\mathbf{1}}\|_\diamond. \tag{37}$$

In the second case, let $0 \in W\left(U^{\otimes M} E_0^{\otimes M}\right)$. Let us consider the situation when M is the first index for which this happens, which means that $0 \notin W\left(U^{\otimes M-1} E_0^{\otimes M-1}\right)$.

If the measurements in question can be perfectly distinguished with M queries, then we have $0 \in \operatorname{conv}(\lambda_1^M, \lambda_1\lambda_d^{M-1}, \lambda_d^M)$ and there exists a probability vector $p = (p_1, p_2, p_3)$ such that

$$p_1\lambda_1^M + p_2\lambda_1\lambda_d^{M-1} + p_3\lambda_d^M = 0. \tag{38}$$

Define a state

$$\rho = p_1\rho_1^{\otimes M} + p_2\left(\rho_1 \otimes \rho_d^{\otimes M-1}\right) + p_3\rho_d^{\otimes M}. \tag{39}$$

We will show that $\operatorname{diag}\left(\rho U^{\otimes M}\right) = 0$. Indeed

$$\begin{aligned} &\operatorname{diag}\left(\rho U^{\otimes M} E_0^{\otimes M}\right) \\ =&\operatorname{diag}\left(p_1\lambda_1^M\rho_1^{\otimes M} + p_2\lambda_1\lambda_d^{M-1}\left(\rho_1\otimes\rho_d^{\otimes M-1}\right) + p_3\lambda_d^M\rho_d^{\otimes M}\right) \\ =&p_1\lambda_1^M \operatorname{diag}(\rho_1^{\otimes M}) + p_2\lambda_1\lambda_d^{M-1}\operatorname{diag}\left(\rho_1\otimes\rho_d^{\otimes M-1}\right) + p_3\lambda_d^M\operatorname{diag}(\rho_d^{\otimes M}) \\ =&\left(p_1\lambda_1^M + p_2\lambda_1\lambda_d^{M-1} + p_3\lambda_d^M\right)\operatorname{diag}(\rho_1^{\otimes M}) = 0. \end{aligned} \tag{40}$$

Thus from Proposition 3 form [29] the condition $\operatorname{diag}\left(\rho U^{\otimes M} E_0^{\otimes M}\right) = 0$ implies that $\|\mathcal{P}_{U^{\otimes M}} - \mathcal{P}_{\mathbf{1}}\|_\diamond = 2$, and hence

$$\|\mathcal{P}_{U^{\otimes M}} - \mathcal{P}_{\mathbf{1}}\|_\diamond = \min_{E \in \mathcal{DU}_d} \|\Phi_{U^{\otimes M} E^{\otimes M}} - \Phi_{\mathbf{1}}\|_\diamond . \tag{41}$$

For $N > M$, the equality $\|\mathcal{P}_{U^{\otimes M}} - \mathcal{P}_{\mathbf{1}}\|_\diamond = 2$ implies that $\|\mathcal{P}_{U^{\otimes N}} - \mathcal{P}_{\mathbf{1}}\|_\diamond = 2$. Therefore

$$\|\mathcal{P}_{U^{\otimes N}} - \mathcal{P}_{\mathbf{1}}\|_\diamond = \min_{E \in \mathcal{DU}_d} \|\Phi_{U^{\otimes N} E^{\otimes N}} - \Phi_{\mathbf{1}}\|_\diamond , \tag{42}$$

which completes the proof. □

Let us discuss the amount of the discriminator's entanglement with environment. Note that the minimal dimension of an auxiliary system needed for optimal discrimination is equal to the rank of the state ρ_0. One special case involves the situation $\|\mathcal{P}_{U^{\otimes N}} - \mathcal{P}_{\mathbf{1}}\|_\diamond < 2$, where $\rho_0 = \frac{1}{2}\rho_1^{\otimes N} + \frac{1}{2}\rho_d^{\otimes N}$. In the best case, when eigenvalues λ_1 and λ_d are not degenerated, the states ρ_1, ρ_d are pure, hence $\operatorname{rank}(\rho_0) = 2$. On the other hand, when the spectrum of UE_0 contains only λ_1 and λ_d, then the rank of ρ_0 can be roughly upper-bounded by $(d-1)^N + 1$.

The situation $\|\mathcal{P}_{U^{\otimes N}} - \mathcal{P}_{\mathbf{1}}\|_\diamond = 2$ is more complicated to analyze, due to lack of an analytic form of the discriminator in a general case. However, finding a pair of not degenerated eigenvalues λ_1, λ_d saturating the latter assumptions of Lemma 2 will lead to $\operatorname{rank}(\rho_0) = 3$, where ρ_0 is defined as in Eq. (39) and we check its optimality in the same manner as in Eq. (40).

