

Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography

Filip B. Maciejewski^{1,2,3}, Zoltán Zimborás^{4,5,6}, and Michał Oszmaniec^{2,3}

¹ University of Warsaw, Faculty of Physics, Ludwika Pasteura 5, 02-093 Warszawa, Poland

² International Centre for Theory of Quantum Technologies, University of Gdansk, Wita Stwosza 63, 80-308 Gdansk, Poland

³ Center for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warszawa, Poland

⁴ Wigner Research Centre for Physics of the Hungarian Academy of Sciences, H-1525 Budapest, P.O.Box 49, Hungary

⁵ BME-MTA Lendület Quantum Information Theory Research Group, Budapest, Hungary

⁶ Mathematical Institute, Budapest University of Technology and Economics, P.O.Box 91, H-1111, Budapest, Hungary

We propose a simple scheme to reduce readout errors in experiments on quantum systems with finite number of measurement outcomes. Our method relies on performing classical post-processing which is preceded by Quantum Detector Tomography, i.e., the reconstruction of a Positive-Operator Valued Measure (POVM) describing the given quantum measurement device. If the measurement device is affected only by an invertible classical noise, it is possible to correct the outcome statistics of future experiments performed on the same device. To support the practical applicability of this scheme for near-term quantum devices, we characterize measurements implemented in IBM's and Rigetti's quantum processors. We find that for these devices, based on superconducting transmon qubits, classical noise is indeed the dominant source of readout errors. Moreover, we analyze the influence of the presence of coherent errors and finite statistics on the performance of our error-mitigation procedure. Applying our scheme on the IBM's 5-qubit device, we observe a significant improvement of the results of a number of single- and two-qubit tasks including Quantum State Tomography (QST), Quantum Process Tomography (QPT), the implementation of non-projective measurements, and certain quantum algorithms (Grover's search and the Bernstein-Vazirani algorithm). Finally, we present results showing improvement for the implementation of certain probability distributions in the case of five qubits.

1 Introduction

In recent years, quantum technologies have been rapidly developing. Scientists and engineers around the world share the hope and fascination caused by the possibility of creating devices that would allow for the manipulation

Filip B. Maciejewski: filip.b.maciejewski@gmail.com

Zoltán Zimborás: zimboras@gmail.com

Michał Oszmaniec: michal.oszmaniec@gmail.com

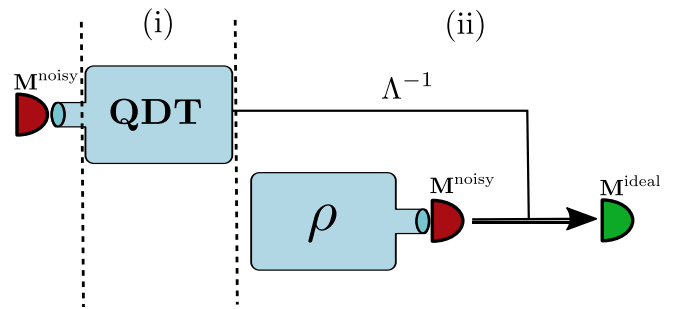


Figure 1: Pictorial representation of our error-mitigation procedure. (i) In the first stage, one performs the tomography of a noisy detector M^{noisy} (red semicircle). (ii) In the next stage, when measuring an arbitrary quantum state ρ , one employs a post-processing procedure on the measured statistics through the application of Λ^{-1} , the inverse of a stochastic noise map obtained in the QDT. This gives access to the statistics that would have been obtained in an ideal detector M^{ideal} (green semicircle).

of delicate quantum states with unprecedented precision [1]. Due to the advent of quantum cloud services (IBM [2, 3], Rigetti [4], DWave [5]), any researcher has a possibility to perform experiments on actual quantum devices. However, if one really hopes for utilizing such near-term devices for real-life applications such as quantum computation [6], quantum simulations [7] or generating random numbers [8], experimental imperfections must be taken into account. Hence, to properly characterize noise occurring in the devices and to develop error correction and mitigation schemes that may help to fight it have become tasks of fundamental importance [9, 10, 11, 12, 13, 14, 15]. In the present work, we address this problem for the noise affecting quantum measurements.

In theoretical considerations about quantum information tasks, a quantum device is often assumed to perform perfect measurements. In practice, this assumption may be significantly violated due to experimental imperfections and noise. Specifically, a broad class of errors occurring in quantum devices may be described as State Preparation And Measurement (SPAM) errors [16, 17]. It follows that if state preparation errors are negligible, one is able to reconstruct a POVM (Positive-Operator Valued Measure) associated with a given measurement device,

and therefore to infer measurement errors. A standard method of such reconstruction is to perform Quantum Detector Tomography (QDT) [18]. Results of the QDT may be used to compare the reconstructed POVM with the ideal one [18] and to analyze the nature of measurement errors [19, 20, 21, 22]. In this work, we propose to do more – we show how to use the knowledge about the POVM associated with a measurement device to correct the results obtained in further experiments performed on this device. Such a correction is possible provided the noise which affects the measurement is of a classical stochastic type and non-degenerate (invertible). In other words, such type of noise is equivalent to invertible classical post-processing of experimental statistics, and it may also be inverted solely by classical post-processing. The general idea of mitigating the effects of such a noise is the following. Assuming that one has access to the aforementioned invertible stochastic map describing a noise, it is straightforward to use the (generally non-stochastic) inversion of this map to classically reverse effects of noise simply by multiplying the vector of experimental statistics (obtained in a further experiment) by this inverted matrix, see Fig. 1. The main aim of this work is to present such a classical correction scheme together with the analysis of its accuracy. Specifically, we show how the deviations from classical noise model and finite-size statistics affect our procedure, by providing upper bounds for the distance of the corrected probability distribution from the ideal noise-reversed scenario.

Besides the theoretical description, we also test our procedure experimentally. First, we present data that suggests that IBM’s and Rigetti’s quantum processors are affected by a significant readout noise that can be described (to a good approximation) by the classical model characterized above. Both architectures consist of superconducting transmon qubits [23], which may suggest that the classical noise may be the dominant form of measurement noise in such devices. Second, we test the correction scheme on the IBM’s five-qubit device *ibmqx4* and conclude that indeed it significantly compensates for the effects of this noise for several one- and two-qubit experiments, including Quantum State Tomography (QST) [6], Quantum Process Tomography (QPT) [6], the implementation of non-projective measurements [24], and two quantum algorithms (Grover’s search [25] and the Bernstein-Vazirani algorithm [26]). Furthermore, we test our scheme for the 5-qubit experiments concerning the implementation of certain probability distributions.

We will now comment on related works. There is a variety of recent research concerning the mitigation of different types of errors in contemporary quantum devices [9, 10, 11, 12, 13, 14, 15]. In particular, in the Appendix of work [12], authors assume the uncorrelated, classical noise model for readout errors and state that they use measurement calibration data to correct ‘assignment errors’. We believe that this might refer to procedure-related to ours (under aforementioned assumptions). Similarly, in work [11] authors mention in the Appendix that they use ‘readout calibration’ to correct expectation values for assignment infidelity.

Quantum Detector Tomography for superconducting

qubits, along with other characterizations of measurement errors, was presented in the Ref. [22] for Rigetti devices. Characterization of SPAM errors for IBM and Rigetti devices has been presented in Ref. [17].

The paper is organized as follows. In Section 2, we present the necessary theoretical concepts, including the POVMs formalism, the QDT scheme and the description of classical noise. Section 3 is devoted to the main idea of this work; it consists of a detailed description of the statistics correction procedure, which is preceded by stating the necessary assumptions. In Section 4, we analyze how violations of the assumptions affect our correction procedure. Section 5 contains a summary of the correction procedure in the form of pictorial representation. The scheme presents practical steps that need to be done in the case of noisy projective measurement in the computational basis. In Section 6, we present experimental results from IBM’s and Rigetti’s devices that provide insight into the magnitude of SPAM errors. Section 7 consist of experimental results from applications of our correction scheme for exemplary quantum information tasks in IBM’s five-qubit device. Finally, we present the conclusions and some possible future research directions in Section 8.

2 Theoretical background

In this section, the necessary theoretical background and mathematical tools are shortly reviewed. First, we define the notion of Positive-Operator Valued Measures, which are useful tools for modeling measurement noise in quantum devices. Second, we describe a Quantum Detector Tomography procedure, which allows reconstructing a POVM associated with a given device. Then we discuss a measure of the distance between quantum measurements known as the operational distance. Finally, we use the formalism of POVMs to precisely define classical noise affecting measurements, which will be used in Section 3 to derive our correction procedure.

2.1 Mathematical description of quantum measurements

We start by a mathematical description of a generalized quantum measurement modeled by a Positive-Operator Valued Measure [27]. A POVM with n outcomes on a d -dimensional Hilbert space may be described by an n -element vector \mathbf{M} of operators M_i (called *effects*), such that

$$\mathbf{M} = (M_1, M_2, \dots, M_n)^T, \quad \forall i \ M_i \geq 0, \quad \sum_{i=1}^n M_i = \mathbb{1}, \quad (1)$$

where M_i ’s are represented by $d \times d$ semi-definite positive matrices, $\mathbb{1}$ is the operator identity and T denotes transposition which we use because the column form of POVMs will be useful later. If a quantum system was initially described by a state ρ , the probability of obtaining the outcome associated with the effect M_i is given by

$$p(i|\rho) = \text{Tr}(\rho M_i). \quad (2)$$

In other words, probabilities of obtaining particular outcomes are equal to the expectation values of the associated effects. In quantum information protocols, the considered measurements are often *projective* [6], which means that the effects M_i fulfill the additional requirements $M_i M_j = \delta_{i,j} M_i$.

In the experimental part of this work, we will focus on projective measurements and their noisy versions, since this fits the set-up of the IBM Q devices. Nevertheless, all our theoretical considerations, including the correction procedure, are formulated for generalized measurements given by Eq. (1).

2.2 Quantum Detector Tomography

Characterization of quantum devices requires knowledge of the POVM associated with a given measurement performed by the device. The procedure for obtaining this is known as Quantum Detector Tomography (QDT) [18]. The general idea of QDT is as follows. Recall that Born's rule associates the probabilities of particular outcomes with the expectation values of effects. If one performs multiple experiments on the set of quantum states which form a basis for Hermitian operators, then one may use the obtained statistics to decompose all effects on that basis via Born's rule. In the case of a d -dimensional Hilbert space, such a basis is d^2 dimensional, and this is the minimal number of different experiments (different quantum states) that must be performed to reconstruct a POVM.

Naturally, in practice one will not have access to perfect statistics, nor to the perfectly prepared quantum states. This may cause that the simple tomographic method returns nonphysical, non-positive POVM elements. To cope with this issue, the method of Maximum Likelihood Estimation (MLE) may be used, which asserts the positivity of the reconstructed quantum measurements [28, 29]. In order to reconstruct POVMs from experimental data, we have thus implemented an iterative algorithm from [28] which converges to the MLE estimator.

2.3 Distances

After reconstructing the POVMs, it is useful to have a distance measure between them in order to compare the reconstructed quantum measurements with the ideal detectors and, naturally, to compare detectors with each other. Distances between quantum measurements are usually closely related to the probability distributions that they generate via Born's rule. Therefore, first, we need a measure of the distance between classical probability distributions. In this work, we will be using so called Total-Variation (TV) distance, which for two arbitrary probability vectors $\mathbf{p} = (p_1, \dots, p_n)^T$ and $\mathbf{q} = (q_1, \dots, q_n)^T$, TV distance between them is defined

as

$$D_{TV}(\mathbf{p}, \mathbf{q}) := \frac{1}{2} \|\mathbf{p} - \mathbf{q}\|_1 = \frac{1}{2} \sum_{i=1}^n |p_i - q_i|, \quad (3)$$

where $\|\cdot\|_1$ denotes l_1 norm. Total-variation is a very stringent measure of distance between probability distributions. It plays important role in proposals for attaining quantum computational advantage [30, 31] and can be also easily utilized to control expectation values of (classical) observables defined on the sample space of interest.

TV-distance is related to the *operational distance* [32, 33, 34] between arbitrary POVMs \mathbf{M} and \mathbf{N} via the equality

$$D_{op}(\mathbf{M}, \mathbf{N}) = \max_{\rho} D_{TV}(\mathbf{p}^{\mathbf{M}}, \mathbf{p}^{\mathbf{N}}), \quad (4)$$

where maximization goes over all quantum states ρ , and $\mathbf{p}^{\mathbf{M}/\mathbf{N}}$ denote probability distributions generated by measurement of quantum state via \mathbf{M}/\mathbf{N} , i.e.,

$$\mathbf{p}^{\mathbf{M}} = (\text{Tr}(\rho M_1), \dots, \text{Tr}(\rho M_n))^T. \quad (5)$$

Hence, the operational distance is in fact the worst-case scenario of the distance between probability distributions generated by performing those measurements. It can be shown [32] that operational distance may be calculated as

$$D_{op}(\mathbf{M}, \mathbf{N}) = \max_x \left\| \sum_{i \in x} (M_i - N_i) \right\|_{\infty}, \quad (6)$$

where the maximization is over all subsets of indices enumerating effects (i.e., all possible sets of outcomes) and $\|\cdot\|_{\infty}$ denotes operator norm¹. Operational distance between POVMs² has an interesting operational interpretation through the formula $D_{op}(\mathbf{M}, \mathbf{N}) = 2p_{\text{disc}}(\mathbf{M}, \mathbf{N}) - 1$, where $p_{\text{disc}}(\mathbf{M}, \mathbf{N})$ is the optimal probability of distinguishing between measurements \mathbf{M} and \mathbf{N} (without using entanglement) [33].

2.4 Classical noise affecting measurements

Now we will describe the model of classical noise that we are considering in this work. Let us denote by $\mathbf{M}^{\text{ideal}}$ an n -outcome quantum measurement that in theory should be associated with our measurement device. In practice, due to the presence of noise, *real* measurement describing our device is some POVM \mathbf{M}^{exp} . In this model, we assume that the relation between $\mathbf{M}^{\text{ideal}}$ and \mathbf{M}^{exp} is given by a (left) stochastic, invertible map Λ , whose element $\Lambda_{i,j} := p(i|j) \in [0, 1]$ are defined by equation

$$\forall i \ M_i^{\text{exp}} = \sum_j p(i|j) M_j^{\text{ideal}}. \quad (7)$$

¹The operator norm is defined in standard way, i.e., for the operator A acting on the vector space V , the operator norm is equal to $\|A\|_{\infty} = \sup_v \{\|Av\|_{\infty} : v \in V, \|v\|_{\infty} = 1\}$.

²We note that operational distance can be also be upper-bounded by another quantity used in the quantum information theory, namely the diamond norm of a difference of POVMs, which can be calculated in terms of Semidefinite Programming [35].

The left-stochastic property of Λ means that $\sum_i \Lambda_{i,j} = \sum_i p(i|j) = 1$. Note that this property asserts that \mathbf{M}^{exp} is a proper POVM. The above equation is somewhat abstract since it provides a description of the noise on the measurement operators level. In order to find its operational interpretation, let us use the fact that since the noise affects only the measurement process, Eq. (7) is fulfilled for an arbitrary quantum state which is measured on the noisy device. For an arbitrary quantum state ρ , let us denote the ideal vector of probabilities that one would've obtained in the ideal device as $\mathbf{p}^{\text{ideal}}$, which should be understood as in Eq (5). Due to the linearity of Born's rule, from Eq. (7) it follows that the vector of probabilities \mathbf{p}_{exp} obtained in an experiment on a noisy device is given by

$$\mathbf{p}_{\text{exp}} = \Lambda \mathbf{p}_{\text{ideal}}. \quad (8)$$

This allows us to give an intuitive interpretation of classical noise. Namely, the application of such noise is equivalent to the classical post-processing of the statistics that one would have obtained in an ideal scenario [36]. Such interpretation has been used previously in the context of the implementation of general POVMs by projective measurements (see works [37, 38, 24]).

3 Scheme of mitigation of readout errors

In this part, we lay out the idea of our procedure to mitigate readout errors. We first state the assumptions under which the method works perfectly. Then, we formally formulate the method itself. Finally, we compute operational distances (see Eq. (6)) between ideal and noisy projective measurements for (i) single-qubit classical readout errors and (ii) uncorrelated classical errors affecting multi-qubit projective measurements. These numbers give us an indication of the magnitude errors that can be corrected by our error mitigation technique in realistic devices.

