University of Warsaw
Faculty of Physics

Joanna Majsak
Record book number: 385858

Mitigation of coherent measurement errors based on characterization of quantum devices

Bachelor's thesis
in the field of Physics

The thesis was written under the supervision of
Dr hab., prof. CFT PAN Michał Oszmaniec
Center for Theoretical Physics
Polish Academy of Sciences

Dr hab., prof. UW Rafał Demkowicz-Dobrzański
Institute of Theoretical Physics
University of Warsaw

Warsaw, September 2022

Summary

In this thesis we propose a way of mitigation of coherent errors present in quantum measurements, in particular in Noisy Intermediate–Scale Quantum (NISQ) devices. The procedure starts with characterization of the error present in a given measurement (called POVM) through Quantum Detector Tomography (QDT). Based on that data, a unitary transformation is chosen to be added to the circuit directly before measurement. The transformation is chosen so that a specific cost function (depending on the tomographically reconstructed POVM and the unitary transformation) is optimized. The choice of 8 cost functions is discussed and justified. 6 of them have operational meaning. 2 are heuristically motivated. A method is presented of finding the explicit form of a local unitary transformation minimizing one of the heuristic functions – the square of the Hilbert–Schmidt norm of the off–diagonal part of a POVM's effects. It is an adaptation to a quantum context of a known method of Approximate Simultaneous Diagonalization (ASD) of a collection of matrices. Using this formalism, the explicit form of a unitary optimizing the other heuristic cost function (inspired by average–case distance between POVMs) is given. The proposed method is tested on data obtained from Detector Overlapping Tomography (DOT) of two state–of–the–art devices: IBM's 127–qubit ibm_washington and Rigetti's 80–qubit ASPEN–M–1.

Keywords

quantum measurement, POVM, measurement error, readout noise, mitigation, quantum gate, device characterization, NISQ, Quantum Detector Tomography, Approximate Simultaneous Diagonalization

Title of thesis in Polish language

Mitygacja koherentnych błędów pomiarowych na podstawie charakteryzacji urządzeń kwantowych

Acknowledgements

I would like to thank my thesis supervisors: Professor Michał Oszmaniec for his time and guidance and Professor Rafał Demkowicz–Dobrzański for his inspiring lectures.

I also extend my gratitude to my colleagues from the Quantin Research Group. In particular I would like to express my appreciation to Filip Maciejewski for his constant assistance and for providing me with relevant data which I use in this thesis.

Contents

Introduction

Background

Since in the 1980s the inspiring vision of quantum machines simulating physics and advancing our understanding of its laws was put forward [1, 2], the field of quantum computing has come a long way. Advances are being made not only towards simulating quantum systems [3]. Now quantum computers have become popular outside of scientific communities. The public imagination is captured by the perspective of the Shor algorithm breaking down modern encryption [4] or the far–reaching promise of quantum computers advancing chemistry [5] and subsequently drug development.

However, error correcting codes, analogous to those in classical computers, are necessary for fault tolerant computation [6]. They require multiple physical qubits to encode one logical qubit. They also necessitate a low error in preparing and measuring those qubits, as well as in operations on them. The required number of qubits and level of noise are prohibitive for implementing error correction in currently available quantum devices. While there is promising ongoing work towards reaching the error correction regime [7], we're not there yet.

We are currently in the era of **Noisy Intermediate-Scale Quantum (NISQ)** devices, called so by Preskill in 2018 [8]. Examples include the largest quantum computers currently advertised by e.g. IBM, Google, Rigetti, which have around 100 qubits. They are all significantly affected by noise, which further limits the kinds of circuits which can be effectively implemented on them.

An intermediate step towards more accurate quantum computation without error correcting codes, tailored for the NISQ devices, is **error mitigation**. It allows for a more accurate estimation of expectation values of observables measured in noisy devices. Some popular examples are: ZNE – zero noise extrapolation (extrapolating results from experiments with varying noise [9]), probabilistic error cancellation – PEC (obtaining the ideal expectation value of an observable from a linear combination of results of experiments on noisy circuits randomly sampled from a quasi–probability distribution corresponding to a decomposition of an ideal gate into noisy gates [9]) and virtual distillation (entangling and measuring multiple copies of a noisy state, obtaining its estimate with reduced error [10]). The effectiveness of some of these error mitigation procedures has been confirmed experimentally, for example on devices based on superconducting qubits [11–16] and trapped-ion qubits [17].

A significant kind of noise in NISQ devices based on superconducting qubits is **measurement (or readout) noise**. Firstly, measurement is present in any computation on a quantum computer. Secondly, the magnitude of readout errors also dominates over state preparation and gate errors in multiple superconducting qubit devices: Google's [18, 19], several of IBM's processors [14–16, 20, 21] and Rigetti's devices [16, 21].

Readout errors have two contributing parts – **classical and coherent (quantum)** errors (see Section 1.6). Classical error is dominant [14–16] and multiple mitigation strategies against it have been proposed.

A well established strategy [14, 16, 22] consists of full characterization of classical noise in a given device through Quantum Detector Tomography (QDT) and then post processing data from experiments, effectively reversing the effects of that noise.

In work expanding on this method [20, 21, 23, 24] attempts were made to reduce the cost in quantum and classical resources, which it requires. Another method based on device characterization was proposed in [25]. A procedure which doesn't rely on the knowledge of the noise model was put forward in [26]. In [27] the authors propose to apply gates before measurement, to effectively cancel out coherent errors, provided ideal gates and infinite experimental statistics. A more detailed description of those methods can be found in Section 1.6.2.

The method of mitigation of coherent errors which we propose is similar to the last example. It aims to supplement the currently available methods of mitigation of classical errors. Our method is based on device characterization, requires high-fidelity gates but doesn't require a high number of experiment runs.

Outline of the thesis

This thesis is structured as follows:

- In Chapter 1 we introduce the necessary mathematical and physical background.
- In Chapter 2 we present our contribution. The proposed method of mitigation of measurement errors with the use of gates is explained.
- In Chapter 3 we test our method on data from IBM's 127–qubit ibm_washington and Rigetti's 80–qubit ASPEN–M–1 devices.
- In Chapter 4 we summarize the thesis and outline directions for future research.

Chapter 1

Theoretical background

In this chapter we introduce the notation used in the following work and the notions needed to understand it. We first present the mathematical notions and then the quantum mechanics background.

1.1 Mathematical notions

1.1.1 Notation

The **Hermitian conjugate** of the operator A will be denoted by $A^\dagger$.

The **complex conjugate** of a complex number a will be denoted by $\overline{a}$.

$\mathrm{O}(n), \mathrm{SO}(n), \mathrm{U}(n), \mathrm{SU}(n)$ denote respectively: the orthogonal, the special orthogonal, the unitary and the special unitary groups of $n \times n$ matrices.

The term in the i–th row and j–th column of a matrix M will be denoted by m_{ij}.