B Explicit form of the matrix A_U

The matrix A_U is a general operation which allows for adaptive information processing in the sequential discrimination scenario of von Neumann measurements. It consists of a sequence of unitary matrices U acting successively on given registers interlacing with classically controlled unitary operations V^i. The explicit form of the matrix A_U is given by

$$\begin{aligned} A_U = & (\mathbb{1}_{1,\ldots N-1} \otimes U \otimes \mathbb{1}_{N+1}) \\ & \left(\sum_{i_1,\ldots,i_{N-1}} |i_1,\ldots,i_{N-1}\rangle\langle i_1,\ldots,i_{N-1}| \otimes V^{(N-1)}_{i_1,\ldots,i_{N-1}} \right) \\ & (\mathbb{1}_{1,\ldots,N-2} \otimes U \otimes \mathbb{1}_{N,N+1}) \\ & \left(\sum_{i_1,\ldots,i_{N-2}} |i_1,\ldots,i_{N-2}\rangle\langle i_1,\ldots,i_{N-2}| \otimes V^{(N-2)}_{i_1,\ldots,i_{N-2}} \right) \\ & \ldots \\ & (\mathbb{1}_1 \otimes U \otimes \mathbb{1}_{3,\ldots N+1}) \\ & \left(\sum_{i_1} |i_1\rangle\langle i_1| \otimes V^{(1)}_{i_1} \right) \\ & (U \otimes \mathbb{1}_{2,\ldots N+1}) . \end{aligned} \tag{43}$$

C Proof of Theorem 4

Proof of Theorem 4. Assume that a black-box measurement ($\mathcal{P}_{\mathbf{1}}$ or $\mathcal{P}_U$) acts on a system extended by the ancilla space $\mathcal{H}_B$ (of some dimension d_1). Without loss of generality we take the pure input state i.e. $\sigma = |\psi_{AB}\rangle\langle\psi_{AB}|$. Let X be a matrix such that $|\psi_{AB}\rangle = \sum_{i,j=1}^{d,d_1} X_{ij}|i\rangle|j\rangle$. The action of the channels $\mathcal{P}_{\mathbf{1}} \otimes \mathbb{1}_B$ and $\mathcal{P}_U \otimes \mathbb{1}_B$ on $|\psi_{AB}\rangle\langle\psi_{AB}|$ can be expressed as

$$
\begin{aligned}
(\mathcal{P}_{\mathbf{1}} \otimes \mathbb{1}_B)\left(|\psi_{AB}\rangle\langle\psi_{AB}|\right) &= \sum_{i=1}^{d} |i\rangle\langle i| \otimes X^T|i\rangle\langle i|\overline{X}, \\
(\mathcal{P}_U \otimes \mathbb{1}_B)\left(|\psi_{AB}\rangle\langle\psi_{AB}|\right) &= \sum_{i=1}^{d} |i\rangle\langle i| \otimes X^T\overline{U}|i\rangle\langle i|U^T\overline{X},
\end{aligned} \tag{44}
$$

where we treat measurements $\mathcal{P}_{\mathbf{1}}$ and $\mathcal{P}_U$ as a measure and prepare channels of the form $\Psi_{\mathcal{M}}(\sigma) = \sum_{i=1}^{n} \operatorname{tr}(M_i\sigma)|i\rangle\langle i|$.

As for any Hermitian operator M and any measurement $\mathcal{P}$ we have

$$
\begin{aligned}
\operatorname{tr}\left(M(\mathcal{P} \otimes \mathbb{1})(\sigma)\right) &= \operatorname{tr}\left(M(\Delta\mathcal{P} \otimes \mathbb{1})(\sigma)\right) \\
&= \operatorname{tr}\left((\Delta \otimes \mathbb{1})(M)(\mathcal{P} \otimes \mathbb{1})(\sigma)\right),
\end{aligned} \tag{45}
$$

where Δ is a dephasing channel. Hence we can restrict our attention to considering measurements $\mathcal{M}$ which effects have block-diagonal structure, that is

$$
\mathcal{M} = \sum_{i=1}^{d} |i\rangle\langle i| \otimes \mathcal{T}_i, \tag{46}
$$

where $\mathcal{T}_i$ is a POVM on $\mathcal{H}_B$ associated with a measure and prepare channel. From Eq.(44) we see that upon obtaining the label i, the state of the auxiliary subsystem is either