3.1 Assumptions

Our error-mitigation scheme relies on the following assumptions:

1. (**Infinite statistics**) We have access to the statistics given by Born's rule (Eq. (2)).
2. (**Classical noise**) The measurement noise occurring in the device has the form of a classical noise described in Section 2 (Eq. (8)).
3. (**Characterized detector**) We have at our disposal a perfect description of that noise.

The approximate validity of these conditions can be motivated by the following arguments. Assumption of infinite statistics may be fulfilled to a good extent simply by increasing the number of experiments one gathers statistics from. Note that violations of this assumption introduce statistical errors that can also be taken into account.

Moreover, in Section 6 (where we present results of QDT for IBM and Rigetti devices) we show that the classical noise model stated in the second condition turns out to be a dominant type of noise in the IBM quantum devices, which uses superconducting transmon qubits. This suggests that classical noise may be a good measurement noise model for devices relying on similar architectures.

Finally, the characterization of noise occurring in a detector may in practice be obtained via Quantum Detector Tomography (see Section 2). Such reconstruction is of good quality, provided one has access to high fidelity preparations of single-qubit quantum states. For example, the fidelities³ of single-qubit gates in IBM quantum devices are typically high (of the order of 99.6% - see Section 6).

We conclude by noting that the violations of the assumptions affect the applicability of our error-mitigation procedure. The effects of such errors are analyzed in Section 4.

3.2 Correction of the statistics

Let us now describe the error-mitigation procedure. In fact, all we need is to notice that since the matrix Λ representing the noise is invertible, we may simply left-multiply the Eq. (8) by its inverse Λ^{-1} , obtaining

$$\mathbf{p}_{\text{ideal}} = \Lambda^{-1} \mathbf{p}_{\text{exp}}. \quad (9)$$

What Eq. (9) tells us, is that provided the knowledge of Λ , one can simply left-multiply \mathbf{p}_{exp} by its inverse, in order to obtain statistics $\mathbf{p}_{\text{ideal}}$ that one would have obtained in the experiments performed on the ideal, non-noisy device. This result indeed appears very natural if one recalls the fact that such classical noise is equivalent to the classical post-processing of ideal statistics (see Section 2).

3.3 Correction of statistics based on Quantum Detector Tomography in (multi-)qubit devices

So far, we have not assumed any particular structure of the measurement $\mathbf{M}^{\text{ideal}}$. In this subsection (and in fact throughout the most part of the paper), we focus on projective measurements in the computational basis in single-qubit and multi-qubit systems. Specifically, we describe in detail how our mitigation method works in these systems. Moreover, we analytically quantify the magnitude of statistical errors that our scheme is capable of correcting under the assumption of independent classical errors affecting the measurements.

3.3.1 Single qubit

For the case of single-qubit projective measurement, without loss of generality we may write effects of a perfect detector POVM $\mathbf{M}^{\text{ideal}} \equiv \mathbf{P}$ in the computational

³State preparation errors may be obtained in the approximately readout-error-independent manner via randomized benchmarking (RB) experiments [39, 40]. Such experiments allow to calculate the average fidelities of quantum gates (that are used to prepare quantum states) in a way that is independent of measurement errors.

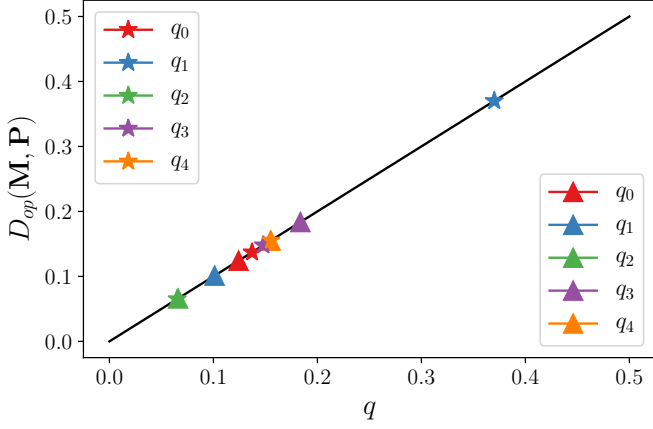


Figure 2: The operational distance $D_{op}(\mathbf{M}, \mathbf{P})$ between the ideal single-qubit projective measurement \mathbf{P} and the measurement with classical noise \mathbf{M} plotted against the probability q of erroneous detection. The figure presents the regime in which $q > p$ (see Eq. (11)), which is the typical situation in experiments. The stars (triangles) correspond to the experimental values of q obtained in the detector tomography experiments on IBM's (Rigetti's) device.

basis as

$$P_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, P_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \quad (10)$$

Let us assume that the measurement that is actually implemented is of the form $\mathbf{M}^{\text{exp}} \equiv \mathbf{M} = \Lambda \mathbf{P}$, for some invertible stochastic transformation Λ . This means that effects of \mathbf{M} can be written as

$$M_1 = \begin{bmatrix} 1-p & 0 \\ 0 & q \end{bmatrix}, M_2 = \begin{bmatrix} p & 0 \\ 0 & 1-q \end{bmatrix}, \quad (11)$$

where $p, q \in [0, 1]$ are probabilities of erroneous detection for outcomes 1 and 2, respectively. It is now straightforward to verify that

$$\Lambda = \begin{bmatrix} 1-p & q \\ p & 1-q \end{bmatrix}. \quad (12)$$

From Λ we can construct the **correction matrix**

$$\Lambda^{-1} = \frac{1}{1-p-q} \begin{bmatrix} 1-q & -q \\ -p & 1-p \end{bmatrix}, \quad (13)$$

which can be used in the future to correct any statistics obtained in one-qubit experiments performed on this device. The adjustment simply amounts to multiplying column vector of statistics from the left by Λ^{-1} . Finally, to illustrate how error probabilities p and q affect the distance of \mathbf{M} from the ideal detector \mathbf{P} , we compute the operational distance

$$D_{op}(\mathbf{M}, \mathbf{P}) = \max\{p, q\}. \quad (14)$$

The plot of this dependence is presented in Fig. 2.

3.3.2 Multiple qubits

Considerations from the previous subsection may be easily generalized for multiple-qubit systems. In general,

detector errors may exhibit correlations between qubits, therefore QDT should be performed via simultaneous measurements on multiple qubits. The problem of tomographic reconstruction is then exponential in the number of qubits K , since it requires preparation of exponentially many quantum states. However, QDT can be performed efficiently if the readout errors affecting the qubits are uncorrelated, i.e.,

$$\Lambda^{(K)} = \bigotimes_{i=1}^K \Lambda_i, \quad \Lambda_i = \begin{bmatrix} 1-p_i & q_i \\ p_i & 1-q_i \end{bmatrix}, \quad (15)$$

where Λ_i is the stochastic matrix describing the readout noise on the i th qubit. If that is the case, it suffices to perform multiple single-qubit QDTs, which makes the problem linear in the number of qubits. Interestingly, the operational distance between the noisy detector $\mathbf{M}^{(K)} = \Lambda^{(K)} \mathbf{P}^{(K)}$ and a multi-qubit projective measurement $\mathbf{P}^{(K)} = \bigotimes_{i=1}^K \mathbf{P}_i$ can be computed analytically (see Appendix A for a detailed derivation)

$$D_{op}(\mathbf{M}^{(K)}, \mathbf{P}^{(K)}) = 1 - \prod_{i=1}^K (1 - D_{op}(\Lambda_i \mathbf{P}_i, \mathbf{P}_i)), \quad (16)$$

where $D_{op}(\Lambda_i \mathbf{P}_i, \mathbf{P}_i) = \max\{p_i, q_i\}$ is the operational distance between ideal and noisy detector on the i th qubit. Importantly for small individual errors (i.e. for $D_{op}(\Lambda_i \mathbf{P}_i, \mathbf{P}_i) \ll 1$) implies that the total error $D_{op}(\mathbf{M}^{(K)}, \mathbf{P}^{(K)})$ is approximately additive

$$D_{op}(\mathbf{M}^{(K)}, \mathbf{P}^{(K)}) \approx \sum_{i=1}^K D_{op}(\Lambda_i \mathbf{P}_i, \mathbf{P}_i), \quad (17)$$

This motivates the usage of our error-mitigation procedure in this regime.

4 Error analysis

So far we have considered idealized scenarios in which noise affecting the measurements was perfectly classical and we could repeat experiments infinitely many times. Now we will provide an analysis of the possible deviations from the ideal model for the problem of reconstruction of measurement statistics. Specifically, we will describe three sources of errors:

1. Not entirely classical noise – it might happen that the noise is not of purely classical nature (as described in Sec. 2), but also have some non-classical unitary rotation part.
2. We have only access to a finite number of experiments, which introduces the statistical noise when we reconstruct probability distributions.
3. Finding the closest probability vector in the case when the correction yields non-physical probability distribution.

We note that we omit here errors resulting from the imperfect tomography of measurements, which we shall address in the future work⁴.

⁴To account for imperfect QDT, we would need a method for

4.1 Non-classical noise

Typically noise affecting quantum measurements cannot be described by Eq. (7). Let us consider the situation, in which the measurement that is actually being implemented is of the form

$$\mathbf{M}^{\text{exp}} = \Lambda \mathbf{M}^{\text{ideal}} + \Delta, \quad (18)$$

where we have decomposed a POVM into a part $\Lambda \mathbf{M}^{\text{ideal}}$ that represents ideal POVM affected by the classical noise (as in Eq. (18)) and Δ which represents every other (non-classical) errors in reconstructed POVM. Of course, having access to \mathbf{M}^{exp} (e.g., from QDT) and $\mathbf{M}^{\text{ideal}}$ (from the theoretical model), decomposition presented in Eq. (18) may be done in arbitrary way. However, in the case of projective d -outcome measurement in the computational basis, the following ansatz seems natural. Namely, we propose to consider a diagonal part of the POVM as containing information about Λ , while all off-diagonal terms should be regarded as non-classical part of the noise Δ (see also Remark 1).

Clearly, non-zero Δ affects our error-mitigation procedure. In the case of a noisy POVM of the form given in Eq. (18), it is impossible to reverse the effects of such noise solely by classical post-processing. Computing the vector of statistics for arbitrary quantum state ρ with the help of Eq. (18) yields (in analogy to Eq.(8))

$$\mathbf{p}_{\text{exp}} = \Lambda \mathbf{p}_{\text{ideal}} + \tilde{\Delta}, \quad (19)$$

where $\tilde{\Delta}$ denotes a generic disturbance of experimental statistics which arises due to the presence of non-classical part of the noise Δ . Its elements may be defined as $\tilde{\Delta}_i := \text{Tr}(\rho \Delta_i)$, where ρ is an *unknown* quantum state and Δ_i corresponds to the non-classical part of i th effect. Applying Λ^{-1} to Eq. (19) gives

$$\Lambda^{-1} \mathbf{p}_{\text{exp}} = \mathbf{p}_{\text{ideal}} + \Lambda^{-1} \tilde{\Delta}, \quad (20)$$

which clearly does not leave us with the ideal statistics, but consists also some additional term $\Lambda^{-1} \tilde{\Delta}$.

However, if non-classical part Δ is small compared to the term $\Lambda \mathbf{M}^{\text{ideal}}$, we propose to neglect the non-classical part Δ and perform our error-mitigation procedure *as if there were only classical noise*. Such action will clearly introduce some error into resulting estimated probability vectors. We quantitatively characterize the error introduced by neglecting non-classical part of the noise by finding an upper bound on $D_{TV}(\Lambda^{-1} \mathbf{p}_{\text{exp}}, \mathbf{p}_{\text{ideal}})$ in terms of the operational distance between the ideal measurement \mathbf{M}^{exp} and $\Lambda \mathbf{M}^{\text{ideal}}$. Using Eq.(19) we obtain

$$\begin{aligned} D_{TV}(\Lambda^{-1} \mathbf{p}_{\text{exp}}, \mathbf{p}_{\text{ideal}}) &= \frac{1}{2} \|\Lambda^{-1} \mathbf{p}_{\text{exp}} - \mathbf{p}_{\text{ideal}}\|_1 = \\ &= \frac{1}{2} \|\Lambda^{-1} \tilde{\Delta}\|_1 \leq \frac{1}{2} \|\Lambda^{-1}\|_{1 \rightarrow 1} \|\tilde{\Delta}\|_1, \end{aligned} \quad (21)$$

where in the inequality in the second line we have used the standard inequality $\|Ax\|_1 \leq \|A\|_{1 \rightarrow 1} \|x\|_1$ valid for

estimation of confidence intervals for the operational distance between POVMs, which we have not developed yet (some progress in this regards have been recently developed for quantum states [41])

all vectors in \mathbb{R}^n and linear transformations⁵ $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$. Performing the optimization over all quantum states (note that $\tilde{\Delta}$ implicitly depends on a quantum state ρ), and using Eq.(4) finally yields the bound

$$D_{TV}(\Lambda^{-1} \mathbf{p}_{\text{exp}}, \mathbf{p}_{\text{ideal}}) \leq \|\Lambda^{-1}\|_{1 \rightarrow 1} D_{op}(\mathbf{M}^{\text{exp}}, \Lambda \mathbf{M}^{\text{ideal}}) \quad (22)$$

We see that upper bound for the error that is introduced by our error-mitigation procedure, due to neglecting the non-classical part of the noise, can be expressed by $1 \rightarrow 1$ norm of Λ^{-1} (i.e. maximal l_1 norm of its columns) and by the operational distance between the actually implemented measurement \mathbf{M}^{exp} (given by Eq. (18)), and the part of POVM which is affected only by classical noise, i.e., $\Lambda \mathbf{M}^{\text{ideal}}$.

Remark 1. The above considerations were based on the decomposition of the imperfect measurements on the 'classical' ($\Lambda \mathbf{M}$) and the 'non-classical' (Δ) part. In general, such decomposition may be done in a somewhat arbitrary way. In principle, one could define an optimization problem, which should minimize, e.g., some suitably-chosen cost function depending on $\|\Lambda^{-1}\|_{1 \rightarrow 1}$ and $\|\Delta\|_1$ so aiming to minimize the upper bound on the error of the error-mitigation procedure.

4.2 Statistical errors

We will now study how finite-statistics errors affect the performance of our method. First of all, let us note that in experiments we do not have access to the probability vector \mathbf{p}_{exp} which we used previously, but only to its *estimator*. In what follows we will focus on the natural Maximum-Likelihood (ML) estimator for which estimates of individual probabilities p_i are given by relative frequencies n_i/N of events observed in the experiment repeated multiple times. Therefore, we shall now call the experimental statistics vector $\mathbf{p}_{\text{exp}}^{\text{est}}$, putting emphasis on the fact that it is an estimator.

The quality of the estimation procedure can be quantified by $\text{Pr}_{\text{err}}(\epsilon) := \text{Pr}(D_{TV}(\mathbf{p}_{\text{exp}}^{\text{est}}, \mathbf{p}_{\text{exp}}) \geq \epsilon)$, i.e., the probability that the TV distance between the estimated statistics vector $\mathbf{p}_{\text{exp}}^{\text{est}}$ and the true vector \mathbf{p}_{exp} is greater than some threshold ϵ . Hence, the value of $1 - \text{Pr}_{\text{err}}(\epsilon)$ may be interpreted as the acceptable confidence level for our estimation. Let us denote by N the number of samples (experimental runs) and by n the number of outcomes of the considered measurement (since we mainly focus on the standard projective measurements, n coincides with the dimension of the considered Hilbert space). In work [42] authors proved that in this scenario with probability at least $1 - \text{Pr}_{\text{err}}$

$$D_{TV}(\mathbf{p}_{\text{exp}}^{\text{est}}, \mathbf{p}_{\text{exp}}) \leq \sqrt{\frac{\log(2^n - 2) - \log(\text{Pr}_{\text{err}})}{2N}} =: \epsilon. \quad (23)$$

The above inequality gives us the upper bound for the TV-distance between estimated statistics (frequencies)

⁵The norm $\|A\|_{1 \rightarrow 1}$ denotes the operator norm of A understood as an operator between in \mathbb{R}^n equipped with l_1 norm. It can be easily compute as it equals to the maximum of l_1 norms of columns of A .

vector $\mathbf{p}_{\text{exp}}^{\text{est}}$ and actual probability vector \mathbf{p}_{exp} as a function of the number of experimental runs and the accepted error probability.