The $\mathrm{diag}(M)$ map (also known as the completely dephasing channel – see Section 1.2.2) returns the diagonal part of a $d \times d$ matrix M, given by:

$$\mathrm{diag}(M) = \sum_{i=1}^{d} m_{ii} |i\rangle\langle i|, \tag{1.1}$$

where $|i\rangle$ are basis vectors of the matrix, written in braket notation.

1.1.2 Definitions

Norms

For a real number $p \geq 1$ the **p-norm** (or ℓ_p norm) of a vector $\mathbf{v} = [v_1, v_2, \ldots v_n]$ from an n–dimensional vector space is defined by:

$$\|\mathbf{v}\|_p = \left(\sum_{i=1}^{n} |v_i|^p \right)^{\frac{1}{p}}. \tag{1.2}$$

For example for $p = 2$ we obtain the Euclidean norm. The **infinity norm** is defined as $\|v\|_\infty = \max_i |v_i|$.

Norms on vector spaces induce **operator norms**. For a linear operator A from the vector space V into W an operator norm $\|\cdot\|_{op}$ can be defined as follows:

$$\|A\|_{op} = \sup\{\|Av\|_W : \|v\|_V \leq 1, v \in V\}, \tag{1.3}$$

where $\|\cdot\|_V$, $\|\cdot\|_W$ are some vector norms on spaces V, W respectively. For example for the $\|\cdot\|_{1-1}$ norm those norms are both chosen to be ℓ_1. For a matrix this norm can be calculated by taking the maximum ℓ_1 norm of its columns.

Hilbert space

A **Hilbert space** $\mathcal{H}$ is a complex vector space equipped with a scalar product. The space is complete in a norm given by this product. In this work we limit ourselves to finite–dimensional spaces, denoted by $\mathcal{H} \cong \mathbb{C}^d$, where d is the dimension. The definitions of the Hilbert–Schmidt inner product and norm given below hold for finite–dimensional spaces, in which they are also called Frobenius inner product and norm.

The **Hilbert-Schmidt inner product** between operators A and B acting on $\mathcal{H} \cong \mathbb{C}^d$ is given by:

$$\langle A|B\rangle_{\mathrm{HS}} = \mathrm{Tr}(A^\dagger B). \tag{1.4}$$

Then the **Hilbert–Schimdt norm** of an operator A acting on $\mathcal{H} \cong \mathbb{C}^d$ is given by:

$$\|A\|_{\mathrm{HS}} = \sqrt{\mathrm{Tr}(A^\dagger A)}. \tag{1.5}$$

For an $n \times n$ matrix M representing A the Hilbert–Schmidt norm can also be calculated as:

$$\|M\|_{\mathrm{HS}} = \sqrt{\sum_{i=1,j=1}^{n} |m_{ij}|^2} = \sqrt{\sum_{i=1}^{n} \lambda_i^2}\,, \tag{1.6}$$

where m_{ij} denotes the matrix element in i-th row and j-th column and λ_i denote the eigenvalues of M. Note the unitary invariance of the Hilbert–Schmidt norm: $\forall_{U\in \mathrm{U(n)}, A\in\mathbb{C}^{n\times n}} \|UA\|_{\mathrm{HS}} = \|AU\|_{\mathrm{HS}} = \|A\|_{\mathrm{HS}}$.

A fact we will often use is that for two operators A, B acting on $\mathcal{H} \cong \mathbb{C}^d$:

$$\|A + B\|_{\mathrm{HS}}^2 = \|A\|_{\mathrm{HS}}^2 + \|B\|_{\mathrm{HS}}^2 + 2\langle A|B\rangle_{\mathrm{HS}}. \tag{1.7}$$

1.2 Formalism of quantum mechanics

Basic notions from quantum mechanics needed to understand this work are introduced briefly here. For a more detailed introduction see e.g. [28].

1.2.1 Quantum states

Quantum states can be represented by operators on a Hilbert space $\mathcal{H} \cong \mathbb{C}^d$. The linear operator $\rho : \mathcal{H} \to \mathcal{H}$ represents a **quantum state** if it fulfills the following two conditions:

$$\mathrm{Tr}(\rho) = 1,\ \rho \geq 0. \tag{1.8}$$

When referring to operators, the ≥ 0 notation indicates that the operator is positive semi–definite. Note that any positive semi–definite operator is necessarily Hermitian and so are operators representing quantum states.

The operators on the finite–dimensional space can then be represented in matrix form. Dirac (bra–ket) notation will be used for vectors. A state ρ is **pure** if there exists a vector $|\psi\rangle \in \mathcal{H} : \langle\psi|\psi\rangle = 1$ such that ρ can be written as:

$$\rho = |\psi\rangle\langle\psi|. \tag{1.9}$$

Or equivalently:

$$\mathrm{Tr}(\rho^2) = 1. \tag{1.10}$$

Otherwise a state is **mixed**.

Note that for a given state ρ the choice of the vector $|\psi\rangle$ is unique only up to multiplication by the phase $\mathrm{e}^{\mathrm{i}\phi}$, where $\phi \in \mathbb{R}$. The collection of normalized vectors $\{|\psi\rangle\}$ differing only by such phase is called a **statevector**. A state's matrix representation is referred to as its **density matrix**. Only pure states can be represented by statevectors.

A **qubit** is a quantum system defined on a 2–dimensional Hilbert space. There are two states of that system, forming its orthonormal basis, whose statevectors are denoted by $|0\rangle$ and $|1\rangle$ and which are called **computational basis states**. Physically they can correspond to some classically interpretable state, such as e.g. being in the ground or excited state. A computational basis can also be defined on a system of more than one qubit, by taking tensor products of all possible basis states of the subsystems. For two qubits the computational basis states are: $|0\rangle \otimes |0\rangle, |0\rangle \otimes |1\rangle, |1\rangle \otimes |0\rangle, |1\rangle \otimes |1\rangle$. For brevity the tensor symbol and multiple kets are often omitted an we can equivalently write e.g. $|00\rangle$ to mean $|0\rangle \otimes |0\rangle$.

For $n > 1$ n–qubit systems these states can also be represented as vectors in 2^n–dimensional space, labelled by integers i:

$$\begin{array}{rccccccc} |i\rangle = & [0 & 0 & \dots & 1 & \dots & 0]^{\mathrm{T}} \\ & \uparrow & & & \uparrow & & \uparrow \\ & 1 & & & i & & 2^n \end{array} \tag{1.11}$$

Note the indexing of i from 1, which is different from the indexing from 0 in the 1–qubit case.

Alternatively these states can be labelled by length–n bit–strings:

$$\mathbf{x} = (x_1, x_2, \dots, x_n) \in \{0,1\}^n, \tag{1.12}$$

where $x_j \in \{0,1\}$ represents whether the j–th qubit is in state $|0\rangle$ or $|1\rangle$. These two conventions are interchangeable. E.g. the vector $|\mathbf{x}\rangle$ labelled by bit–string $\mathbf{x} = [1,0]$ represents the same state as the vector $|i = 2\rangle = [0\,0\,1\,0]^{\mathrm{T}}$.