$$
|x_i\rangle\langle x_i| = p_i^{-1} X^\top |i\rangle\langle i|\overline{X}, \tag{47}
$$

when measurement $\mathcal{P}_{\mathbf{1}}$ was performed, or it is given by

$$
|y_i\rangle\langle y_i| = q_i^{-1} X^\top \overline{U}|i\rangle\langle i|U^T\overline{X} \tag{48}
$$

if $\mathcal{P}_U$ was implemented. In the above formulas p_i, q_i are responsible for normalization. We assume that $p_i > 0$ and $q_i > 0$ (otherwise the specific outcome i does not occur). We see that states $|x_i\rangle\langle x_i|$, $|y_i\rangle\langle y_i|$ are pure and therefore the optimal measurements $\mathcal{T}_i = \{T_1^{(i)}, T_2^{(i)}, T_?^{(i)}\}$ will be simply given by

$$
\begin{aligned}
T_1^{(i)} &= \gamma_1(\mathbb{1} - |y_i\rangle\langle y_i|), \\
T_2^{(i)} &= \gamma_2(\mathbb{1} - |x_i\rangle\langle x_i|), \\
T_?^{(i)} &= \mathbb{1} - T_1 - T_2,
\end{aligned} \tag{49}
$$

for some choice of $\gamma_{1,2}$ which guarantees the non-negativity of $T_?^{(i)}$.

The probability of success in unambiguous discrimination of pure states $|x\rangle, |y\rangle$ with unequal a priori probabilities $\eta, 1-\eta$ is given by [45]

$$
p_{succ}^{u}(x, y, \eta) = \begin{cases} 1 - \eta - (1-\eta)c^2 & \text{for } \eta < \frac{c^2}{1+c^2} \\ 1 - 2c\sqrt{\eta(1-\eta)} & \text{for } \frac{c^2}{1+c^2} \leq \eta \leq \frac{1}{1+c^2} \\ 1 - (1-\eta) - \eta c^2 & \text{for } \frac{1}{1+c^2} < \eta, \end{cases} \tag{50}
$$

where $c = |\langle x|y\rangle|$.

We will use the following upper bound

$$
p_{succ}^{u}(x, y, \eta) \leq 1 - 2c\sqrt{\eta(1-\eta)}, \tag{51}
$$

which can be verified directly by elementary calculations.

Let $\rho = XX^\dagger$. The overlap c_i between states of the auxiliary subsystem is given by

$$c_i = |\langle x_i|y_i\rangle| = |\langle i|\overline{X}X^\top\overline{U}|i\rangle|/\sqrt{p_iq_i} = |\langle i|\rho U|i\rangle|/\sqrt{p_iq_i}, \tag{52}$$

while a priori probabilities of $|x_i\rangle, |y_i\rangle$ upon obtaining label i are $\eta_i = \frac{p_i}{p_i+q_i}, 1-\eta_i = \frac{q_i}{p_i+q_i}$. Taking the above into account, we get that probability of success in unambiguous measurement on auxiliary subsystem, given that label i was observed, can be bounded from above by

$$p^u_{succ}(x_i, y_i, \eta_i) \leq 1 - 2c_i\frac{\sqrt{p_iq_i}}{p_i+q_i} = 1 - \frac{2|\langle i|\rho U|i\rangle|}{p_i+q_i}. \tag{53}$$

Therefore, the overall probability of success is bounded by

$$\begin{aligned} p_\mathrm{u}(\mathcal{P}_\mathbf{1}, \mathcal{P}_U) &= \max_{|\psi_{AB}\rangle}\sum_i \Pr(\text{label}=i)p^u_{succ}(x_i,y_i,\eta_i) \\ &\leq \max_\rho \sum_i \frac{1}{2}(p_i+q_i)\left(1-\frac{2|\langle i|\rho U|i\rangle|}{p_i+q_i}\right) \\ &= 1 - \min_\rho\sum_i |\langle i|\rho U|i\rangle|. \end{aligned} \tag{54}$$

What is more, the above bound is tight. The situation when $\mathcal{P}_\mathbf{1}$ and $\mathcal{P}_U$ are perfectly distinguishable is trivial to check. In the case when $\mathcal{P}_\mathbf{1}$ and $\mathcal{P}_U$ are not perfectly distinguishable, then there exists a state ρ which will give equal probabilities p_i and q_i for each label i. This statement follows from Lemma 2, from which we take $\rho = \rho_0$. □