Let us finally stress that we are not interested in the statistical error itself as it is inherently present in any estimation scheme, whenever we have on our disposal only a finite number of samples. Instead, we want to understand how finite-sample statistics affect our procedure. Specifically, we want to bound the TV distance of our corrected estimated statistics $\Lambda^{-1}\mathbf{p}_{\text{exp}}^{\text{est}}$ to the ideal probability vector $\mathbf{p}_{\text{ideal}}$. This can be easily done in a manner analogous to the derivation leading to Eq. (22). The final bound is the following

$$\begin{aligned} D_{TV}(\Lambda^{-1}\mathbf{p}_{\text{exp}}^{\text{est}}, \mathbf{p}_{\text{ideal}}) &\leq \\ D_{TV}(\Lambda^{-1}\mathbf{p}_{\text{exp}}^{\text{est}}, \Lambda^{-1}\mathbf{p}_{\text{exp}}) + D_{TV}(\Lambda^{-1}\mathbf{p}_{\text{exp}}, \mathbf{p}_{\text{ideal}}) &\leq \\ \|\Lambda^{-1}\|_{1 \rightarrow 1} \epsilon + \|\Lambda^{-1}\|_{1 \rightarrow 1} D_{op}(\mathbf{M}^{\text{exp}}, \Lambda \mathbf{M}^{\text{ideal}}) &=: \delta, \end{aligned} \quad (24)$$

where the first inequality is a consequence of the triangle inequality, and the derivation of the second inequality is fully analogous to that of Eq. (22). In the upper bound above, the first term in the sum accounts for the statistical errors while the second term accounts for the non-classical part of the noise.

4.3 Non-physical probability vectors

Due to the coherent and statistical errors described above, it may happen that the corrected vector of the estimated statistics $\Lambda^{-1}\mathbf{p}_{\text{exp}}^{\text{est}}$ will not be a proper probability vector. Generally, the vector obtained in our procedure may contain negative elements that will, however, sum up to 1 (this is because the inverse of left-stochastic matrix still has columns of which elements sum up to 1). When this is the case, we propose to find the *corrected* vector $\mathbf{p}_{\text{exp}}^*$ which is the closest probability vector to $\Lambda^{-1}\mathbf{p}_{\text{exp}}^{\text{est}}$ in Euclidean norm⁶. In other words, if obtained vector $\Lambda^{-1}\mathbf{p}_{\text{exp}}^{\text{est}}$ is not a proper probability vector, we use the *corrected* vector $\mathbf{p}_{\text{exp}}^*$ that is the solution to the following problem

$$\mathbf{p}_{\text{exp}}^* = \underset{\forall_i p_i \geq 0, \sum_{i=1}^n p_i = 1}{\text{argmin}} \left(\|\Lambda^{-1}\mathbf{p}_{\text{exp}}^{\text{est}} - \mathbf{p}\|_2 \right), \quad (25)$$

where $\|\cdot\|_2$ denotes the Euclidean norm. The above problem can be easily solved by convex optimization solvers like *cvxopt* [43]. Furthermore, the error α introduced by this method may be quantified in terms of TV-distance between new corrected vector $\mathbf{p}_{\text{exp}}^*$ and non-physical $\Lambda^{-1}\mathbf{p}_{\text{exp}}^{\text{est}}$

$$\alpha := \frac{1}{2} \|\mathbf{p}_{\text{exp}}^* - \Lambda^{-1}\mathbf{p}_{\text{exp}}^{\text{est}}\|_1. \quad (26)$$

In order to account for the overall error of our procedure, we modify the Eq. (24) by putting $\mathbf{p}_{\text{exp}}^*$ in place of $\Lambda^{-1}\mathbf{p}_{\text{exp}}^{\text{est}}$. The usage of the triangle inequality gives

⁶We note that the choice of this norm in the error-mitigation procedure is somewhat arbitrary. We have chosen the Euclidean for sake of the mathematical convenience (the relevant optimization problem can be easily solved).

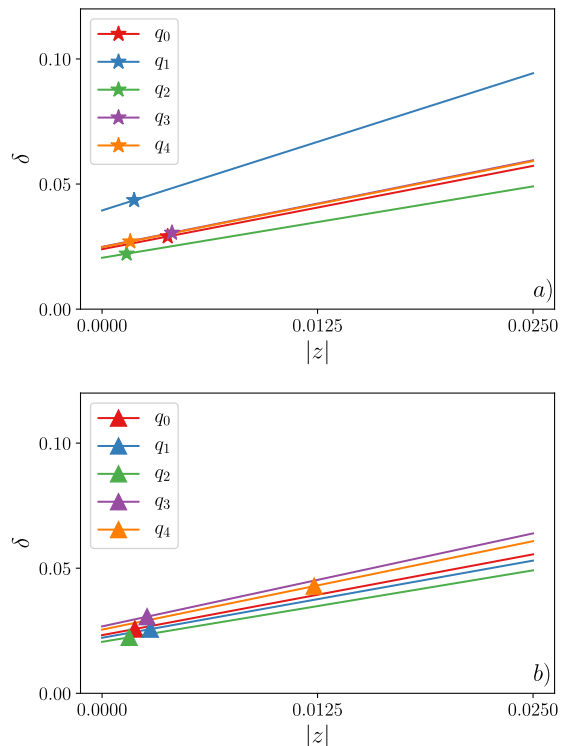


Figure 3: Dependence of the error upper bound δ on the magnitude of the coherent error $|z|$ for noisy single-qubit measurements. We depict the interval of relevant $|z|$ values obtained from the detector tomography performed on a) IBM's, b) Rigetti's devices. Stars and triangles correspond to the actual experimental values of $|z|$. Statistical error is fixed at value $\epsilon = 0.018$, which corresponds to setting $\text{Pr}_{\text{err}} = 0.01$ when $N = 8192$.

the following upper bound for the overall error in our procedure

$$D_{TV}(\mathbf{p}_{\text{exp}}^*, \mathbf{p}_{\text{ideal}}) \leq \delta + \alpha, \quad (27)$$

where δ and α are defined in Eq. (24) and Eq. (26), respectively. The quantity $\delta + \alpha$ can be considered as the overall error that our procedure yields for the problem of estimation of the probability vector $\mathbf{p}_{\text{ideal}}$. This upper bound on $D_{TV}(\mathbf{p}_{\text{exp}}^*, \mathbf{p}_{\text{ideal}})$ is a function of four variables: the number of experimental runs N used in the procedure of probability estimation, accepted error probability Pr_{err} , the reconstructed measurement \mathbf{M}^{exp} , and the vector of statistics obtained in a given experiment $\mathbf{p}_{\text{exp}}^{\text{est}}$.

4.4 When is the mitigation successful?

We would like to address now a crucial question. Namely, when can we consider our error-mitigation procedure to be successful? The answer to this can be given by analyzing the upper bound $\delta + \alpha$ on the error of estimation of probability vector that can be potentially obtained by the error-mitigation scheme (resulting from the presence of coherent errors and finite-statistics errors).

Imagine that we have on our disposal only the noisy quantum measurement \mathbf{M}^{exp} that produces N samples

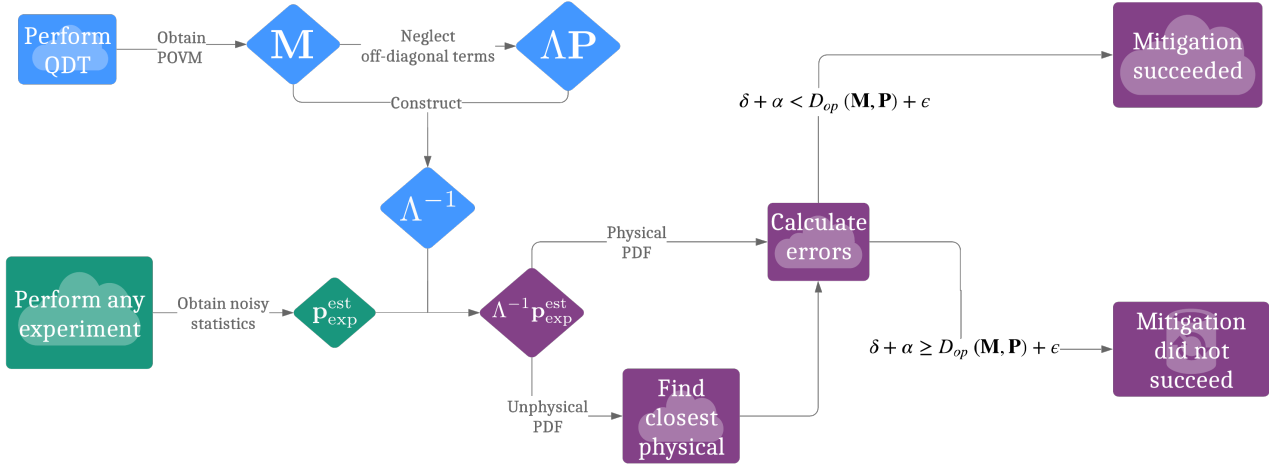


Figure 4: Pictorial representation of the mitigation scheme for the case of projective measurements. Blue boxes correspond to the pre-processing stage. In this stage, it is first necessary to perform the tomography of the noisy detector. This gives access to the classical description of the noisy POVM \mathbf{M}^{exp} . Neglecting the off-diagonal terms of its effects gives the classical part of the readout noise Λ , from which one obtains the correction matrix Λ^{-1} . Green boxes correspond to performing an arbitrary experiment on the noisy device. The error-mitigation scheme itself is presented by the purple boxes. Upon correcting experimental statistics $\mathbf{p}_{\text{exp}}^{\text{est}}$, one needs to determine if $\Lambda^{-1}\mathbf{p}_{\text{exp}}^{\text{est}}$ is a physical probability vector. If not, one performs the additional step of finding the physical probability vector closest to $\Lambda^{-1}\mathbf{p}_{\text{exp}}^{\text{est}}$. The final step is to calculate the upper bound on the error $D_{TV}(\mathbf{p}_{\text{exp}}^{\text{est}}, \mathbf{p}^{\text{ideal}})$ given by $\delta + \alpha$ (Eq. (27)). The decision whether such a mitigation is successful or not is based on the comparison of this bound with $D_{op}(\mathbf{P}, \mathbf{M}) + \epsilon$.

from the distribution \mathbf{p}_{exp} . Let $\mathbf{p}_{\text{exp}}^{\text{est}}$ denote the empirical estimator of this probability vector. By the virtue of considerations given in earlier subsections, the error resulting from imperfect measurement and the finite number of samples (for the assumed error probability Pr_{err}) can be upper bounded as follows

$$D_{TV}(\mathbf{p}_{\text{exp}}^{\text{est}}, \mathbf{p}^{\text{ideal}}) \leq D_{op}(\mathbf{M}^{\text{exp}}, \mathbf{M}^{\text{ideal}}) + \epsilon \quad (28)$$

We then propose to consider our mitigation successful if the upper bound in Eq.(27), on the error $D_{TV}(\mathbf{p}_{\text{exp}}^{\text{est}}, \mathbf{p}^{\text{ideal}})$ that can be introduced by the error-mitigation procedure, is smaller than the upper bound on $D_{TV}(\mathbf{p}_{\text{exp}}^{\text{est}}, \mathbf{p}^{\text{ideal}})$ discussed above (for the same number of samples N and error probability Pr_{err}). In other words we propose the following rule

$$\delta + \alpha < D_{op}(\mathbf{M}^{\text{exp}}, \mathbf{M}^{\text{ideal}}) + \epsilon \Rightarrow \text{mitigation successful} . \quad (29)$$

Remark 2. Of course, the actual distance from the perfect probability distribution after post-processing is highly dependent on the quantum state that is measured. Importantly, it might happen that though inequality in (29) holds, the particular quantum state will give statistics that after post-processing are *worse* than without it in terms of the distance from the perfect probability distribution. We analyze this problem in some detail in Appendix B where we provide more support of the heuristic procedure presented above.

Let us illustrate the above ideas on the simple example of single-qubit projective measurement. We assume that in detector tomography, noisy POVM describing our device is given by

$$M_1 = \begin{bmatrix} 1-p & z \\ \bar{z} & q \end{bmatrix}, \quad M_2 = \begin{bmatrix} p & -z \\ -\bar{z} & 1-q \end{bmatrix}, \quad (30)$$

Unlike the case of purely classical noise discussed in Section 3, there are also off-diagonal terms which represent non-classical part of the noise. If their magnitude is much smaller than that of the diagonal elements, we may approximate our POVM by $\Lambda\mathbf{P}$ (where the appropriate stochastic matrix is given in Eq. (12)) at the cost of introducing some additional errors. According to the analysis presented earlier, in order to bound this error we need to compute $\|\Lambda^{-1}\|_{1 \rightarrow 1}$ and $D_{op}(\mathbf{M}, \Lambda\mathbf{P})$. Straightforward computation give us the following expressions

$$\|\Lambda^{-1}\|_{1 \rightarrow 1} = \frac{1 + |p - q|}{|p + q - 1|}, \quad D_{op}(\mathbf{M}, \Lambda\mathbf{P}) = |z|. \quad (31)$$

Inserting these expressions into Eq. (24) we obtain the following upper bound on the error of estimation of the probability distribution with the usage of our error mitigation scheme

$$D_{TV}(\Lambda^{-1}\mathbf{p}_{\text{exp}}^{\text{est}}, \mathbf{p}^{\text{ideal}}) \leq \delta = \frac{1 + |p - q|}{|p + q - 1|} (|z| + \epsilon), \quad (32)$$

where statistical error ϵ is determined by Eq. (23) and depends on the assumed level of confidence and the number of samples N used in the estimation procedure. For example, if we set $N = 8192$, which is maximal number of experimental runs for a single quantum circuit in IBM Q devices, and $\text{Pr}_{\text{err}} = 0.01$, we obtain $\epsilon \approx 0.018$. Importantly in Eq. (32) we have assumed that the estimated probability vector $\Lambda^{-1}\mathbf{p}_{\text{exp}}^{\text{est}}$ is a probability distribution add hence we could set $\alpha = 0$ in Eq. (26). The visualization of overall error quantifier δ is shown in Fig. 3.

5 Summary of the method

In the following, we aim to correct the output statistics obtained from a noisy measurement. The error-

mitigation procedure applied to projective measurements is realized by performing the following sequence of practical steps, the graphical illustration of which is given in Fig 4.

1. Perform Quantum Detector Tomography (QDT) experiments. The estimation of the detector's POVM \mathbf{M} can be done using, e.g., an algorithm which converges to the Maximum Likelihood Estimation, like the one of [29].
2. Neglect all the off-diagonal terms of the elements of \mathbf{M} , obtaining $\Lambda\mathbf{P}$, where \mathbf{P} is the ideal projective measurement we want to implement.
3. Reverse the stochastic matrix Λ defined in Eq. (7) to obtain the "correction matrix" Λ^{-1} .
4. After performing arbitrary experiment on the device characterized in the above way, multiply the vector of obtained frequencies $\mathbf{p}_{\text{exp}}^{\text{est}}$ by Λ^{-1} .
5. Check if the $\Lambda^{-1}\mathbf{p}_{\text{exp}}^{\text{est}}$ is a proper probability vector, i.e., its elements are positive and sum up to 1.
 - If yes, set $\mathbf{p}_{\text{exp}}^* = \Lambda^{-1}\mathbf{p}_{\text{exp}}^{\text{est}}$ and proceed to the next step.
 - If no, solve the problem formulated in Eq (25), obtaining a vector of corrected statistics $\mathbf{p}_{\text{exp}}^*$.
6. Calculate $\delta + \alpha$, i.e., the upper bound on the total error magnitude $D_{TV}(\mathbf{p}_{\text{exp}}^*, \mathbf{p}_{\text{ideal}})$. Compare this with $D_{op}(\mathbf{P}, \mathbf{M}) + \epsilon$, i.e., with the upper bound on the error occurring without correction (Eq. 28).
 - If $\delta + \alpha \geq D_{op}(\mathbf{P}, \mathbf{M}) + \epsilon$, then the error-mitigation is considered not successful.
 - If $\delta + \alpha < D_{op}(\mathbf{P}, \mathbf{M}) + \epsilon$, then the error-mitigation is considered successful.