1.2.2 Quantum gates, unitary transformations and channels

In analogy to logic gates on classical computers, quantum gates on qubits can be defined. A **quantum gate** is a building block from which **unitary transformations** (or unitaries) can be built. A unitary transformation acting on a qubit can be represented by a unitary matrix $U \in \mathrm{SU}(2)$. A transformation on n qubits is represented by a unitary matrix belonging to $\mathrm{SU}(2^n)$. After a transformation U acts on a pure state $|\psi\rangle$, the state $|\psi'\rangle = U|\psi\rangle$ is obtained. Its density matrix is $|\psi'\rangle\langle\psi'| = U|\psi\rangle\langle\psi|U^\dagger$. In general for any initial state ρ, after a unitary transformation U acts on it, the state $\rho' = U\rho U^\dagger$ is obtained.

A notion more general than a unitary transformation is that of quantum channel. A **quantum channel** is a linear map acting between spaces of linear operators on Hilbert spaces. It has to be completely positive and trace preserving (CPTP). The action of a unitary transformation U on a state ρ can be described by the channel Ad_U:

$$\mathrm{Ad}_U(\rho) = U\rho U^\dagger. \tag{1.13}$$

The **completely dephasing channel** acting on an operator cancels out its off-diagonal terms in computational basis, while leaving the diagonal terms intact. As mentioned in 1.1.1, the map $\mathrm{diag}(\cdot)$ (1.1) can be interpreted as the completely dephasing channel.

1.2.3 Quantum measurements

An m–outcome measurement on a system defined on $\mathcal{H} \cong \mathbb{C}^d$ is a collection of n operators $\mathcal{H} \to \mathcal{H}$ (called **effects**) $\mathsf{M} = \{M^{(k)}\}_{k=1}^{m}$ satisfying:

$$\forall_k M^{(k)} \geq 0, \sum_{k=1}^{m} M^{(k)} = \mathbb{1}_d. \tag{1.14}$$

Such a measurement is referred to as a **POVM** - positive operator-valued measure. A special case of a POVM is the **projective** or **von Neumann** measurement. Its effects additionally satisfy the condition:

$$\forall_{k,l} M^{(k)} M^{(l)} = \delta_{kl} M^{(k)}. \tag{1.15}$$

An example of a projective POVM, which is often used in present–day applications, is the **computational basis** POVM. Each of its effects is a projection onto one computational basis state of the system. E.g. for 2 qubits the effects are: $|00\rangle\langle 00|, |01\rangle\langle 01|, |10\rangle\langle 10|, |11\rangle\langle 11|$. Note that for a computational basis POVM the number of its effects m is equal to the number of the system's dimensions d. It is a projective measurement and we will denote a computational basis POVM by P.

Note that, using the convention from Eq. (1.12), the computational basis POVM effects can equivalently be labelled by the bit–strings corresponding to the states onto which they project: $\mathsf{P} = \{P^{(\mathbf{x})}\}_{\mathbf{x}=\{0,1\}^n}$.

Born's rule

Born's rule states that the probability of obtaining the outcome associated with the effect $M^{(k)}$ when measuring a state ρ is given by:

$$p(k|\rho) = \mathrm{Tr}(\rho M^{(k)}). \tag{1.16}$$

The measurement of state ρ with an m-outcome POVM M generates the **probability vector** or distribution:

$$\mathbf{p} = [p_1, p_2, ...p_k]^{\mathrm{T}} = [\mathrm{Tr}(\rho M^{(1)}), \mathrm{Tr}(\rho M^{(2)}), ...\mathrm{Tr}(\rho M^{(k)})]^{\mathrm{T}}. \tag{1.17}$$

Operationally this vector represents how often each outcome would be obtained experimentally, provided an infinite number of experimental runs.

1.2.4 Quantum circuits

Using the building blocks introduced above, a quantum circuit can be constructed. Such circuits are often represented in image form. We will use conventional symbols for states, gates and measurements. The image should be read from left to right. An example of such a circuit is shown in 1.1.

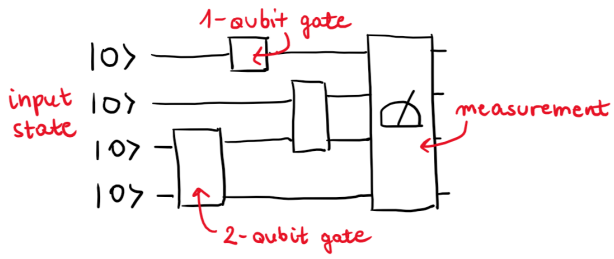


Figure 1.1: Example of a quantum circuit

1.3 Bloch ball representation

1.3.1 State representation

It is often useful to think of the state of a qubit in the **Bloch ball** representation. Any density matrix ρ of a qubit can be decomposed in the basis of the Pauli matrices $\sigma_x, \sigma_y, \sigma_z$ and the identity matrix $\mathbb{1}_2$. The basis is orthogonal with respect to the Hilbert-Schmidt inner product. The decomposition is as follows:

$$\rho = \frac{1}{2}(\mathbb{1}_2 + n_x\sigma_x + n_y\sigma_y + n_z\sigma_z) = \frac{1}{2}(\mathbb{1}_2 + \langle \boldsymbol{\sigma}|\mathbf{n}\rangle), \tag{1.18}$$

where $\mathbf{n} \in \mathbb{R}_3$, $0 \leq |\mathbf{n}| \leq 1$, $\boldsymbol{\sigma} = [\sigma_x, \sigma_y, \sigma_z]$, $\langle \boldsymbol{\sigma}|\mathbf{n}\rangle = \sigma_x n_x + \sigma_y n_y + \sigma_z n_z$. The **Bloch vector n** defines points in the 3–dimensional Bloch ball. Pure states are represented by the surface of the ball and mixed states by its interior.

More generally any 2×2 complex matrix $M \in \mathcal{M}_{2\times 2}(\mathbb{C})$ can be decomposed in a similar way (see: [29]):

$$M = \frac{1}{2}(\mathrm{Tr}(M\mathbb{1}_2)\mathbb{1}_2 + \mathrm{Tr}(M\sigma_x)\sigma_x + \mathrm{Tr}(M\sigma_y)\sigma_y + \mathrm{Tr}(M\sigma_z)\sigma_z) = \frac{1}{2}(\mathrm{Tr}(M)\mathbb{1}_2 + \langle \boldsymbol{\sigma}|\mathbf{n}\rangle), \tag{1.19}$$

where $|\mathbf{n}\rangle = [\mathrm{Tr}(M\sigma_x), \mathrm{Tr}(M\sigma_y), \mathrm{Tr}(M\sigma_z)]^{\mathrm{T}} \in \mathbb{C}_3$. In quantum mechanics we deal with Hermitian matrices representing states. As mentioned above, for a Hermitian matrix M the vector is real: $|\mathbf{n}\rangle \in \mathbb{R}_3$.