D Proof of Theorem 5

Proof of Theorem 5. We will assume that the unitary matrix U is optimal, *i.e.* $\Upsilon(U) = \Theta(U)$. Take an arbitrary input state $\sigma = |\psi_{A,B}\rangle\langle\psi_{A,B}|$. Let us denote

$$\begin{aligned} |x_i\rangle &= p_i^{-1/2}\left(\langle i|\otimes \mathbb{1}_{N+1}\right)A_\mathbf{1}|\psi_{A,B}\rangle \\ |y_i\rangle &= q_i^{-1/2}\left(\langle i|\otimes \mathbb{1}_{N+1}\right)A_U|\psi_{A,B}\rangle, \end{aligned} \tag{55}$$

where A_U and $A_\mathbf{1}$ are defined as in Appendix B and p_i, q_i are responsible for normalization. Repeating the calculation from the single-shot scenario from the proof in Appendix C we can upper-bound the probability of successful discrimination as follows

$$\begin{aligned} p_\mathrm{u}(\Psi_U, \Psi_\mathbf{1}) &\leq 1 - \min_{|\psi_{A,B}\rangle}\sum_i\left|\langle\psi_{A,B}|A_\mathbf{1}^\dagger\left(|i\rangle\langle i|\otimes\mathbb{1}_{N+1}\right)A_U|\psi_{A,B}\rangle\right| \\ &\leq 1 - \min_{|\psi_{A,B}\rangle}\left|\sum_i\langle\psi_{A,B}|A_\mathbf{1}^\dagger\left(|i\rangle\langle i|\otimes\mathbb{1}_{N+1}\right)A_U|\psi_{A,B}\rangle\right| \\ &= 1 - \min_{|\psi_{A,B}\rangle}\left|\langle\psi_{A,B}|A_\mathbf{1}^\dagger A_U|\psi_{A,B}\rangle\right|. \end{aligned} \tag{56}$$

From the work [37] we know that there exists a state $|\phi\rangle$ such that for all $|\psi_{A,B}\rangle$ it holds that

$$|\langle\psi_{A,B}|A_\mathbf{1}^\dagger A_U|\psi_{A,B}\rangle| \geq |\langle\phi|U^{\otimes N}|\phi\rangle|. \tag{57}$$

Moreover, using optimality of U and Lemma 1, the state $|\phi\rangle$ can be chosen to satisfy $|\langle\phi|U^{\otimes N}|\phi\rangle| = \min_\rho\sum_i|\langle i|\rho U^{\otimes N}|i\rangle|$. This leads to the desired inequality

$$p_\mathrm{u}(\Psi_U, \Psi_\mathbf{1}) \leq 1 - \min_\rho\sum_i|\langle i|\rho U^{\otimes N}|i\rangle|. \tag{58}$$

□

E Proof of Theorem 6

Proof of Theorem 6. Let us fix the input state $\sigma \in \mathcal{D}(\mathbb{C}^d)$. Without loss of generality we can restrict our attention to measurements $\mathcal{M}$ with diagonal effects (see Eq. (45) in Appendix C).

From the unambiguity condition we obtain that $M_{\mathbf{1}} \perp \operatorname{supp}(\mathcal{P}_U(\sigma))$ and $M_U \perp \operatorname{supp}(\mathcal{P}_{\mathbf{1}}(\sigma))$. Therefore, the optimal measurements can be always chosen as projectors onto disjoint subsets Γ, Λ of $\{1, \ldots, d\}$. The formula for the success probability reads

$$\tilde{p}_{\mathrm{u}}\left(\mathcal{P}_{\mathbf{1}}, \mathcal{P}_U; \sigma, \Gamma, \Lambda\right) = \frac{1}{2}\operatorname{tr}(P_\Gamma \sigma) + \frac{1}{2}\operatorname{tr}(Q_\Lambda \sigma). \tag{59}$$

Importantly, the input state σ satisfies $\sigma \perp P_\Lambda$ and $\sigma \perp Q_\Gamma$. For fixed subsets Λ, Γ, due to linearity, the maximum over σ equals $\|\mathbb{P}_{\Gamma,\Delta}(P_\Gamma + Q_\Lambda)\mathbb{P}_{\Gamma,\Delta}\|$, where $\|\cdot\|$ denotes the operator norm and $\mathbb{P}_{\Gamma,\Delta}$ is the orthogonal projector onto $\operatorname{Span}\left(\{U|i\rangle\}_{i\in\Gamma^c}\right) \cap \operatorname{Span}\left(\{|j\rangle\}_{j\in\Lambda^c}\right)$. By optimizing over disjoint subsets $\Lambda, \Gamma \subset \{1, \ldots, d\}$ we obtain the result.

□