Remark 3. We note that if we wanted to mitigate the readout errors in the case of arbitrary ideal measurement $\mathbf{M}^{\text{ideal}}$, we would have to modify the second step of our method. Specifically, we would have to propose a decomposition of the noise on the classical ($\Lambda\mathbf{M}^{\text{ideal}}$) and non-classical (Δ) part (see discussion in Remark 1).

6 Device characterization

In this section, we present experimental results⁷ that confirm the (approximate) validity of the physical assumptions stated in Section 3 in the IBM's device. These assumptions have to be satisfied with our error-mitigation procedure to work. We first present the randomized benchmarking data for IBM's devices, which shows that single-qubit gates errors are small and make Quantum Detector Tomography feasible. Then the results of QDTs performed on IBM devices are provided. This shows that readout errors in these systems are indeed mostly of classical nature (see Section 2). We observe that the readout errors between physically connected pairs of qubits

⁷Dates of execution of all presented experiments are given in Appendix C.

are mostly uncorrelated (see Fig. 5 for visual presentation of the connectivity of *ibmqx4* device). Furthermore, throughout the section, we present analogous data for exemplary five qubits in Rigetti's 16-qubit device *Aspen-4-16Q-A*. However, in the case of this device, the assumption of perfect detector reconstruction is significantly violated due to the high infidelities of the single-qubit quantum gates.

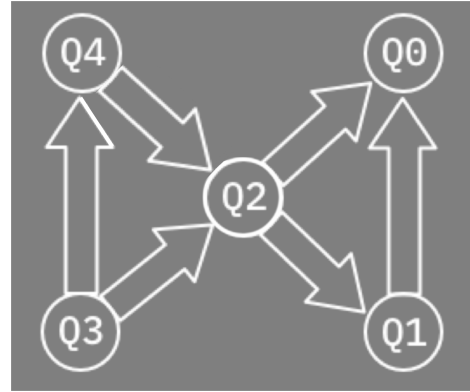


Figure 5: The scheme of connectivity for the qubits in the *ibmqx4* backend. The beginning (end) of the arrows denote the control (target) qubits of the directly accessible physical CNOT gates. Source: [44].

Remark 4. The experimental results presented in this work were obtained on publicly available devices. For this reason, it was not possible to perform all the experiments during a single calibration period. Consequently, the data presented throughout the paper may consist of values obtained in experiments performed on different days. Importantly, to correct statistics in all experiments performed in a given calibration period, we have used the data from QDT specific to that period. We note that results of QDT were varying across different calibration periods. However, those fluctuations do not change qualitative, nor quantitative conclusions derived throughout our work.

6.1 Perfect state preparation

In order to be able to perform reliable Quantum Detector Tomography, one needs to assure accurate state preparation. Since the tensor product of single-qubit quantum gates suffices to prepare quantum states spanning the whole operator space, it is enough to analyze only single-qubit gate errors. Randomized benchmarking experiments [39, 40] allow for inferring such errors in a manner approximately independent on state-preparation and measurement errors. In Table 1 we list average gate fidelities F_{RB} for *ibmqx4* device. The errors are relatively small (i.e., $1 - F_{RB}$ is of order 0.1%) which indicates that the assumption of perfect state preparation is reasonable. The average gate fidelities in Rigetti's device are, however, worse, which makes it difficult to precisely gauge the errors introduced by our error-mitigation procedure.

We would like to remark that in general the problem of rigorously separating measurement, gate, and state-

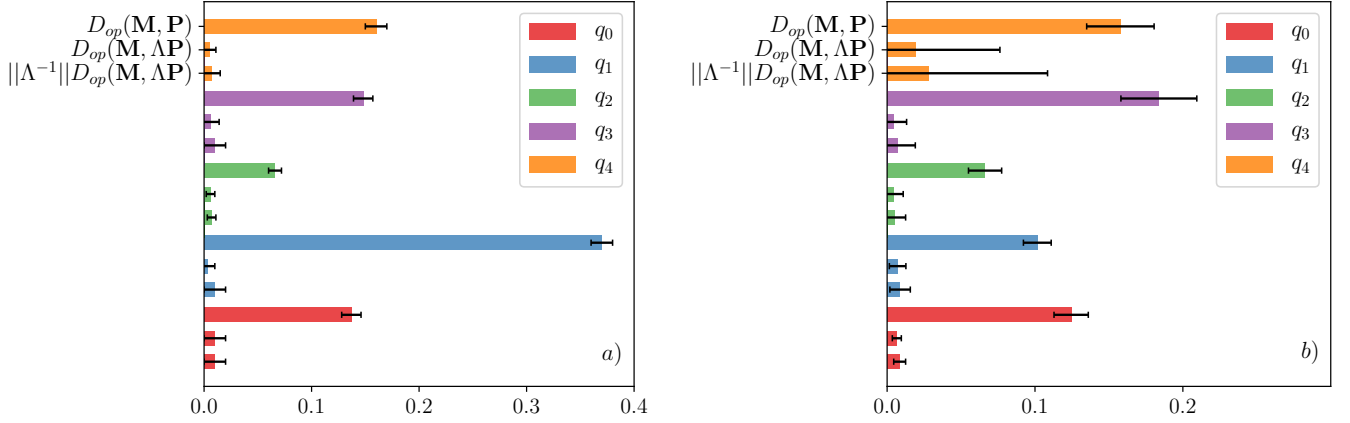


Figure 6: Characterization of single-qubit measurements on a) IBM's *ibmqx4* and b) Rigetti's *Aspen-4-16Q-A*. For each qubit, the first bar corresponds to the operational distance between the perfect projective measurement and the single-qubit POVM reconstructed via QDT, denoted by $D_{op}(\mathbf{M}, \mathbf{P})$. The second bar corresponds to the distance between the reconstructed POVM and the POVM obtained by neglecting its off-diagonal terms, denoted as $D_{op}(\mathbf{M}, \Delta\mathbf{P})$. The third bar corresponds to the upper bound for the overall error δ of our error-mitigation procedure, which does not include statistical errors. Each tomography experiment consisted of 6 quantum circuits: the preparation of all Pauli eigenstates. Each circuit was implemented 8192 times. Furthermore, each QDT experiment was repeated 4 times, in order to estimate the standard deviations, the corresponding 3σ bars are shown on the plot.

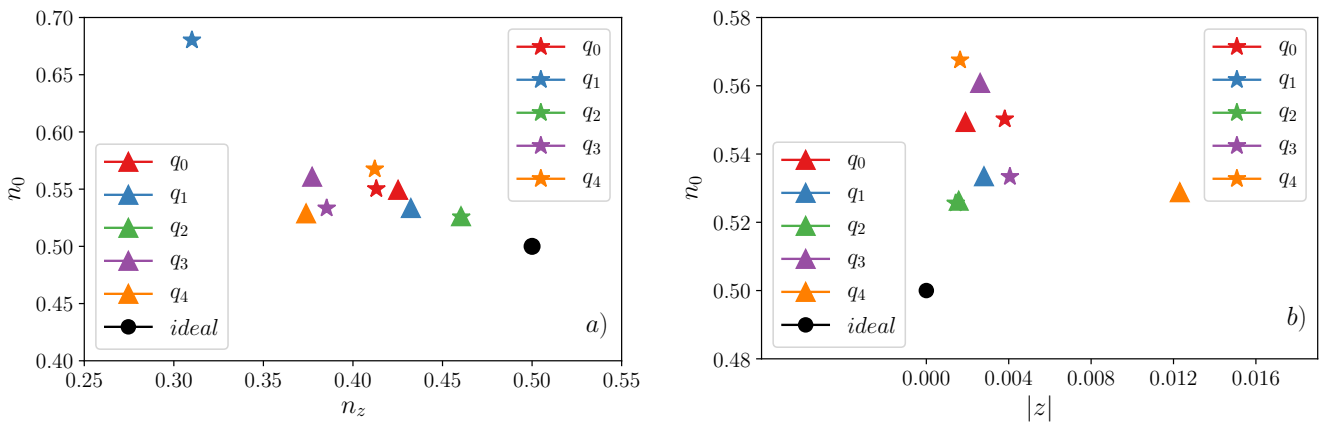


Figure 7: Visualization of single-qubit detectors for IBM's *ibmqx4* (stars) and Rigetti's *Aspen-4-16Q-A* (triangles). We parametrize the first effect of a measurement as $M_1 = \sum_{k \in \{0,x,y,z\}} n_k \sigma_k$, for $\sigma_0 := \mathbb{1}$. We set $|z| := \sqrt{n_x^2 + n_y^2}$ to be the magnitude of coherent errors. In the figures, we plot the two-dimensional projections of the four-dimensional parameter space of M_1 : part a) depicts the identity (n_0) and the σ_z (n_z) components, whereas part b) shows the identity (n_0) and the coherent ($|z|$) components.

Qubit	F_{RB}	Qubit	F_{RB}
q_0	0.998	q_0	0.993
q_1	0.997	q_1	0.965
q_2	0.996	q_2	0.914
q_3	0.996	q_3	0.924
q_4	0.996	q_4	0.937

a) b)

Table 1: Average single-qubit gate fidelities for a) *ibmqx4*, and b) *Aspen-4-16Q-A*. Data obtained via a) qiskit [2] and b) Forest [4].

preparation errors is a difficult task (for recent progress in this direction, see [45, 46, 47]). Without performing this division, we cannot precisely estimate the accuracy of our error-mitigation procedure. Let us emphasize, however, that for many practical applications (like variational quantum algorithms for example [48]) one can simply implement our scheme and test if it works in practice.

6.2 Quantum Detector Tomography

Here we present results of detector tomographies performed on all 5 qubits of IBM’s five-qubit device *ibmqx4* and on the first 5 qubits of Rigetti’s 16-qubit device *Aspen-4-16Q-A*.

6.2.1 Single qubit

In Fig. 6 we present the results of QDT performed on individual qubits in IBM’s and Rigetti’s devices. Three observations can be made. First, readout noise is significant for both platforms. We note that although it cannot be directly compared to the single-qubit gate errors from RB experiments, in the case of IBM *ibmqx4* average gate fidelities are so large that the readout-noise can be considered as the predominant type of noise on the level of individual qubits (at least for short quantum circuits). Second, the distance between actual POVM and its diagonal form is in all cases small, which confirms the assertion that classical noise is the dominant type of noise in both devices. Lastly, we observe that the upper bound for our correction error $\|\Lambda^{-1}\|_{1 \rightarrow 1} D_{op}(\mathbf{M}^{\text{exp}}, \Lambda \mathbf{M}^{\text{ideal}})$ (see Eq. (22)) is always significantly smaller than the distance of the noisy POVM from the ideal detector, which suggests that the mitigation scheme should be beneficial⁸.

We also visualize the results of QDT of single-qubit projective measurements on Fig 7. To this end, we parametrize the first effect of a measurement as $M_1 = \sum_{k \in \{0,x,y,z\}} n_k \sigma_k$, and $\sigma_0 := \mathbb{1}$. Then, we set $|z| := \sqrt{n_x^2 + n_y^2}$ to be the magnitude of coherent errors. In order for M_1 to remain element of POVM, its coefficients must satisfy $n_0^2 - 1 \leq |z|^2 + n_z^2 \leq n_0^2$.

⁸We note that this upper bound can be used under the assumption of infinite statistics. We would like to remark that similar conclusions can be also made with regards to more conservative indicators of errors that take into account a limited number of performed experiments.

Qubits	$D_{op}(\mathbf{M}^{2q}, \mathbf{M}^{1q \otimes 1q})$
$q_1 q_0$	0.03 ± 0.01
$q_2 q_0$	0.02 ± 0.01
$q_2 q_1$	0.19 ± 0.04
$q_3 q_2$	0.439 ± 0.04
$q_3 q_4$	0.038 ± 0.007
$q_4 q_2$	0.05 ± 0.01

Qubits	$D_{op}(\mathbf{M}^{2q}, \mathbf{M}^{1q \otimes 1q})$
$q_0 q_1$	0.02 ± 0.01
$q_1 q_2$	0.02 ± 0.01
$q_2 q_3$	0.04 ± 0.02

Table 2: Operational distances between the two-qubit POVMs that are obtained via the two-qubit QDT and the ones that are obtained via the tensor product of single-qubit POVMs. These values are provided for all physically connected qubit pairs in IBM’s *ibmqx4* (upper table) and for three exemplary physically connected pairs of Rigetti’s *Aspen-4-16Q-A* (lower table). Each tomography was repeated 4 times in order to estimate standard deviations, the corresponding 3σ intervals are also presented in table.

6.2.2 Two qubits

We have also performed QDT for two-qubit projective measurements on Rigetti’s and IBM’s devices. In Figs. 8a, c we present results of detector characterization performed on the basis of joint two-qubit measurement tomography. We also give results of the alternative approach in which two qubit POVMs are reconstructed from tensor products of single-qubit POVMs (see Figs. 8b, d). In both cases, the data resembles single-qubit scenario - coherent errors are relatively small and correction error $\|\Lambda^{-1}\|_{1 \rightarrow 1} D_{op}(\mathbf{M}^{\text{exp}}, \Lambda \mathbf{M}^{\text{ideal}})$ (relevant for the infinite statistics scenario) is always much smaller than operational distance of the POVM from ideal detector.

We have also investigated the correlations between measurement errors. Tab. 2 shows operational distances between POVMs reconstructed via two-qubit tomography and these obtained via tensor product of single-qubit measurements. We observe that for most pairs in IBM’s device, the correlations between readout errors are small. However, with two exceptions of the pairs $q_3 q_1$ and $q_2 q_1$. This suggests that mitigation for those pairs should be performed based on joint two-qubit QDT. In the case of Rigetti’s device, examined pairs do not show significant level of correlations.

7 Applications on IBM devices

We have performed numerous experiments on IBM *ibmqx4* quantum device to demonstrate the usefulness of our error mitigation method for a number of quantum information tasks. In what follows we first describe briefly the theoretical basis of each of the tasks performed. Then, we proceed to the detailed description of the experimental results.