1.3.2 State evolution

Unitary evolution of a qubit's state can be thought of as rotating its Bloch vector. The rotation is represented by some 3×3 real matrix O with determinant 1. The group of such matrices is called the special orthogonal group of dimension 3 and denoted SO(3).

As mentioned before, a unitary evolution of a qubit state ρ is represented by an element U of the group SU(2): $\rho \to U\rho U^\dagger$. For any such element, an element O of SO(3) can be found which describes the same unitary evolution of the Bloch vector: $\mathbf{n} \to O\mathbf{n}$. We can parametrize $U \in \mathrm{SU}(2)$ in the following way:

$$U = \begin{bmatrix} c & \bar{s} \\ -s & \bar{c}, \end{bmatrix} \tag{1.20}$$

where the complex coefficients c, s satisfy $|c|^2 + |s|^2 = 1$. Then the corresponding matrix $O \in \mathrm{SO}(3)$ takes the form (see: [30]):

$$O = \begin{bmatrix} \mathrm{Re}(c^2 - \bar{s}^2) & \mathrm{Im}(c^2 + \bar{s}^2) & -2\mathrm{Re}(cs) \\ -\mathrm{Im}(c^2 - \bar{s}^2) & \mathrm{Re}(c^2 + \bar{s}^2) & 2\mathrm{Im}(cs) \\ 2\mathrm{Re}(cs) & 2\mathrm{Im}(cs) & |c|^2 - |s|^2 \end{bmatrix}. \tag{1.21}$$

1.4 Distances

This section describes distances between the quantum objects introduced above (probability vectors, POVMs, channels), which will be used in what follows. Considering distances between objects is particularly important in the NISQ regime, e.g. to quantify how close an implementation is to an ideal object.

The **Total Variation (TV) distance** between probability vectors $\mathbf{p}, \mathbf{q}$ is defined as:

$$\mathrm{D}_{\mathrm{TV}}(\mathbf{p}, \mathbf{q}) = \frac{1}{2}\sum_{k=1}^{m} |p_k - q_k|. \tag{1.22}$$

Operationally this distance defines the optimal distinguishability between probability distributions. For an optimal strategy, it can be determined whether an experimental sample came from distribution $\mathbf{p}$ or $\mathbf{q}$ with probability of success $p = \frac{1}{2}(1 + \mathrm{D}_{\mathrm{TV}}(\mathbf{p}, \mathbf{q}))$, assuming equal prior distribution.

The TV distance gives rise to the **operational distance** between measurements M, N:

$$\mathrm{D}_{\mathrm{op}}(\mathsf{M}, \mathsf{N}) = \max_{\rho} \mathrm{D}_{\mathrm{TV}}(\mathbf{p}_{\mathsf{M}}(\rho), \mathbf{p}_{\mathsf{N}}(\rho)), \tag{1.23}$$

where $\mathbf{p}_{\mathsf{M}}(\rho), \mathbf{p}_{\mathsf{N}}(\rho)$ are probability vectors generated by measuring state ρ with M and N respectively. Maximization is over all states ρ, meaning that the operational distance quantifies the **worst–case** distance between distributions generated by the two compared POVMs.

An alternative way to measure the distance between two POVMs was recently given in [31]. Instead of the worst–case, an **average–case distance** between POVMs M and N was proposed:

$$\mathrm{D}_{\mathrm{avg}}(\mathsf{M}, \mathsf{N}) = \frac{1}{2d} \sum_{k=1}^{m} \sqrt{\mathrm{Tr}[(M^{(k)} - N^{(k)})^2] + \mathrm{Tr}^2[M^{(k)} - N^{(k)}]}. \tag{1.24}$$

Intuitively, this distance can be thought of as the average TV distance between the probability vectors $\mathbf{p}_{\mathsf{M},\psi}, \mathbf{p}_{\mathsf{N},\psi}$ generated by measurements M and N and quantum states ψ, averaged over pure quantum states sampled from a uniform distribution induced by the Haar measure on $\mathrm{U}(d)$.

To compare two states ρ, σ, **state fidelity** is defined:

$$\mathrm{F}_{\mathrm{state}}(\rho, \sigma) = (\mathrm{Tr}(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}))^2. \tag{1.25}$$

This distance is used to define **channel fidelity** between Λ and $\mathcal{E}$, given state ρ, as (see: [32]):

$$\mathcal{F}(\Lambda(\rho), \mathcal{E}(\rho)) = \mathrm{F}_{\mathrm{state}}(\Lambda(\rho), \mathcal{E}(\rho)). \tag{1.26}$$

Recall from (1.13) that the action of a unitary transformation can be described by a unitary channel and thus this distance can be used to compare an ideal transformation to its experimental implementation. In current devices the gate fidelities are often established experimentally, through randomized benchmarking (RB) [33] or cross entropy benchmarking (XEB) [18]. Specifically, the error ϵ is estimated and the fidelity is given as $\mathcal{F} = 1 - \epsilon$.

1.5 Quantum Detector Tomography

In real devices the implementation of measurements deviates from their ideal model. It is desirable to be able to characterize the experimentally available detector. **Quantum Detector Tomography (QDT)** is a strategy designed for this purpose. Along with Quantum State Tomography and Quantum Process Tomography, it allows for full characterization of a given experiment [34]. The goal of QDT is to obtain the explicit classical description of a POVM's effects. The full description can be obtained by preparing a complete or overcomplete set of Hermitian operators on a given Hilbert space and then measuring each of them many times, to obtain sufficiently exact statistics. For an n–qubit 2^n–dimensional Hilbert space at least 2^{2n} basis states are needed. From the experimental statistics the POVM's effects can be reconstructed by various methods, including convex optimization [35], linear inversion [36], or Maximum Likelihood Estimation [14], [16]. Note that for this method accurate state preparation is required.

The number of required circuits in this method scales exponentially with the size of the system, which quickly becomes experimentally infeasible. Recently in [23] this problem was addressed. The authors proposed to reconstruct k–qubit POVMs in an n–qubit device, using $\mathcal{O}(k2^k \log n)$ circuits. The reconstruction was restricted to only diagonal terms of a POVMs effects. The method was called **Diagonal Detector Overlapping Tomography (DDOT)**, in reference to the Overlapping Tomography of states [37], which inspired it.

A further development of this work is in progress, awaiting publication [21]. The **Detector Overlapping Tomography (DOT)** is a generalization of the method described above, which allows to fully characterize

k–qubit POVMs, including the off-diagonal terms of effects. Similarly to DDOT the number of circuits required is exponential in k and logarithmic in n. This method was used to obtain the experimental data which will later be analyzed in this work from the most advanced quantum devices currently made available. It is outlined in more detail in Section 3.1.1.

1.6 Measurement errors

Two types of errors which have been observed in QDT results are described below, accompanied by a brief review of methods of their mitigation.

1.6.1 Types of error

Classical error

To understand classical error in measurements, we will start with a simple example. Consider a device which is designed to perform a 1–qubit computational basis measurement. We denote this measurement by P. Its effects have the following form:

$$P^{(0)} = |0\rangle\langle 0| = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \; P^{(1)} = |1\rangle\langle 1| = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \tag{1.27}$$

Assume now that the POVM M was reconstructed in QDT of this device and its effects have the form:

$$M^{(0)} = \begin{bmatrix} 1-p & 0 \\ 0 & q \end{bmatrix}, \; M^{(1)} = \begin{bmatrix} p & 0 \\ 0 & 1-q \end{bmatrix}. \tag{1.28}$$

Note that this means that with probability p the measurement result 0 was obtained when it should have been 1 and likewise with probability q it was 1 instead of 0. This is analogous to classical bit flip. Thus such error is called **classical error**. If only classical error is present in a POVM, its effects are diagonal matrices in computational basis.

Classical error can be fully characterized by a **noise matrix** Λ, whose elements are:

$$\Lambda_{kl} = p(k|l) = \mathrm{Tr}(|l\rangle\langle l| M^{(k)}) = M^{(k)}_{ll}. \tag{1.29}$$

Λ is stochastic, that is: $\forall_k \sum_l \Lambda_{kl} = 1$. For the 1–qubit example above Λ would take the form:

$$\Lambda = \begin{bmatrix} 1-p & q \\ p & 1-q \end{bmatrix}. \tag{1.30}$$

All of the POVM's effects can be constructed from Λ:

$$\forall_k M^{(k)} = \sum_l \Lambda_{kl} P^{(l)}. \tag{1.31}$$

This relation is meant when we write that:

$$\mathsf{M} = \Lambda\mathsf{P}. \tag{1.32}$$

Coherent error

Coherent (or quantum) errors are also present in devices, whenever the POVM effects obtained from QDT have nonzero off–diagonal elements in computational basis. This can be written, using notation from Eq. (1.32) as:

$$\mathsf{M} = \Lambda\mathsf{P} + \Delta. \tag{1.33}$$

Here $\Delta = \{\Delta^{(k)}\}_{k=1}^m$ is the coherent part of noise: $\forall_k \Delta^{(k)} = M^{(k)} - \mathrm{diag}(M^{(k)})$.

1.6.2 Mitigation

Mitigation of classical error with the inverse of noise matrix

Classical error can be mitigated using just classical post–processing. Notice from (1.31) and linearity of Born's rule that the probability vector obtained in the experiment $\mathbf{p}^{\mathbf{exp}} = [\mathrm{Tr}(\rho M^{(0)}), ..., \mathrm{Tr}(\rho M^{(m)})]^{\mathrm{T}}$ is related to the probability vector which would be obtained in an ideal device $\mathbf{p}^{\mathrm{ideal}} = [\mathrm{Tr}(\rho P^{(0)}), ..., \mathrm{Tr}(\rho P^{(m)})]^{\mathrm{T}}$ by the following relationship:

$$\mathbf{p}^{\mathrm{exp}} = [\mathrm{Tr}(\rho \sum_l \Lambda_{0l} P^{(l)}), ..., \mathrm{Tr}(\rho \sum_l \Lambda_{ml} P^{(l)})]^{\mathrm{T}} = \Lambda \mathbf{p}^{\mathrm{ideal}}. \tag{1.34}$$

Thus to mitigate classical error one simply needs to multiply the experimental probability vector by the inverse of the noise matrix (assuming it is invertible) to obtain the ideal probability vector:

$$\mathbf{p}^{\mathrm{ideal}} = \Lambda^{-1} \mathbf{p}^{\mathrm{exp}}. \tag{1.35}$$

This can yield nonphysical results, however it's an issue which can be easily dealt with. For details see [14, 16], where this method was introduced. It is worth noting that it is currently implemented in the qiskit SDK on IBM's quantum devices [22].

The error of this mitigation method resulting from the presence of coherent noise can be quantified by considering the TV distance between the ideal probability vector and that obtained from the mitigation procedure. This distance can be upper–bounded (see: [23]):

$$\mathrm{D}_{\mathrm{TV}}(\Lambda^{-1}\mathbf{p}^{\mathrm{exp}}, \mathbf{p}^{\mathrm{ideal}}) \leq \|\Lambda^{-1}\|_{1-1} \cdot \mathrm{D}_{\mathrm{op}}(\mathsf{M}, \mathsf{P}). \tag{1.36}$$

Other mitigation methods

Because the cost in quantum and classical resources of this method grows exponentially with system size, various approaches to reducing this cost have been made. They are based on employing noise models in which the noise characterization and inversion can be conducted on smaller subsystems without significant loss of effectiveness [20, 21, 23]. Additionally in [24] a model is used based on Continuous Time Markov Process, inspired by PEC. Noise matrices are sampled from a suitable quasi-probability distribution. The method targets a specific probability distribution, which limits its universality.

A different, two-step, mitigation procedure based on device characterization, targeting both classical and coherent measurement errors was proposed in [25]. After device characterization in the first step the POVM effects obtained from QDT are decomposed into weighted sum of tensor product of the efect rotated by local unitaries and some hermitian global operator. Then local unitaries from that decomposition are placed before measurement. In second step, post–processing, the knowledge of the first step is leveraged to obtain a mitigated experimental probability distribution.

Another method of mitigation of general measurement error was proposed in [26]. Unlike most alternative methods, it doesn't rely on the knowledge of the noise model, it only assumes accurate ground state preparation. It focuses on error mitigation in the estimation of Pauli observables (which form the basis of any observable). In this method random Pauli bit flips are uniformly applied in a circuit prior to measurement. Using the knowledge of those Pauli transformations, in post–processing a bias–free mitigated Pauli expectation value can be retrieved.

Some methods of mitigation of just coherent errors have also been proposed. In [27] the authors propose mitigation of coherent errors by adding a gate to the circuit before measurement. They propose 3 methods: firstly, applying a Z gate half of the experiments and identity gate the other half; secondly "XY" twirling and thirdly Pauli twirling ("IXYZ"). In the two latter methods random circuits are generated from the respective gate sets. With appropriate classical post processing this method would yield POVMs with no coherent errors, assuming ideal gates and infinite statistics. The authors validate the method in a simulation.

1.7 Errors in NISQ devices

Below we gather the reported information on the order of magnitude of errors in some present–day quantum devices.