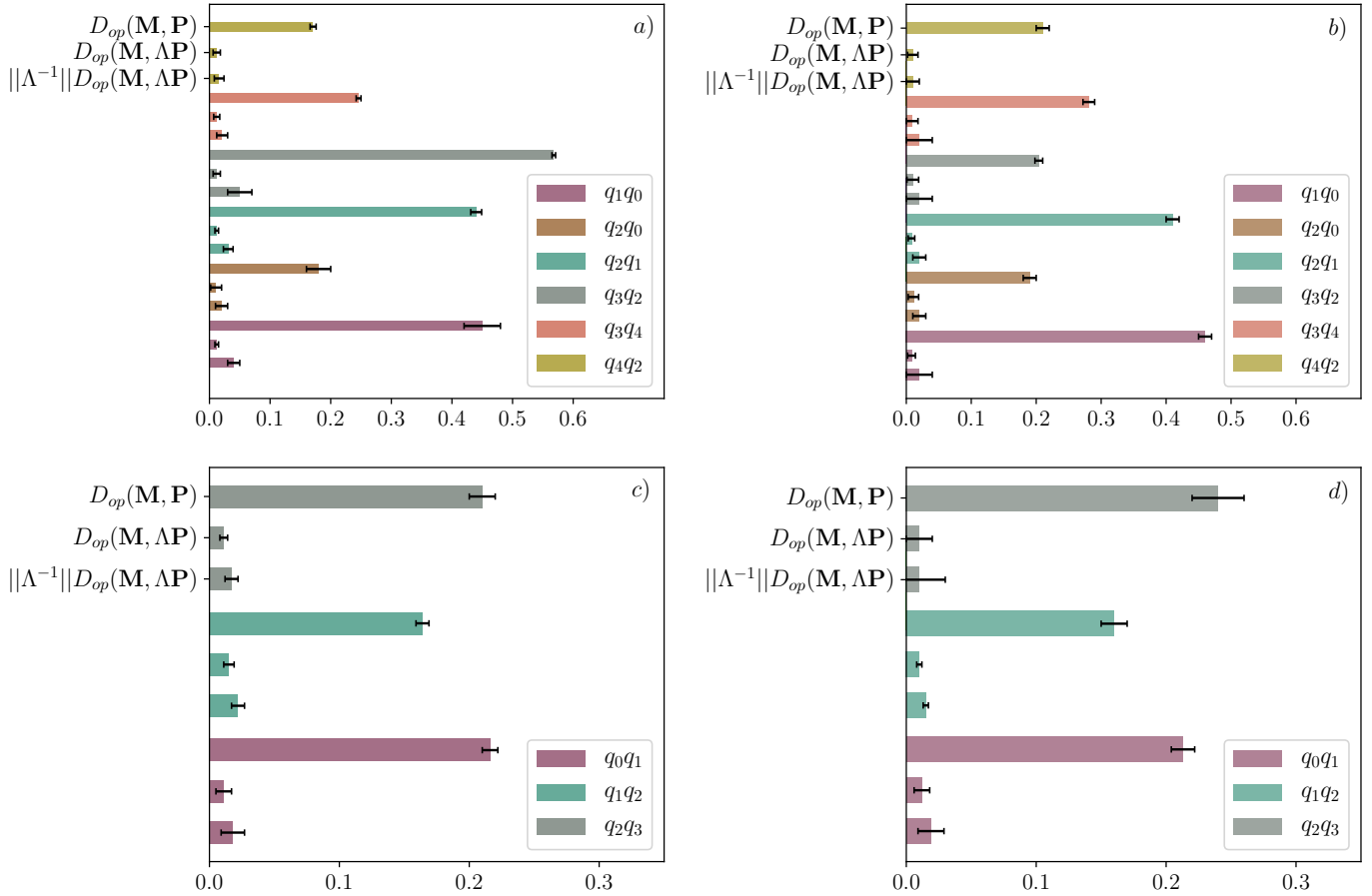


Figure 8: Barplot of the characterization of two-qubit measurements on a,b) all physically connected pairs of IBM's *ibmqx4* and c,d) three exemplary physically connected pairs of Rigetti's *Aspen-4-16Q-A*. The presented data is analogous to Fig. 6. On plots a,c), the results were obtained via a joint two-qubit detector tomography, while on plots b,d), the results were obtained via two single-qubit QDTs and the two-qubit POVMs created via tensor products. Each tomography experiment consisted of 36 circuits: all tensor products of Pauli eigenstates. Each circuit was implemented 8192. Furthermore, each QDT experiment was repeated 4 times, in order to estimate the standard deviations, the corresponding 3σ bars are shown on the plot.

7.1 Theoretical description

7.1.1 Quantum State Tomography and Quantum Process Tomography

Both Quantum State Tomography (QST) and Quantum Process Tomography (QPT) are based on the same idea as Quantum Detector Tomography, with the clear difference that in the case of QST a quantum state is to be reconstructed, while a QPT provides the characterization of a quantum process, i.e., channel. In our experiments, the tomographic reconstruction was done using the algorithm from [49], which is available in qiskit.

To compare the target pure state $|\psi\rangle\langle\psi|$ with the tomographically reconstructed ρ we used state fidelity [6],

$$F^{state}(|\psi\rangle\langle\psi|, \rho) = \text{Tr}(|\psi\rangle\langle\psi|\rho). \quad (33)$$

Likewise, in compare the target unitary channel U to-be-ideally-implemented with the actual quantum channel Λ obtained in the process tomography, we used the entanglement fidelity [50, 51], which may be calculated as

$$F^{ent} = \text{Tr}(\Phi_U \rho_\Lambda), \quad (34)$$

where Φ_U is the Choi-matrix representation of the unitary channel and ρ_Λ is the Choi-matrix representation of Λ .

7.1.2 Implementation of non-projective measurements

In our work, we have used generalized quantum measurements as models for noisy projective measurements. However, it is certainly not the only application of POVMs in quantum information. Indeed, POVMs can outperform projective measurements in some tasks, such as unambiguous state discrimination [52, 53], minimal error state discrimination [52, 53], quantum tomography [54], port-based teleportation of quantum states [55] or quantum computing [56]. Therefore, it may happen that one actually wants to *implement* a certain generalized measurement.

The standard method to do so is via the Naimark's extension [27, 57] and requires adding an ancilla system and performing a projective measurement on a whole system. We have implemented in this way three different single-qubit POVMs. We have assessed the quality of implementation by performing quantum measurement tomography and computing the operational distance, see Eq. (6).

7.1.3 Quantum algorithms – Grover's search and the Bernstein-Vazirani algorithm

Grover's [25] and Bernstein-Vazirani [26] algorithms are the canonical examples of quantum algorithms operating withing an oracular (black box) model of computation. Grover's search aims to find unknown element y encoded in the application of the unitary gate U_y defined via $U_y|x\rangle = (1 - 2\delta_{x,y})|x\rangle$, where $\delta_{x,y}$ is the Kronecker's delta. BV algorithm uses Hadamard transform to find (in a single query) a hidden string $s \in \mathbb{Z}_2^n$ encoded in a n -qubit unitary transformation V_s defined via $V_s|x\rangle = (-1)^{s \cdot x}|x\rangle$, where $s \cdot x = \sum_{i=1}^n s_i x_i \pmod{2}$.

We have based our implementation of these routines on the expository work [58]. Both algorithms were implemented on three qubits, one of which was an ancilla. In that case, Grover's algorithm required only a single query to the oracle, hence it could have been realized in a single quantum circuit. The figure of merit for the quality of implementation in the case of both algorithms is a single number – the probability of obtaining a particular outcome. We have used this number to benchmark our error mitigation scheme.

7.1.4 Probability distributions

We have implemented various five qubit circuits in the IBM device aiming to generate specific probability distributions upon measuring all the qubits. As a figure of merit we have used Total Variation distance (see Eq. (3)) between target probability distribution, and the one estimated from relative frequencies. In order to test our error mitigation method, we have implemented these probability distributions with and without the error-mitigation procedure.

7.2 Experimental results

Now we present experimental results that demonstrate the practical effectiveness of our error-mitigation procedure. For each presented experiment, QDT from which we inferred correction matrix Λ^{-1} , has been performed in the same calibration period as the corrected experiments. Moreover, for tasks involving measuring multiple-qubits, we compare mitigation which assumes a lack of correlation between qubit readout errors, with the one which accounts for such correlations.

7.2.1 Quantum State Tomography and Quantum Process Tomography

In what follows we present results of the quantum state and quantum process tomographies performed on single-qubit systems and quantum state tomographies performed on two-qubit systems. The procedure of tomography involves performing multiple experiments that are then used to reconstruct the objects in question. We use an error-mitigation procedure to correct the statistics (probability vectors) in these experiments in order to enhance the quality of the tomographic reconstruction.

In Fig. 9 we present results for different quantum state and process tomographies on individual qubits in IBM's *ibmqx4* device. It is worth noting that although results highly depend on the input state, our correction reduces infidelities of reconstructed states in every tested case.

We observe the improvement also for two-qubit state tomographies, as shown in Fig. 10. We compare the performance of the error-mitigation procedures based on the measurement reconstruction obtained from two-qubit QDTs and the one obtained from the tensor product of single-qubit QDTs. Interestingly, although one would expect that for highly correlated readout errors (in particular the pair q_2q_1), accounting for those correlations in readout errors should provide an improvement over the

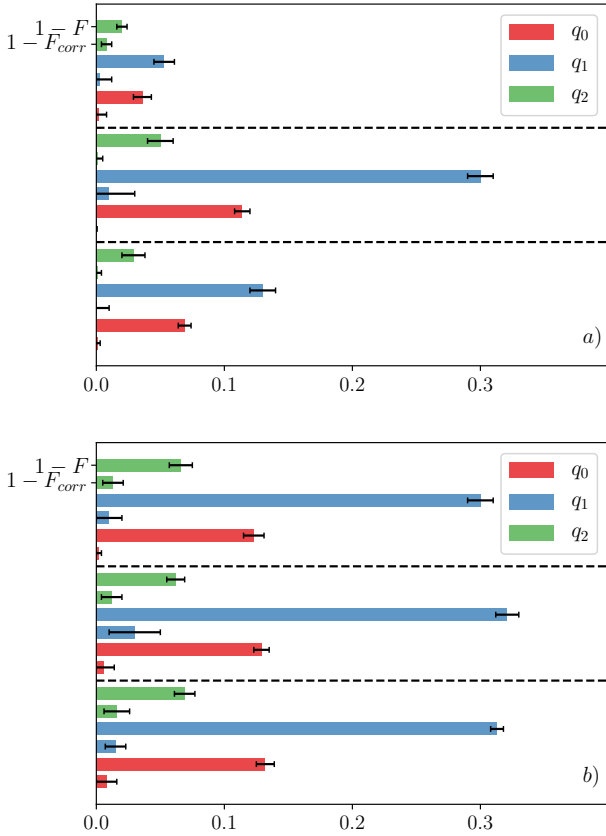


Figure 9: Infidelities obtained from the reconstruction of a) single-qubit states and b) single-qubit channels. Different colors refer to different qubits. The first bar in each pair corresponds to uncorrected statistics, while the second bar corresponds to the corrected one. The three regions separated by dashed lines refer to different Haar-random [59] quantum states (processes) to be reconstructed. Each tomography experiment consisted of several quantum circuits (3 for QST, 12 for QPT), and each quantum circuit was implemented 8192 times. Furthermore, each tomography experiment was repeated 5 times in order to estimate statistical errors, the corresponding 3σ intervals are shown on the barplot as black lines. The correction was based on a QDT which used an over-complete set of states (all Pauli eigenstates), with each circuit implemented 32768 times.

mitigation based on not-correlated tomography, we observed such improvement in the case of only one out of three tested quantum states.

We remark on the possible causes of the result reported above. First, in the course of correction of statistics, we sometimes obtain nonphysical probability vectors. As already pointed out in Section 4.3 dealing with such situations potentially introduce additional errors. These errors can be different for mitigation schemes based on different tomographic reconstructions. The second probable cause is the insufficient number of experiments necessary to perform a reliable QDT of a two-qubit measurement (the sample complexity of this problem is definitely higher as there are much more parameters needed to describe the two-qubit measurement compared to two single-qubit measurements).

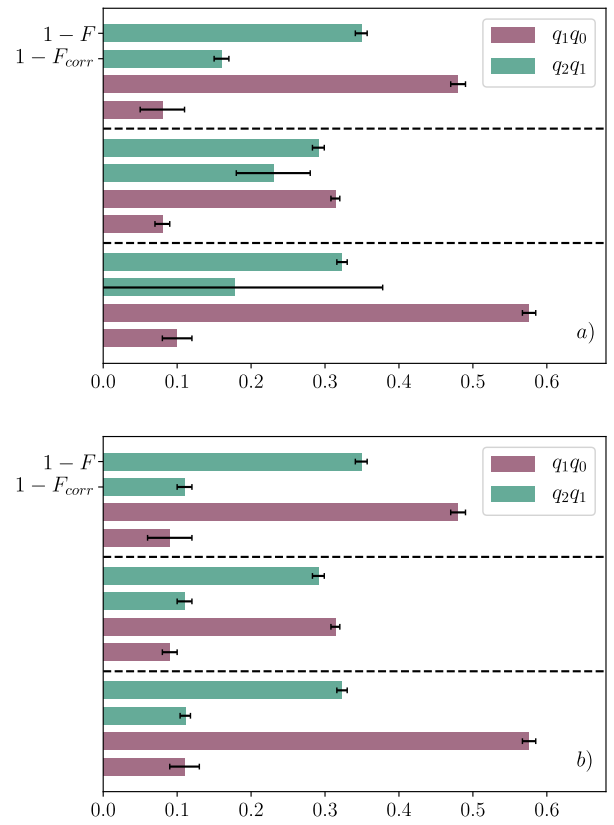


Figure 10: Infidelities obtained from the tomographic reconstruction of Haar-random [59] two-qubit quantum states using error-mitigation procedures based on a) uncorrelated single-qubit QDTs and b) correlated two-qubit QDT. The convention for presenting data is analogous to that of Fig. 9. Each tomography experiment consisted of 9 quantum circuits, corresponding to local measurements performed in different Pauli bases. Each quantum circuit was implemented 8192 times. Furthermore, each tomography experiment was repeated 5 times in order to estimate statistical errors, the corresponding 3σ intervals are shown on the barplot as black lines. The QDT used in the error-mitigation procedure was implemented using an over-complete set of states (all Pauli eigenstates), with each circuit implemented 32768 times.

7.2.2 Implementation of non-projective measurements

For the implementation of non-projective measurements via Naimark's extension, we observe the general improvement of the quality of reconstructed POVMs when the error-mitigation scheme is used (see Fig. 11). Moreover, the difference between error-mitigation procedure based on the non-correlated and on correlated detector tomography is negligible this time. However, we can offer a simpler explanation than in the case of two-qubit QSTs described above. Namely, at the time when those particular experiments were performed, correlations between errors for pair q_2q_1 were relatively small (compared to data regarding quantum process tomography and quantum detector tomography). This points to the importance of performing systematic device characterization and calibration.

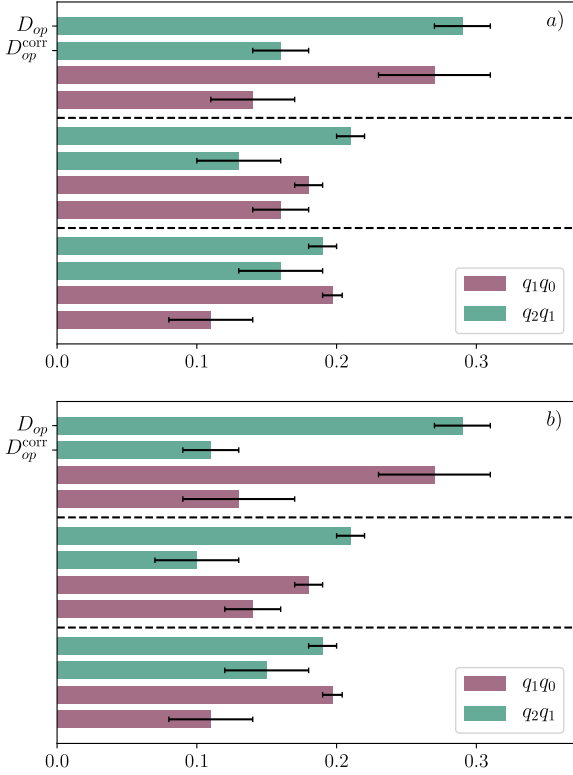


Figure 11: Operational distances for POVMs implemented in *ibmqx4* via Naimark's extension. Different colors correspond to different pairs of qubits. The first bar in each pair depicts the operational distance between the to-be-implemented POVM and the one obtained in the measurement tomography based on uncorrected statistics (D_{op}); while the second bar shows the distance between the ideal POVM and the one reconstructed from a tomography based on statistics corrected by our scheme (D_{op}^{corr}). The three regions separated by dashed lines correspond to different to-be-implemented POVMs: the Haar-random 4-outcome POVM, the tetrahedral measurement, and the trine measurement [59]. Each POVM reconstruction required 4 quantum circuits (Pauli 'x+', 'y+', 'z+' and 'z-' states) and each quantum circuit was implemented 8192 times. Reconstructions were repeated 5 times in order to estimate standard deviations, the corresponding 3σ intervals are shown here as black bars. The mitigation or measurement errors by our method was done based on a) separate single-qubit QDTs, and b) joint two-qubit QDT. The QDTs were based on the preparation of Pauli 'x+', 'y+', 'z+' and 'z-' states for single qubits, and all of their tensor-product combinations for pairs of qubits. Each circuit for the QDT was implemented 32768 times.

7.2.3 Quantum algorithms – Grover's search and the Bernstein-Vazirani algorithm

From the data presented in Table 3 we observe that our error mitigation method is favorable for the implementation of both Grover and BV algorithms. In the case of correction performed on the highly correlated pair q_2q_1 (Grover's search), the mitigation scheme accounting for the correlations of readout errors provided a significant advantage over the scheme using non-correlated tomography.

We remark that that relative improvement is much higher in the case of Grover's algorithm than in BV's.