Unless stated otherwise, by measurement error we mean deviation in absolute value of matrix elements of POVM's effects from those in an ideal computational basis POVM. For classical errors this can be interpreted as the bit–flip probability. The gate errors are experimentally estimated through randomized or cross entropy benchmarking (as mentioned in Section 1.4).

IBM devices

In [14] the results of QDT in the 5–qubit devices IBM Q 5 Tenerife (ibmqx4) and IBM Q 5 Yorktown (ibmqx2) are reported for single qubit POVMs, where measurement was done on one qubit or in parallel with others. In both cases the typical classical measurement error was of the order of 1%, reaching 6% and the coherent error was of the order of 1%, reaching 3%.

In[20] 4– and 8–qubit subsets of 16–qubit ibmq_melbourne device were characterized. In the 4–qubit subset: average classical error over all qubits was 5%, reaching 9% in one qubit, in the 8–qubit subset: average 9%, up to 30%. No information on coherent measurement errors was provided.

In [15] the results of characterization of single–qubit POVMs of 6 out of the 27 qubits of ibmq_montreal were provided. The classical errorwas on average around 2,5%, reaching up to 6%, the coherent error was on average around 0.1%, reaching up to 3%.

The IBM Quantum website [38] reports for each device the average bit–flip probability on each qubit. No information on coherent error is provided. For the state–of–the–art ibm_washington 127–qubit device the bit–flip probability is on average 3.5% but for a few qubits reaches over 40%. The single–qubit gate error is on average around 0.2% but reaches 15% for one qubit. The two–qubit gates have on average the error of 6.5%.

Rigetti devices

In [16] 16-qubit device Aspen-4-16Q-A was characterized. Classical error was on average around 7%, up to 18%. Coherent error – around 0,3%, up to 1%. However the authors point our that since the average single–gate error in this device was around 5%, the assumption of ideal gate needed for QDT is higly violated.

The Rigetti website [**rigetti**] reports gate error of the state–of–the–art Rigetti Aspen-M-1 device. The median error on single–qubit gates is 0.2%, of two–qubit gates: 6.3% for CZ, 5.4% for XY.

Google devices

In [18], which is well–known for its claim of achieving quantum supremacy, the characterization data of the Sycamore device is provided. The classical readout errors are on average around 3.5%, reaching up to 10%. Single gate errors are on average around 0.16%, reaching up to 0.8%. Two–qubit gate errors are a few times higher but stay below 1%. In [19] very similar results are reported.

Summary

Readout noise dominates over single–qubit gate noise in present day superconducting–qubit devices. Classical noise is the dominant type of readout noise in present devices. While for some qubits it was reported to reach 40%, such values are rare outliers. In experiments one can choose to not use the most noisy qubits (as is done e.g. in [21]), so we consider the average values of errors more significant than those outliers.

Chapter 2

Theory of mitigation of measurement errors with the use of unitary transformations

This chapter describes our contribution. First we propose a method of mitigating the noise in a POVM by adding a unitary transformation into a circuit before measurement. We then set up the choice of an appropriate transformation as an optimization problem. We propose operationally meaningful cost functions for this problem:

- the worst– and average–case distances of a POVM from its dephased version,
- the worst– and average–case distances of a POVM from the computational basis POVM,
- the upper–bounds on the worst– and average–case TV distances between a probability vector obtained from the mitigation of classical errors using the method described in 1.6.2 and the probability vector which would be obtained in an ideal experiment.

We derive the upper–bound on the average–case TV distance between an ideal probability vector and that obtained from mitigation of classical errors, in analogy to the known worst–case upper–bound (1.36). Later we propose two heuristically motivated cost functions:

- one which quantifies a POVM's distance from its dephased version,
- one which quantifies a POVM's distance from the computational basis POVM.

We describe a known analytical method of Approximate Simultaneous Diagonalization of a set of matrices. We provide an alternative proof for this method's validity by translating it into the quantum–inspired Bloch ball formalism. Using similar techniques we provide methods of analytical optimization of both of the heuristically motivated cost functions. Finally we briefly discuss the physical assumptions made in this chapter.

2.1 Using unitaries before measurement to modify a POVM

Using QDT, one can obtain the characterization of a quantum measurement available experimentally in a given device. With that knowledge, such a measurement can be effectively modified to mitigate errors.
If a unitary $U^\dagger$ is added to a circuit before measurement, an input state ρ evolves into $U^\dagger \rho U$. According to Born's rule (1.16), the probability of obtaining outcome k upon performing measurement M on this state is:

$$p(k|U^\dagger \rho U) = \mathrm{Tr}(U^\dagger \rho U M^{(k)} = \mathrm{Tr}(\rho U M^{(k)} U^\dagger. \tag{2.1}$$

Thus it can be thought of as measuring the same state ρ, but with a different POVM, whose effects $UM^{(k)}U^\dagger$ were *rotated* by a unitary operator U. In the following work an operator $M^{(k)}$ rotated by a unitary U will mean $UM^{(k)}U^\dagger$. A POVM M rotated by U will mean a POVM whose effects have been rotated by U and will be denoted $U\mathsf{M}U^\dagger$. Note that in this convention adding a unitary U before measurement means rotating its effects by $U^\dagger$. It follows that in order to mitigate measurement errors, after characterizing the device, an appropriate unitary can be added to a circuit directly before measurement. The mitigation strategy is schematically shown in Figure 2.1.

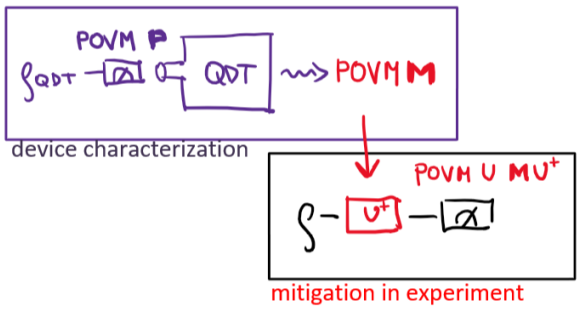


Figure 2.1: The proposed method of mitigation of measurement errors. The form of the experimentally available POVM M is obtained from QDT. With this knowledge, in further experiments a unitary $U^\dagger$ is inserted before measurement, such that M effectively becomes $U\mathsf{M}U^\dagger$.

2.2 Optimization problem

The task of finding an appropriate unitary to mitigate measurement error can be set up as an optimization problem. For a given POVM M on an n–qubit system, we will aim to optimize a chosen cost function f over some subset S of 2^n – dimensional unitary matrices $\mathrm{SU}(2^n)$:

$$\begin{aligned} &\underset{U}{\text{optimize}} && f(U\mathsf{M}U^\dagger) \\ &\text{s.t.} && U \in \mathrm{S} \subseteq \mathrm{SU}(2^n). \end{aligned} \tag{2.2}$$

2.2.1 Global or local unitary gates

In principle the unitary matrix U can be chosen from the whole set of unitary matrices on the n – qubit system, $\mathrm{SU}(2^n)$. Operationally it would mean implementing a global unitary on all of the n qubits in the system.
A different approach would be to look for U composed of local 1–qubit gates, $U \in \mathrm{SU}(2^n) : U = U_1 \otimes ... \otimes U_n$.
An intermediate approach could consider choosing unitaries local to some subset of the n–qubit system.
As mentioned in 1.7, in NISQ devices the fidelity of local 1–qubit gates is significantly higher than for multiple–qubit ones. While adding a global unitary before a multi–qubit POVM could theoretically reduce

the cost function more effectively than a just a local unitary or set of local gates, in practice the noise introduced by such a global unitary could be greater than the benefit of the mitigation procedure. For this reason we will focus on local gates.

2.3 Choice of cost functions

In this section we consider two main strategies of mitigating measurement errors - approaching a diagonal POVM or computational basis POVM. We define cost functions motivated by those strategies.

2.3.1 Approaching a diagonal POVM

As described in Section 1.6.2, there exist effective methods of mitigating classical measurement errors. In particular, a method often in use is the post–processing of experimental data with the use of the inverse of the noise matrix Λ. However, the presence of any coherent noise increases the error in this method. Thus it is desirable to modify a POVM M into $U\mathsf{M}U^\dagger$ so that its effects are *as close as possible* to diagonal matrices. Then the mitigation of classical errors through post-processing can be more effective. This joint mitigation strategy is schematically presented in Figure 2.2.

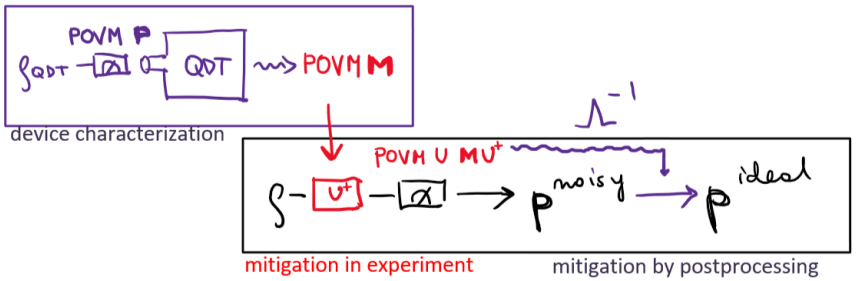


Figure 2.2: Strategy of combined method of mitigation of measurement errors. After device characterization coherent errors are mitigated with the use of unitary before measurement. For the resulting POVM, the method 1.6.2 of mitigating classical errors through post–processing is used.

We can use the operationally meaningful worst– and average–case distances introduced in Section 1.4 between the considered POVM and its dephased (diagonal) version to define two cost functions to be minimized:

$$\mathrm{D}_{\mathrm{op}}^{\mathrm{dep}}(U\mathsf{M}U^\dagger) \coloneqq \mathrm{D}_{\mathrm{op}}(U\mathsf{M}U^\dagger, \mathrm{diag}(U\mathsf{M}U^\dagger)), \tag{2.3}$$

$$\mathrm{D}_{\mathrm{avg}}^{\mathrm{dep}}(U\mathsf{M}U^\dagger) \coloneqq \mathrm{D}_{\mathrm{avg}}(U\mathsf{M}U^\dagger, \mathrm{diag}(U\mathsf{M}U^\dagger)). \tag{2.4}$$

The next proposed cost function quantifies how big the off–diagonal terms of the rotated POVM's effects are:

$$\mathrm{Off}(U\mathsf{M}U^\dagger) \coloneqq \sum_{k=1}^{m} \|UM^{(k)}U^\dagger - \mathrm{diag}(UM^{(k)}U^\dagger)\|_{\mathrm{HS}}^2. \tag{2.5}$$

The heuristic reasoning for choosing this cost function is that its minimization should intuitively reduce the norms of the off–diagonal terms of the POVM's effects. In Section 2.4.3 we provide a method for minimization of this function over the set of unitaries local to one qubit, taking advantage of the properties of the Hilbert-Schmidt norm.

The next two cost functions we propose stem directly from the method of mitigation of classical errors through post–processing described in Section 1.6.2. Recall from (1.36) that the upper–bound on the TV

distance between the probability vector obtained from this mitigation method and the ideal probability vector is given by:

$$\mathrm{upb}_{\mathrm{TV}}(U\mathsf{M}U^\dagger) \coloneqq \|\Lambda^{-1}\|_{1-1} \cdot \mathrm{D}_{\mathrm{op}}(U\mathsf{M}U^\dagger, \mathsf{P}). \tag{2.6}$$

An analogous upper–bound can be derived (see Appendix A.1) for the average–case error of this mitigation method. Intuitively speaking, this is an upper–bound on the average TV distance between the probability vector obtained from this mitigation method and the ideal probability vector, averaged over pure states:

$$\mathrm{upb}_{\mathrm{avg}}(U\mathsf{M}U^\dagger) \coloneqq \frac{1}{2\sqrt{m(m+1)}} \sum_{j=1}^{m} \sqrt{\mathrm{Tr}(\theta^{(j)2}) + \mathrm{Tr}^2(\theta^{(j)})}, \tag{2.7}$$

where $\theta^{(j)} = \sum_{k=1}^{m} \left[\Lambda^{-1}_{jk} \cdot [UM^{(k)}U^\dagger - \mathrm{diag}(UM^{(k)}U^\dagger)]\right]$ and m is the number of outcomes of M, here equal to the number of dimensions of the system. In both of the equations above Λ denotes the stochastic matrix defined in (1.30).

2.3.2 Approaching the computational basis POVM

An alternative mitigation strategy can be adopted. Instead of approaching a diagonal POVM, we can try to rotate a POVM M in such a way that it approaches the ideal, computational basis POVM P. We propose the following worst– and average–case distances to be minimized:

$$\mathrm{D}^{\mathrm{comp}}_{\mathrm{op}}(U\mathsf{M}U^\dagger) \coloneqq \mathrm{D}_{\mathrm{op}}(U\mathsf{M}U^\dagger, \mathsf{P}), \tag{2.8}$$

$$\mathrm{D}^{\mathrm{comp}}_{\mathrm{avg}}(U\mathsf{M}U^\dagger) \coloneqq \mathrm{D}_{\mathrm{avg}}(U\mathsf{M}U^\dagger, \mathsf{P}). \tag{2.9}$$

The next cost function is motivated heuristically. We propose to maximize the sum of overlaps of a POVM's M effects with the POVM's P effects:

$$\mathrm{Ovl}^{\mathrm{comp}}(U\mathsf{M}U^\dagger) \coloneqq \sum_{k=1}^{m} \langle \mathbf{k}|UM^{(k)}U^\dagger|\mathbf{k}\rangle. \tag{2.10}$$

Intuitively it should quantify how close a POVM M is to computational basis. In Section 2.5 we provide a method for minimization of this function.