Algorithm	Standard	Corr ($1q \otimes 1q$)	Corr ($2q$)
Grover's	0.58 ± 0.01	0.70 ± 0.02	0.79 ± 0.02
BV	0.55 ± 0.02	0.63 ± 0.02	0.61 ± 0.02

Table 3: The guess probabilities for the three-qubit Grover's search and BV algorithms implemented on *ibmqx4*. The third column shows the mitigation based on uncorrelated single-qubit QDTs, and the fourth column shows mitigation where the correlations between readout errors were accounted for. For both algorithms, we used the qubits q_0 , q_1 , and q_2 . The measurement was done on the q_2q_1 pair for Grover's search and on the pair q_1q_0 for the BV algorithm. For Grover's search, the hidden element was chosen to be '11', while for BV it was '01'. Each experiment was repeated 5 times in order to estimate standard deviations, the corresponding 3σ intervals are given in the table. The mitigation was based on a minimal QDT (Pauli 'x+', 'y+', 'z+' and 'z-' states), with each circuit implemented 32768 times.

	Uncorrelated	Correlated
lower bound for $D_{op}(\mathbf{M}, \mathbf{P}) + \epsilon$	0.748	0.745
upper bound for δ	0.291	0.256

Table 4: Errors for five-qubit measurements in *ibmqx4*. The first row presents the *lower bound*^a for the operational distances between noisy and ideal measurements *including* statistical errors ϵ with the number of repetitions $N = 72 \times 8192 = 589824$ and accepted error probability $\text{Pr}_{err} = 0.01$ (see Eq. (23)). The second row gives the *upper bound* on δ defined in Eq. (24), which bounds the errors in the observed statistics. The second column corresponds to a tomographic reconstruction based on the tensor product of five single-qubit QDTs, while in the third column the joint QDT for a pair q_2q_1 was included. The presented data is for POVMs which were used for the correction procedure, hence we do not provide standard deviations.

^aCalculating operational distance (Eq. (6)) via search over all possible combinations of indices quickly becomes computationally infeasible. However, a lower bound can be found by exhaustive search over a sufficiently large subset of all combinations. On the other hand, assuming a tensor structure of POVMs in consideration, an upper bound can be easily obtained using the subadditivity of the operational distance (see Appendix A for details). Finally, note that to estimate whether the inequality in Eq. (29) is satisfied, indeed the upper bound for LHS and the lower bound for RHS is needed.

This may result from two factors. First, the circuit for Grover's algorithm has lower depth than the circuit for the BV algorithm, therefore, in this case, gate-errors might have a smaller impact on the final result. Second, for the implemented instance of Grover's algorithm, the expected result was '11' (which is the state most prone to the noise observed in IBM's device), while for the BV algorithm the theoretically correct result was '01'.

Remark 5. We note that in Ref. [58], authors obtained values ≈ 0.65 and ≈ 0.386 for the correct guessing probabilities in the same instances of Grover's and BV algorithms. Both of these values differ significantly from the ones observed by us without implementing the correction procedure. Such discrepancy is another confirmation that indeed device's performance varies over long periods of time.

Name	Standard	Corrected	α
Uniform	0.110 ± 0.006	0.100 ± 0.007	0
NOT	0.66 ± 0.02	0 ± 0	0.36 ± 0.09
Mixed	0.196 ± 0.006	0.031 ± 0.008	0.019 ± 0.005

Name	Corrected	α
Uniform	0.03 ± 0.02	0
NOT	0.004 ± 0.023	0.04 ± 0.04
Mixed	0.022 ± 0.007	0.023 ± 0.007

Table 5: Results of experiments for implementing specific probability distributions on the *ibmqx4* processor. We use the total variation distance between the target and the reconstructed probability distribution in order to assess the quality of the implementation. The upper table presents results of for the uncorrected case and the case when applying a correction procedure that does not take into account the correlations in the readout errors, while the lower table provides the results of a correction procedure where the correlations between a single qubit pair (q_2q_1) measurements were taken into account. Each experiment consisted of a single quantum circuit, which was run $72 * 8192 = 589824$ times. This high number of repetitions was necessary in order to minimize the statistical errors. Furthermore, each experiment was repeated four times in order to estimate the statistical errors. The errors given in the tables have magnitude 3σ , where σ is the estimated standard deviation. The correction procedure used in the experiments was based on over-complete (all Pauli eigenstates) single-qubit QDTs with each circuit implemented 32768 times. The two-qubit QDT for pair q_2q_1 was done using all tensor-product combinations of Pauli eigenstates on both qubits, with each circuit implemented the same number of times.

7.2.4 Probability distributions

Finally, we present results on the implementation of probability distributions on five qubits. We start by giving the operational distances and bounds on the correction errors for the five-qubit detector in Table 4. From the data presented there, it is clear that even for 5 qubits we are still in the regime in which correction is beneficial (see the criterion given in Eq. (29)).

The probability distributions tested here were generated by performing a simple circuit (acting on the standard input state $|0\rangle^{\otimes 5}$) followed by a measurement in the computational basis. We have chosen to implement the following three probability distributions: The first distribution was the uniform one and was implemented by the tensor product of five Hadamard gates. The second probability distribution, named 'NOT', was implemented by the product of five X gates. For the last probability distribution, named by us 'Mixed', the circuit consisted of two X gates on q_0 and q_2 , two Hadamards on q_3 and q_4 , and the identity gate on q_1 .

Results of the correction of the probability distributions are presented in Tab. 5. There are a few interesting features to be observed. First, let us note that in the case of the uniform probability distribution, our correction procedure is ineffective if we do not account for correlations between qubit readout errors. However, if we do account for such correlations in only a *single* pair of qubits (the highly correlated pair q_2q_1), it is clear

that correction procedure provides a high improvement⁹. Second, in the case of the 'NOT' probability distribution, the correlation-ignorant correction has attained the perfect estimation – yet with remarkably high distance α between first-guess non-physical probability vector and its closest physical neighbor. This moves us very close to the regime in which the correction is unreliable according to the criterion given in Eq. (29). Accounting for correlations this time proves to improve results also remarkably highly, this time with only slightly distant first-guess non-physical distribution. Finally, the method provides an advantage for the 'Mixed' distribution in both cases. However, interestingly, this time the difference of improvement after accounting for correlations is negligible.

8 Conclusions and further research directions

8.1 Conclusions

In this work, we have presented a scheme for the mitigation of readout errors which is suitable for noisy and imperfect quantum devices. Our method uses solely classical post-processing that depends on the results obtained in the course of Quantum Detector Tomography. In principle, it can be applied on any quantum hardware provided the readout errors are classical and one has access to approximately perfect detector tomography. We have analyzed how our error-mitigation procedure is affected by non-classical noise and finite statistics (at the stage of estimation of probability distributions). We have also presented a comprehensive study confirming the (approximate) validity of the classicality of errors in the publicly available prototypes of quantum chips based on architectures with superconducting transmon qubits.

Our method was tested on a variety of quantum tasks and algorithms implemented on IBM's five-qubit quantum processor. We observed that our scheme, despite relying on purely classical processing of the obtained results, yields substantial improvements in the performance for a number of in a single-, two- and five-qubit experiments on the *ibmqx4* device. We have compared the error mitigation scheme which accounts for correlations in readout errors with the one that does not account for them. Taking into account correlations proved to be crucial in the case of experiments with single probability vectors (Grover's algorithm for two qubits and the probability distributions in the five-qubit case). We point at the study of readout error correlations and accounting for them as a future research direction.

We are convinced that our method and its future extensions will find applications to mitigate readout errors in realistic near-term quantum devices that, despite being inherently noisy and imperfect, are expected to be

⁹ It is important to note, that although from Table 2 one can see that there is also another highly correlated pair, q_3q_2 , our experiments for the five-qubit case were performed at the time in which only the pair q_2q_1 has shown a significant level of correlations. See Appendix C for the dates of execution of all experiments.

capable of effectively solving problems of practical relevance. One can argue that performing quantum detector tomography and implementing our error-mitigation procedure will not be feasible for systems involving many qubits (due to the exponential growth of the size of this problem). However, many quantum algorithms tailored to near-term applications (such as quantum approximate optimization algorithm [48] or variational quantum eigensolvers [60]) can be performed by implementing just a polynomial number of few-qubit quantum measurements. We expect that our scheme can be particularly useful in such scenarios.

8.2 Further research

Finally, we state a number of possible future research directions.

The first question concerns the possibility of *physical* correction of coherent readout errors. For the case of a single-qubit measurement there exists a natural possibility of correction of the coherent measurement errors. Namely, if the readout errors are uncorrelated one could obtain single-qubit POVMs with diagonal effects by implementing suitably chosen unitaries at the end of every quantum circuit. Moreover, in the case of uncorrelated errors, such physical correction is in principle possible for an arbitrary number of qubits (provided that readout errors dominate over gate errors). Determining whether this procedure will work for realistic near-term devices is an important research question.

Another important problem is to give confidence intervals (with respect to the operational distance defined in Eq. (4)) for the problem of quantum detector tomography. This kind of results is required to give estimates for the sample complexity of QDT and hence are crucial in assessing the feasibility of QDT on realistic near-term quantum devices. So far, such results have been obtained only for the case of quantum state tomography and the trace distance as the figure of merit [41, 61].

Last but not least, it is desirable to extend our error-mitigation procedure to larger systems with complicated geometry of the physical connections between qubits. To realize this goal it is necessary to efficiently perform detector tomography of multi-qubit devices and understand if the readout correction is possible without the necessity to perform estimation of the full probability distribution.

Code availability

We are making Python code implementing the ideas presented in this work publicly available in the form of GitHub repository: <https://github.com/fbm2718/QREM>. We intend to further develop the repository in the future, adding new functionalities and tutorial Jupiter notebooks, which we hope will make our method more useful in practice.

Acknowledgments

We thank Á. Budai, R. Demkowicz-Dobrzański, M. Horodecki, H. Pashayan, A. Pályi, and J. Tworzydło for

interesting, fruitful discussions. We are especially grateful to Ł. Pawela for running our code on the Rigetti device. M.O. acknowledges the support of Homing programme of the Foundation for Polish Science. F.B.M acknowledges the support by the Foundation for Polish Science through IRAP project (contract no.2018/MAB/5). Homing and IRAP programmes are co-financed by EU within Smart Growth Operational Programme. Z.Z. was supported by the Hungarian National Research, Development and Innovation Office (NKFIH) through Grants No. K124351, K124152, K124176 KH129601, and the Hungarian Quantum Technology National Excellence Program (Project No. 2017-1.2.1-NKP-2017-00001); and he was also partially funded by the János Bolyai and the Bolyai+ Scholarships. We acknowledge the use of the IBM Quantum Experience and Rigetti Forest SDK for this work. The views expressed are those of the authors and do not reflect the official policy or position of IBM or the IBM Quantum Experience team, or the Rigetti team.

NOTE ADDED

Upon completion of this manuscript, we became aware of a recent work [62] that proposed an analogous scheme for the mitigation of readout errors. Furthermore, we note that IBM's qiskit package has been recently updated to include the readout error correction scheme, which seems to rely on an analogous procedure.

References

- [1] John Preskill. “Quantum Computing in the NISQ era and beyond”. In: *Quantum* 2 (Aug. 2018), p. 79. ISSN: 2521-327X. DOI: <https://doi.org/10.22331/q-2018-08-06-79>.
- [2] Héctor Abraham et al. *Qiskit: An Open-source Framework for Quantum Computing*. 2019. DOI: [10.5281/zenodo.2562110](https://doi.org/10.5281/zenodo.2562110).
- [3] IBM. <https://quantumexperience.ng.bluemix.net/qx/>. Access: 2018.12.28. URL: <https://quantumexperience.ng.bluemix.net/qx/>.
- [4] Rigetti. <https://www.rigetti.com/forest>. Access: 2018.12.28. URL: <https://www.rigetti.com/forest>.
- [5] D-Wave. <https://cloud.dwavesys.com/qubist/>. [Access: 2018.12.28]. URL: <https://cloud.dwavesys.com/qubist/>.
- [6] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. DOI: [10.1017/CBO9780511976667](https://doi.org/10.1017/CBO9780511976667).
- [7] I. M. Georgescu, S. Ashhab, and Franco Nori. “Quantum simulation”. In: *Reviews of Modern Physics* 86 (Jan. 2014), pp. 153–185. DOI: [10.1103/RevModPhys.86.153](https://doi.org/10.1103/RevModPhys.86.153). arXiv: [1308.6253](https://arxiv.org/abs/1308.6253) [quant-ph].
- [8] Kentaro Tamura and Yutaka Shikano. *Quantum Random Numbers generated by the Cloud Superconducting Quantum Computer*. 2019. arXiv: [1906.04410](https://arxiv.org/abs/1906.04410) [quant-ph].

- [9] Ying Li and Simon C. Benjamin. “Efficient variational quantum simulator incorporating active error minimisation”. In: *Phys Rev X* 7.021050 (2017), p. 021050. DOI: <https://doi.org/10.1103/PhysRevX.7.021050>. arXiv: 1611.09301 [quant-ph].
- [10] Abhinav Kandala et al. “Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets”. In: *Nature* 549 (2017), pp. 242–246. DOI: [10.1038/nature23879](https://doi.org/10.1038/nature23879). arXiv: 1704.05018 [quant-ph].
- [11] Abhinav Kandala et al. “Extending the computational reach of a noisy superconducting quantum processor”. In: *Nature* 567.491 (2019), p. 491. DOI: <https://doi.org/10.1038/s41586-019-1040-7>. arXiv: 1805.04492 [quant-ph].
- [12] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. “Error Mitigation for Short-Depth Quantum Circuits”. In: *Phys. Rev. Lett.* 119.18, 180509 (Nov. 2017), p. 180509. DOI: [10.1103/PhysRevLett.119.180509](https://doi.org/10.1103/PhysRevLett.119.180509). arXiv: 1612.02058 [quant-ph].
- [13] Suguru Endo, Simon C. Benjamin, and Ying Li. “Practical Quantum Error Mitigation for Near-Future Applications”. In: *Physical Review X* 8, 031027 (2018), p. 031027. DOI: [10.1103/PhysRevX.8.031027](https://doi.org/10.1103/PhysRevX.8.031027). arXiv: 1712.09271 [quant-ph].
- [14] Vickram N. Premakumar and Robert Joynt. “Error Mitigation in Quantum Computers subject to Spatially Correlated Noise”. In: *arXiv e-prints* (2018), arXiv:1812.07076. arXiv: 1812.07076 [quant-ph].
- [15] X. Bonet-Monroig et al. “Low-cost error mitigation by symmetry verification”. In: *Phys. Rev. A* 98 (2018), p. 062339. DOI: [10.1103/PhysRevA.98.062339](https://doi.org/10.1103/PhysRevA.98.062339). arXiv: 1807.10050 [quant-ph].
- [16] Joshua Combes et al. “Logical Randomized Benchmarking”. In: *arXiv e-prints* (2017). arXiv: 1702.03688 [quant-ph].
- [17] Mingyu Sun and Michael R. Geller. “Efficient characterization of correlated SPAM errors”. In: *arXiv e-prints* (2018). arXiv: 1810.10523 [quant-ph].
- [18] J. S. Lundeen et al. “Tomography of quantum detectors”. In: *Nature Physics* 5 (Nov. 2008), p. 27. URL: <http://dx.doi.org/10.1038/nphys1133>.
- [19] Lijian Zhang et al. “Mapping coherence in measurement via full quantum tomography of a hybrid optical detector”. In: *Nature Photonics* 6 (May 2012), p. 364. DOI: [10.1038/nphoton.2012.107](https://doi.org/10.1038/nphoton.2012.107).
- [20] Lijian Zhang et al. “Recursive quantum detector tomography”. In: *New Journal of Physics* 14, 115005 (Nov. 2012), p. 115005. DOI: [10.1088/1367-2630/14/11/115005](https://doi.org/10.1088/1367-2630/14/11/115005). arXiv: 1207.3501 [quant-ph].
- [21] J. J. Renema et al. “Modified detector tomography technique applied to a superconducting multiphoton nanodetector”. In: *Opt. Express* 20.3 (2012), pp. 2806–2813. DOI: [10.1364/OE.20.002806](https://doi.org/10.1364/OE.20.002806). URL: <http://www.opticsexpress.org/abstract.cfm?URI=oe-20-3-2806>.
- [22] J. Z. Blumoff et al. “Implementing and Characterizing Precise Multiqubit Measurements”. In: *Phys. Rev. X* 6 (3 2016), p. 031041. DOI: [10.1103/PhysRevX.6.031041](https://doi.org/10.1103/PhysRevX.6.031041). URL: <https://link.aps.org/doi/10.1103/PhysRevX.6.031041>.
- [23] Jens Koch et al. “Charge-insensitive qubit design derived from the Cooper pair box”. In: *Phys. Rev. A* 76 (4 2007), p. 042319. DOI: [10.1103/PhysRevA.76.042319](https://doi.org/10.1103/PhysRevA.76.042319). URL: <https://link.aps.org/doi/10.1103/PhysRevA.76.042319>.
- [24] Michał Oszmaniec, Filip B. Maciejewski, and Zbigniew Puchała. “Simulating all quantum measurements using only projective measurements and postselection”. In: *Physical Review A* 100.1 (2019). ISSN: 2469-9934. DOI: [10.1103/PhysRevA.100.012351](https://doi.org/10.1103/PhysRevA.100.012351). URL: <http://dx.doi.org/10.1103/PhysRevA.100.012351>.
- [25] Lov K. Grover. “A fast quantum mechanical algorithm for database search”. In: *arXiv e-prints* (1996), quant-ph/9605043. arXiv: [quant-ph/9605043](https://arxiv.org/abs/quant-ph/9605043) [quant-ph].
- [26] E. Bernstein and U. Vazirani. “Quantum complexity theory”. In: *Proc. of the Twenty-Fifth Annual ACM Symposium on Theory of Computing (STOC '93)*. 1993, 11–20. DOI: [10.1145/167088.167097](https://doi.org/10.1145/167088.167097).
- [27] Asher Peres. *Quantum theory: Concepts and methods*. Vol. 57. Springer Science & Business Media, 2006. DOI: <https://doi.org/10.1007/0-306-47120-5>.
- [28] Zdeněk Hradil et al. “3 Maximum-Likelihood Methods in Quantum Mechanics”. In: *Quantum State Estimation*. Ed. by Matteo Paris and Jaroslav Řeháček. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 59–112. ISBN: 978-3-540-44481-7. DOI: [10.1007/978-3-540-44481-7_3](https://doi.org/10.1007/978-3-540-44481-7_3).
- [29] Jaromír Fiurášek. “Maximum-likelihood estimation of quantum measurement”. In: *Physical Review A* 64, 024102 (Aug. 2001), p. 024102. DOI: [10.1103/PhysRevA.64.024102](https://doi.org/10.1103/PhysRevA.64.024102). arXiv: [quant-ph/0101027](https://arxiv.org/abs/quant-ph/0101027) [quant-ph].
- [30] Aram W. Harrow and Ashley Montanaro. “Quantum computational supremacy”. In: *Nature* 549 (Sept. 2017), pp. 203–209. DOI: [10.1038/nature23458](https://doi.org/10.1038/nature23458). arXiv: 1809.07442 [quant-ph].
- [31] Hakop Pashayan, Stephen D. Bartlett, and David Gross. “From estimation of quantum probabilities to simulation of quantum circuits”. In: *Quantum* 4 (2020), p. 223. ISSN: 2521-327X. DOI: [10.22331/q-2020-01-13-223](https://doi.org/10.22331/q-2020-01-13-223). URL: <http://dx.doi.org/10.22331/q-2020-01-13-223>.

- [32] Miguel Navascués and Sandu Popescu. “How Energy Conservation Limits Our Measurements”. In: *Phys. Rev. Lett.* 112, 140502 (Apr. 2014), p. 140502. DOI: [10.1103/PhysRevLett.112.140502](https://doi.org/10.1103/PhysRevLett.112.140502). arXiv: [1211.2101](https://arxiv.org/abs/1211.2101) [quant-ph].
- [33] Z. Puchała et al. “Strategies for optimal single-shot discrimination of quantum measurements”. In: *Phys. Rev. A* 98.042103 (2018), p. 042103. DOI: [10.1103/PhysRevA.98.042103](https://doi.org/10.1103/PhysRevA.98.042103). arXiv: [1804.05856](https://arxiv.org/abs/1804.05856) [quant-ph].
- [34] Zbigniew Puchała et al. “Multiple-shot and unambiguous discrimination of von Neumann measurements”. In: *arXiv e-prints*, arXiv:1810.05122 (2018), arXiv:1810.05122. arXiv: [1810.05122](https://arxiv.org/abs/1810.05122) [quant-ph].
- [35] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. DOI: [10.1017/9781316848142](https://doi.org/10.1017/9781316848142).
- [36] Erkkka Haapasalo, Teiko Heinosaari, and Juha-Pekka Pellonpää. “Quantum measurements on finite dimensional systems: relabeling and mixing”. In: *Quant. Inf. Process.* 11.6 (2012), pp. 1751–1763. DOI: [10.1007/s11128-011-0330-2](https://doi.org/10.1007/s11128-011-0330-2).
- [37] M. Oszmaniec et al. “Simulating Positive-Operator-Valued Measures with Projective Measurements”. In: *Physical Review Letters* 119.19, 190501 (Nov. 2017), p. 190501. DOI: [10.1103/PhysRevLett.119.190501](https://doi.org/10.1103/PhysRevLett.119.190501). arXiv: [1609.06139](https://arxiv.org/abs/1609.06139) [quant-ph].
- [38] Leonardo Guerini et al. “Operational framework for quantum measurement simulability”. In: *Journal of Mathematical Physics* 58.9, 092102 (2017), p. 092102. DOI: [10.1063/1.4994303](https://doi.org/10.1063/1.4994303). arXiv: [1705.06343](https://arxiv.org/abs/1705.06343) [quant-ph].
- [39] E. Knill et al. “Randomized benchmarking of quantum gates”. In: *Physical Review A* 77, 012307 (Jan. 2008), p. 012307. DOI: [10.1103/PhysRevA.77.012307](https://doi.org/10.1103/PhysRevA.77.012307). arXiv: [0707.0963](https://arxiv.org/abs/0707.0963) [quant-ph].
- [40] Jay M. Gambetta et al. “Characterization of Addressability by Simultaneous Randomized Benchmarking”. In: *Phys. Rev. Lett.* 109, 240504 (Dec. 2012), p. 240504. DOI: [10.1103/PhysRevLett.109.240504](https://doi.org/10.1103/PhysRevLett.109.240504). arXiv: [1204.6308](https://arxiv.org/abs/1204.6308) [quant-ph].
- [41] Madalin Guta et al. “Fast state tomography with optimal error bounds”. In: *arXiv e-prints* (2018), arXiv:1809.11162. arXiv: [1809.11162](https://arxiv.org/abs/1809.11162) [quant-ph].
- [42] T. Weissman et al. “Inequalities for the L1 Deviation of the Empirical Distribution”. In: *Technical Report HPL-2003-97R1, Hewlett-Packard Labs* (Aug. 2003).
- [43] M. S. Andersen, J. Dahl, and L. Vandenbergh. *CVXOPT: A Python package for convex optimization, version 1.2*. 2019. URL: <https://cvxopt.org/>.
- [44] IBM. *Qiskit Github repository*. Access: 2019.07.09. URL: <https://github.com/Qiskit/ibmq-device-information/tree/master/backends>.
- [45] Robin Blume-Kohout et al. “Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit”. In: *arXiv e-prints*, arXiv:1310.4492 (2013), arXiv:1310.4492. arXiv: [1310.4492](https://arxiv.org/abs/1310.4492) [quant-ph].
- [46] Seth T. Merkel et al. “Self-consistent quantum process tomography”. In: *Phys. Rev. A* 87.6, 062119 (2013), p. 062119. DOI: [10.1103/PhysRevA.87.062119](https://doi.org/10.1103/PhysRevA.87.062119). arXiv: [1211.0322](https://arxiv.org/abs/1211.0322) [quant-ph].
- [47] Michael D. Mazurek et al. “Experimentally bounding deviations from quantum theory in the landscape of generalized probabilistic theories”. In: *arXiv e-prints*, arXiv:1710.05948 (2017), arXiv:1710.05948. arXiv: [1710.05948](https://arxiv.org/abs/1710.05948) [quant-ph].
- [48] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. “A Quantum Approximate Optimization Algorithm”. In: *arXiv e-prints*, arXiv:1411.4028 (2014), arXiv:1411.4028. arXiv: [1411.4028](https://arxiv.org/abs/1411.4028) [quant-ph].
- [49] John A. Smolin, Jay M. Gambetta, and Graeme Smith. “Maximum Likelihood, Minimum Effort”. In: *Phys. Rev. Lett.* 108.070502 (2012), p. 070502. DOI: [10.1103/PhysRevLett.108.070502](https://doi.org/10.1103/PhysRevLett.108.070502). arXiv: [1106.5458](https://arxiv.org/abs/1106.5458) [quant-ph].
- [50] Benjamin Schumacher. “Sending quantum entanglement through noisy channels”. In: *arXiv e-prints*, quant-ph/9604023 (1996), quant-ph/9604023. arXiv: [quant-ph/9604023](https://arxiv.org/abs/quant-ph/9604023) [quant-ph].
- [51] Pawel Horodecki, Michal Horodecki, and Ryszard Horodecki. “General teleportation channel, singlet fraction and quasi-distillation”. In: *arXiv e-prints*, quant-ph/9807091 (1998), quant-ph/9807091. arXiv: [quant-ph/9807091](https://arxiv.org/abs/quant-ph/9807091) [quant-ph].
- [52] Anthony Chefles. “Unambiguous discrimination between linearly independent quantum states”. In: *Physics Letters A* 239.6 (1998), pp. 339–347. DOI: [10.1016/S0375-9601\(98\)00064-4](https://doi.org/10.1016/S0375-9601(98)00064-4). arXiv: [quant-ph/9807022](https://arxiv.org/abs/quant-ph/9807022) [quant-ph].
- [53] Stephen M. Barnett and Sarah Croke. “Quantum state discrimination”. In: *Advances in Optics and Photonics* 1.2 (2009), p. 238. DOI: [10.1364/AOP.1.000238](https://doi.org/10.1364/AOP.1.000238). arXiv: [0810.1970](https://arxiv.org/abs/0810.1970) [quant-ph].
- [54] Joseph M. Renes et al. “Symmetric informationally complete quantum measurements”. In: *Journal of Mathematical Physics* 45.6 (2004), pp. 2171–2180. DOI: [10.1063/1.1737053](https://doi.org/10.1063/1.1737053). arXiv: [quant-ph/0310075](https://arxiv.org/abs/quant-ph/0310075) [quant-ph].
- [55] Satoshi Ishizaka and Tohya Hiroshima. “Asymptotic Teleportation Scheme as a Universal Programmable Quantum Processor”. In: *Phys. Rev. Lett.* 101.24, 240501 (2008), p. 240501. DOI: [10.1103/PhysRevLett.101.240501](https://doi.org/10.1103/PhysRevLett.101.240501). arXiv: [0807.4568](https://arxiv.org/abs/0807.4568) [quant-ph].

- [56] Andrew M. Childs and Wim van Dam. “Quantum algorithms for algebraic problems”. In: *Rev. Mod. Phys.* 82 (1 2010), pp. 1–52. DOI: [10.1103/RevModPhys.82.1](https://doi.org/10.1103/RevModPhys.82.1). URL: <https://link.aps.org/doi/10.1103/RevModPhys.82.1>.
- [57] John Preskill. “Quantum computing and the entanglement frontier”. In: *arXiv e-prints* (Mar. 2012). arXiv: [1203.5813](https://arxiv.org/abs/1203.5813) [quant-ph].
- [58] Patrick J. Coles et al. “Quantum Algorithm Implementations for Beginners”. In: *arXiv e-prints* (2018). arXiv: [1804.03719](https://arxiv.org/abs/1804.03719) [cs.ET].
- [59] Additional experimental data, such as the exact form of the implemented and reconstructed operators, is accessible online in the github repository – https://github.com/fbm2718/mitigation_paper2019. For the Python code implementing the ideas from the work, see GitHub repository: <https://github.com/fbm2718/QREM>.
- [60] Nikolaž Moll et al. “Quantum optimization using variational algorithms on near-term quantum devices”. In: *Quantum Science and Technology* 3 (2018), p. 030503. DOI: [10.1088/2058-9565/aab822](https://doi.org/10.1088/2058-9565/aab822). arXiv: [1710.01022](https://arxiv.org/abs/1710.01022) [quant-ph].
- [61] Jinzhao Wang, Volkher B. Scholz, and Renato Renner. “Confidence Polytopes in Quantum State Tomography”. In: *Phys. Rev. Lett.* 122.19, 190401 (2019), p. 190401. DOI: [10.1103/PhysRevLett.122.190401](https://doi.org/10.1103/PhysRevLett.122.190401). arXiv: [1808.09988](https://arxiv.org/abs/1808.09988) [quant-ph].
- [62] Yanzhu Chen et al. “Detector tomography on IBM quantum computers and mitigation of an imperfect measurement”. In: *Physical Review A* 100.5 (2019). ISSN: 2469-9934. DOI: [10.1103/PhysRevA.100.052315](https://doi.org/10.1103/PhysRevA.100.052315). URL: <http://dx.doi.org/10.1103/PhysRevA.100.052315>.

A Proofs of technical statements

Here we prove technical results that were given without proofs in the main text. First, we present the proof of equality (16), which we recall here for the Reader’s convenience

$$D_{op}(\mathbf{M}^{(K)}, \mathbf{P}^{(K)}) = 1 - \prod_{i=1}^K (1 - D_{op}(\Lambda_i \mathbf{P}_i, \mathbf{P}_i)) . \quad (35)$$

In the above formula $\mathbf{M}^{(K)} = (\otimes_{i=1}^K \Lambda_i) \mathbf{P}^{(K)}$ is the noisy version of the multi-qubit projective measurement $\mathbf{P}^{(K)} = \otimes_{i=1}^K \mathbf{P}_i$ (see Section 3 for more details). Directly from the definition of the operational distance given in Eq. (4), one obtains

$$D_{op}(\mathbf{M}^{(K)}, \mathbf{P}^{(K)}) = \max_{\rho} D_{TV}(\mathbf{p}^{\mathbf{M}^{(K)}}, \mathbf{p}^{\mathbf{P}^{(K)}}) , \quad (36)$$

where $\mathbf{p}^{\mathbf{M}}$ is the probability vector of probabilities generated by the statistics of measurements of \mathbf{M} on a quantum state ρ . Using the linearity of the Born rule in quantum states and the fact that both $\mathbf{M}^{(K)}$ and $\mathbf{P}^{(K)}$ are diagonal in the computational basis in $(\mathbb{C}^2)^{\otimes K}$, one can immediately prove that the optimal state in (36) can be also chosen to be *diagonal* in the computational basis. Moreover, from the linearity of the Born rule it also follows that $D_{TV}(\mathbf{p}^{\mathbf{M}^{(K)}}, \mathbf{p}^{\mathbf{P}^{(K)}})$ is a convex function of ρ and therefore the maximum in (36) is always attained for *pure states*. Combining these two observations, we see that the optimal input state ρ^* will be a particular computational basis state: $\rho^* = |\underline{i}\rangle\langle\underline{i}| = |i_1\rangle\langle i_1| \otimes |i_2\rangle\langle i_2| \otimes \dots \otimes |i_K\rangle\langle i_K|$. The optimization over such states becomes now easy since the projective measurement $\mathbf{P}^{(K)}$ acts as the identity transformation on probability distributions (i.e., on diagonal quantum states). Therefore, we get

$$D_{op}(\mathbf{M}^{(K)}, \mathbf{P}^{(K)}) = \max_{\underline{i}} D_{TV}(\otimes_{i=1}^K \Lambda_i |\underline{i}\rangle\langle\underline{i}|, |\underline{i}\rangle\langle\underline{i}|) = 1 - \prod_{i=1}^K (1 - \max\{p_i, q_i\}) , \quad (37)$$

which, by the virtue of (14) concludes the proof of (35).

The second technical result used without a proof in the main part of the paper was subadditivity of operational distance, i.e.,

$$D_{op}(\mathbf{M}^{(1)} \otimes \mathbf{M}^{(2)}, \mathbf{N}^{(1)} \otimes \mathbf{N}^{(2)}) \leq D_{op}(\mathbf{M}^{(1)}, \mathbf{N}^{(1)}) + D_{op}(\mathbf{M}^{(2)}, \mathbf{N}^{(2)}) , \quad (38)$$

where $\mathbf{M}^{(1)} \otimes \mathbf{M}^{(2)}$ and $\mathbf{N}^{(1)} \otimes \mathbf{N}^{(2)}$ are POVMs on a compound quantum system $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ whose effects are build from single party measurements according to the following prescription:

$$(\mathbf{M}^{(1)} \otimes \mathbf{M}^{(2)})_{(i,j)} := M_i^{(1)} \otimes M_j^{(2)} . \quad (39)$$

In order to prove (37) we will again use Eq. (4) which yields

$$D_{op}(\mathbf{M}^{(1)} \otimes \mathbf{M}^{(2)}, \mathbf{N}^{(1)} \otimes \mathbf{N}^{(2)}) = \max_{\rho} \frac{1}{2} \sum_{i,j} \left| \text{tr} \left(\rho M_i^{(1)} \otimes M_j^{(2)} - \rho N_i^{(1)} \otimes M_j^{(2)} \right) \right|. \quad (40)$$

Using triangle inequality we obtain that for all quantum states state ρ

$$\frac{1}{2} \sum_{i,j} \left| \text{tr} \left(\rho M_i^{(1)} \otimes M_j^{(2)} - \rho N_i^{(1)} \otimes M_j^{(2)} \right) \right| \leq \frac{1}{2} \sum_{i,j} \left| \text{tr} \left(\rho M_i^{(1)} \otimes [M_j^{(2)} - N_j^{(2)}] \right) \right| + \frac{1}{2} \sum_{i,j} \left| \text{tr} \left(\rho [M_i^{(1)} - N_i^{(1)}] \otimes N_j^{(2)} \right) \right|. \quad (41)$$

To proceed we need to introduce auxiliary probability distributions and quantum states

$$p_i^{(2)} = \text{tr} \left(M_i^{(1)} \otimes \mathbb{I} \rho \right), \quad p_j^{(1)} = \text{tr} \left(\mathbb{I} \otimes N_j^{(2)} \rho \right), \quad (42)$$

$$\rho_i^{(2)} = \text{tr}_1 \left(\sqrt{M_i^{(1)}} \otimes \mathbb{I} \rho \sqrt{M_i^{(1)}} \otimes \mathbb{I} \right) / p_i^{(2)}, \quad \rho_j^{(1)} = \text{tr}_2 \left(\mathbb{I} \otimes \sqrt{N_j^{(2)}} \rho \mathbb{I} \otimes \sqrt{N_j^{(2)}} \right) / p_j^{(1)}. \quad (43)$$

After simple computations using these auxiliary objects we obtain

$$\frac{1}{2} \sum_{i,j} \left| \text{tr} \left(\rho M_i^{(1)} \otimes M_j^{(2)} - \rho N_i^{(1)} \otimes M_j^{(2)} \right) \right| \leq \frac{1}{2} \sum_{i,j} p_i^{(2)} \left| \text{tr} \left(\rho_i^{(2)} [M_j^{(2)} - N_j^{(2)}] \right) \right| + \frac{1}{2} \sum_{i,j} p_j^{(1)} \left| \text{tr} \left(\rho_j^{(1)} [M_i^{(1)} - N_i^{(1)}] \right) \right|. \quad (44)$$

Finally, directly from the definition of operational distance (see Eq. (4)) we obtain

$$\frac{1}{2} \sum_{i,j} p_i^{(2)} \left| \text{tr} \left(\rho_i^{(2)} [M_j^{(2)} - N_j^{(2)}] \right) \right| = \frac{1}{2} \sum_i p_i^{(2)} \sum_j \left| \text{tr} \left(\rho_i^{(2)} [M_j^{(2)} - N_j^{(2)}] \right) \right| \leq D_{op}(\mathbf{M}^{(2)}, \mathbf{N}^{(2)}) \quad (45)$$

and analogously

$$\frac{1}{2} \sum_{i,j} p_j^{(1)} \left| \text{tr} \left(\rho_j^{(1)} [M_i^{(1)} - N_i^{(1)}] \right) \right| \leq D_{op}(\mathbf{M}^{(1)}, \mathbf{N}^{(1)}). \quad (46)$$

We conclude the proof by inserting the above two inequalities into the right hand side of Eq. (41).

B State dependence of correction

In Section 3, we developed a scheme for the mitigation of readout errors which relies solely on classical post-processing of the vector of experimental statistics. We have shown that this procedure mitigates noise perfectly, provided the measuring device is affected by a certain type of classical noise. In Section 4, we analyzed what effects on the mitigation scheme arises from the deviations from the adopted noise model and from the finite-size statistics. In particular, we have pointed out that the procedure is highly on the quantum state being measured. Specifically, it might happen that correction actually worsens the results compared to the non-post-processed case. To address this issue, we proposed a criterion for deciding whether the mitigation has been successful (Eq. (29)), which we will now rewrite here for convenience of the Reader

$$\delta + \alpha < D_{op}(\mathbf{M}^{\text{exp}}, \mathbf{M}^{\text{ideal}}) + \epsilon \Rightarrow \text{mitigation succesful}. \quad (47)$$

In what follows we provide provide numerical arguments for approximate correctness of the above rule for single-qubit detectors and two-qubit uncorrelated detectors.

First, let us note that the above inequality is in fact comparison of two bounds. On the LHS, we have the upper bound for the error for our mitigation procedure, i.e., the upper bound on TV-distance (Eq. (3)) between post-processed statistics $\mathbf{p}_{\text{exp}}^*$ and the ideal statistics $\mathbf{p}_{\text{ideal}}$ that one would have obtained on the non-noisy detector. On the RHS, we have an upper bound for TV-distance between non-post-processed statistics $\mathbf{p}_{\text{exp}}^{\text{est}}$, and the $\mathbf{p}_{\text{ideal}}$. This is the *worst possible error* one would expect on the noisy detector without mitigation. Of course, in the real experiment, we do not have access to any particular $\mathbf{p}_{\text{ideal}}$, which makes such general figures of merit particularly useful. However, in the case of numerical simulations, when $\mathbf{p}_{\text{ideal}}$ can be computed, we can use a quantum state sensitive criterion analogous to Eq. (47), namely

$$D_{TV}(\mathbf{p}_{\text{exp}}^*, \mathbf{p}_{\text{ideal}}) < D_{TV}(\mathbf{p}_{\text{exp}}^{\text{est}}, \mathbf{p}_{\text{ideal}}) \Rightarrow \text{mitigation succesful}. \quad (48)$$

It is now natural to ask how often the above criterion is satisfied, when Haar-random quantum states are measured by the noisy detector \mathbf{M} . To address this question, we perform the following numerical procedure.

1. Generate L Haar-random pure quantum states.

For each quantum state:

- i. Calculate probability vectors \mathbf{p}_{exp} and $\mathbf{p}_{\text{ideal}}$, that quantum state generates, when measured by noisy detector \mathbf{M} and ideal detector \mathbf{P} , respectively.
- ii. Sample N times from the probability distribution \mathbf{p}_{exp} , obtaining $\mathbf{p}_{\text{exp}}^{\text{est}}$.
- iii. Compute $\Lambda^{-1}\mathbf{p}_{\text{exp}}^{\text{est}}$ and check if it is physical probability vector. If no, solve problem defined in Eq. (25), obtaining $\mathbf{p}_{\text{exp}}^*$ and calculate α defined in Eq. (26). If yes, identify $\mathbf{p}_{\text{exp}}^* = \mathbf{p}_{\text{exp}}^{\text{est}}$ and set $\alpha = 0$.
- iv. Calculate $D_{TV}(\mathbf{p}_{\text{exp}}^*, \mathbf{p}_{\text{ideal}})$ and $D_{TV}(\mathbf{p}_{\text{exp}}^{\text{est}}, \mathbf{p}_{\text{ideal}})$. If inequality in Eq. (47) is satisfied, the mitigation was successful in that case.

2. Calculate fraction f of quantum states for which mitigation was successful.

We have studied how fraction f changes with growing ratio $\frac{\delta + \alpha}{D_{op}(\mathbf{M}, \mathbf{P}) + \epsilon}$ in the case of single-qubit detectors. To this aim, we have implemented the above algorithm for the range of magnitudes of off-diagonal terms of POVM's elements, while keeping diagonal terms from actual IBM's data. Results are shown in Fig. 12. For all qubits the fraction of corrected statistics crosses 50% in the region in which $\delta + \alpha \approx D_{op}(\mathbf{M}, \mathbf{P}) + \epsilon$, which backs up rule give in Eq. (47). Furthermore, the fraction of successful error mitigations for actual experimental data lies between 88% and 99%, which suggests that for single-qubit experiments our protocol should be helpful in most cases.

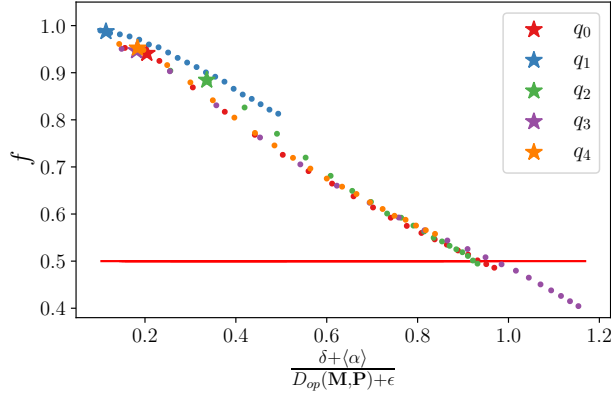


Figure 12: Dependence between the fraction f for which the mitigation is successful and the ratio $\frac{\delta + \alpha}{D_{op}(\mathbf{M}, \mathbf{P}) + \epsilon}$. Each point corresponds to different non-classical part z of the noise. For each point, fraction f was calculated for $L = 10000$ Haar-random quantum states. The sampling size for the probability estimation was fixed at $N = 8192$. The probability of error in the statistical part of the bound was fixed at $\Pr_{\text{err}} = 0.01$. The dependence of α on the particular case has been overcome by taking the average $\langle \alpha \rangle$ over all ($L = 10000$) implementations. Stars denote actual POVMs obtained via QDT on IBM's quantum device.

For the two-qubit case we have tested POVMs created via the tensor product of single-qubit POVMs with increasing off-diagonal terms, again leaving diagonal terms from experimental data. Results of simulation are shown on Figure 13. Similarly to single-qubit case, fraction of successful mitigations crosses 50% around the regime where Eq. (47) becomes unsatisfied. Furthermore, for actual experimental data f takes values between 98.86% and 99.99%, which suggests that for *uncorrelated* pairs of qubits our mitigation procedure should be successful for generic quantum states.

C Additional experimental data

In the following section, we present explicit forms of matrices representing POVMs reconstructed in QDTs discussed in Section 6. Furthermore, we present the dates of execution of all our experiments presented throughout the main part of the paper. For further experimental data, we refer the Reader to the online repository [59].

C.1 Quantum Detector Tomography

Here we present the explicit form of the first effects of the exemplary POVMs reconstructed in single-qubit detector tomographies on all five qubits of IBM's *ibmqx4* and first five qubits of Rigetti's *Aspen-4-16Q-A*. The second effect of each the POVMs can then be automatically obtained as the complement to identity.

a. IBM

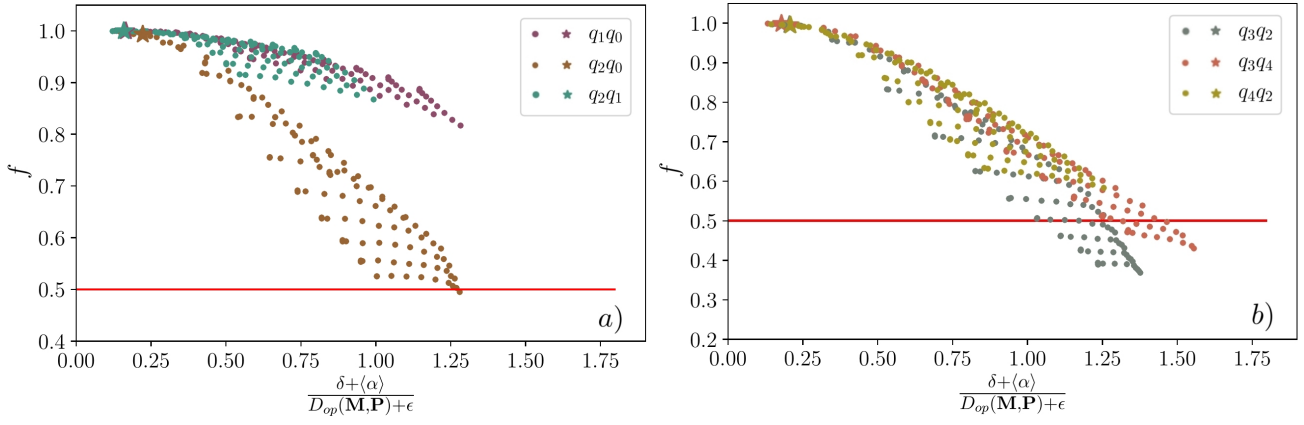


Figure 13: Dependence between the fraction f for which the mitigation is successful and the ratio $\frac{\delta + \langle \alpha \rangle}{D_{op}(\mathbf{M}, \mathbf{P}) + \epsilon}$. The data convention is fully analogous to that of Fig 12. Fraction f was calculated for $L = 10000$ Haar-random quantum states, each probability vector was sampled $N = 8192$ times and the probability of error in the statistical part of the bound was fixed at $\text{Pr}_{err} = 0.01$. a) and b) correspond to different pairs of qubits, which were separated for clarity. POVMs were created by the tensor product of single-qubit measurements with increasing non-classical parts.

$$\begin{aligned}
 M_1^{q_0} &= \begin{pmatrix} 0.963 & 0.004 \\ 0.004 & 0.137 \end{pmatrix}, & M_1^{q_1} &= \begin{pmatrix} 0.99 & 0.002 - 0.001i \\ 0.002 + 0.001i & 0.37 \end{pmatrix}, \\
 M_1^{q_2} &= \begin{pmatrix} 0.986 & -0.001 \\ -0.001 & 0.065 \end{pmatrix}, & M_1^{q_3} &= \begin{pmatrix} 0.919 & 0.003 - 0.003i \\ 0.003 + 0.003i & 0.148 \end{pmatrix}, \\
 M_1^{q_4} &= \begin{pmatrix} 0.98 & -0.002i \\ 0.002i & 0.155 \end{pmatrix}.
 \end{aligned} \tag{49}$$

b. Rigetti

$$\begin{aligned}
 M_1^{q_0} &= \begin{pmatrix} 0.975 & -0.002 \\ -0.002 & 0.124 \end{pmatrix}, & M_1^{q_1} &= \begin{pmatrix} 0.966 & 0.002 + 0.002i \\ 0.002 - 0.002i & 0.101 \end{pmatrix}, \\
 M_1^{q_2} &= \begin{pmatrix} 0.987 & 0.001 - 0.001i \\ 0.001 + 0.001i & 0.066 \end{pmatrix}, & M_1^{q_3} &= \begin{pmatrix} 0.938 & 0.002 + 0.001i \\ 0.002 - 0.001i & 0.184 \end{pmatrix}, \\
 M_1^{q_4} &= \begin{pmatrix} 0.903 & 0.012 - 0.001i \\ 0.012 + 0.001i & 0.155 \end{pmatrix}.
 \end{aligned} \tag{50}$$

C.2 Dates of the experiments

Table 6 contains the dates of execution of experiments performed on IBM's device *ibmqx4*. Fidelities on Rigetti's device given in Tab. 1 were checked via Forest [4] on June 30, 2019, and all tomographic reconstructions on Rigetti's device were done on May 30, 2019.

Location	Date
Tab. 1	February 15
Tab. 2	April 28
Tab. 3	March 12
Tab. 5	May 30
Fig. 6	April 28
Fig. 7	April 28
Fig. 8	April 28
Fig. 9	April 28
Fig. 10	April 28
Fig. 11	February 21

Table 6: The dates of execution of all our experiments performed on IBM's *ibmqx4*. We note that in the case of Tab. (1) the data was obtained via qiskit [2]. All experiments were done in 2019.