2.3.3 Cost functions summary

The Table 2.1 summarizes the proposed cost functions for our mitigation method is shown.

interpretation	worst–case	average–case	heuristic
measure of coherent error	$\mathrm{D}^{\mathrm{dep}}_{\mathrm{op}}(U\mathsf{M}U^\dagger)$ (2.3)	$\mathrm{D}^{\mathrm{dep}}_{\mathrm{avg}}(U\mathsf{M}U^\dagger)$ (2.4)	$\mathrm{Off}^{\mathrm{diag}}(U\mathsf{M}U^\dagger)$ (2.5)
distance from computational basis POVM	$\mathrm{D}^{\mathrm{comp}}_{\mathrm{op}}(U\mathsf{M}U^\dagger)$ (2.8)	$\mathrm{D}^{\mathrm{comp}}_{\mathrm{avg}}(U\mathsf{M}U^\dagger)$ (2.9)	$\mathrm{Ovl}^{\mathrm{comp}}(U\mathsf{M}U^\dagger)$ (2.10)
upper–bound on error in classical noise mitigation	$\mathrm{upb}_{\mathrm{TV}}(U\mathsf{M}U^\dagger)$ (2.6)	$\mathrm{upb}_{\mathrm{avg}}(U\mathsf{M}U^\dagger)$ (2.7)	

Table 2.1: Summary of cost functions in the optimization problem of finding the appropriate unitary for mitigating measurement error. All but one function are to be minimized. $\mathrm{Ovl}^{\mathrm{comp}}(U\mathsf{M}U^\dagger)$ is to be maximized.

2.4 Minimizing the function $\mathrm{Off}^{\mathrm{diag}}(U\mathsf{M}U^\dagger)$

Here we discuss the methods of minimizing $\mathrm{Off}^{\mathrm{diag}}(U\mathsf{M}U^\dagger) = \sum_{k=1}^m \|UM^{(k)}U^\dagger - \mathrm{diag}(UM^{(k)}U^\dagger)\|_{\mathrm{HS}}^2$ (2.5) over a subset of unitary matrices. We briefly review some known methods. Later we provide an alternative proof for a known analytical method by translating it into the quantum mechanical formalism. Finally we use a similar technique to propose an analytical method of minimizing the heuristic cost function over a different subset of unitaries.

2.4.1 Existing methods

Simultaneous diagonalization

A set of matrices $\mathsf{A} = \{A^{(k)}\}_{k=1}^m$ is **simultaneously diagonizable** if there exists an invertible matrix P such that $\forall_k : PAP^{-1}$ is a diagonal matrix. We then say that P diagonalizes this set. A set of matrices is simultanously diagonizable if and only if all those matrices commute pairwise. Note that P is not unique, since also ΠDP diagonalizes the set of matrices, where by Π we denote a permutation matrix and by D – a diagonal invertible matrix.

If a POVM's effects were simultaneously diagonizable, minimizing the cost function $\mathrm{Off}^{\mathrm{diag}}(U\mathsf{M}U^\dagger)$ could be achieved by finding U which rotates its effects into a basis in which they're diagonal: $\mathrm{Off}^{\mathrm{diag}}(U\mathsf{M}U^\dagger) = 0$. For 2–effect POVMs it is always possible to simultaneously diagonalize both of their effects because they commute. However, for POVMs with more effects, the notion of **Approximate Simultaneous Diagonalization (ASD)** (or Approxiamte Joint Diagonalization) has to be introduced. ASD is a problem with applications in digital signal processing, among others in blind source separation (see e.g. references in [39]). ASD has also been used in physics to find an efficient basis for solving the vibrational Schrödinger equation [40].

Approximate Simultaneous Diagonalization

The problem of ASD of a set of non–commuting $d \times d$ matrices A has often been been set up (see e.g. [41–43]) as the problem of optimization over $P \in \mathrm{G}(d)$ ($\mathrm{G}(d)$ can denote $\mathrm{O}(d)$, $\mathrm{U}(d)$, $\mathrm{SO}(d)$, $\mathrm{SU}(d)$) of a cost function $f(P\mathsf{A}P^\dagger)$, quantifying *how close to diagonal* each of the matrices in the set is after rotation by P. In particular a very popular cost function to be minimized is $\mathrm{Off}^{\mathrm{diag}}(P\mathsf{A}P^\dagger)$ (see e.g. [39, 43–45]). Note that this becomes the same problem that we have set up in 2.2 with the cost function (2.5), if the matrices in A are chosen to be positive semi–definite and $\mathrm{G}(d) = \mathrm{SU}(d)$.

A popular approach to this optimization problem is the **Jacobi method**. The cost function is iteratively optimized over sets of plane rotations $G_{ij} \in \mathrm{SU}(d)$ (so–called Givens rotations). Here a Givens rotation is defined as an identity $d \times d$ matrix except for four entries of a unitary 2×2 matrix $R \in \mathrm{SU}(2)$ at positions determined by indices i, j:

$$G_{ij} \coloneqq \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & \mathbf{0} & \\ & & g_{ii} & \cdots & g_{ij} & & \\ & & \vdots & \ddots & \vdots & & \\ & & g_{ji} & \cdots & g_{jj} & & \\ & \mathbf{0} & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & \mathbf{0} & \\ & & r_{11} & \cdots & r_{12} & & \\ & & \vdots & \ddots & \vdots & & \\ & & r_{21} & \cdots & r_{22} & & \\ & \mathbf{0} & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}. \tag{2.11}$$

In each iteration (labeled by $l \in \{1, ..., N\}$) of the method the indices i_l, j_l are chosen according to some **pivot strategy** (see below). Then the $G_{i_l j_l}$ optimizing $f(G_{i_l j_l} \mathsf{A}_{l-1} G_{i_l j_l}^\dagger)$ is found and each of the the matrices

$A_{l-1}^{(k)}$ in A_{l-1} is rotated: $A_{l-1}^{(k)} \to A_l^{(k)} = G_{i_l j_l} A_{l-1}^{(k)} G_{i_l j_l}^\dagger$, where A_l denotes the set A after l–th iteration and $\mathsf{A}_0 = \mathsf{A}$. In the next iteration the optimization is carried out for a different choice of i, j. The final matrix in the optimization procedure is $P = G_{i_N j_N} \dots G_{i_1 j_1}$.

Multiple pivot strategies for the Jacobi method have been proposed. One example is the cyclic strategy, e.g. **cyclic–by–row**. In this strategy index pairs (i, j) chosen for subsequent iterations form the sequence: $(1,2), (1,3), \dots, (1,m), (2,3), \dots, (2,d), \dots (d-1,d), (1,2), (1,3) \dots$

For fixed i, j the explicit expression for $G_{ij} \in \mathrm{SU}(d)$ minimizing the function $\mathrm{Off}^{\mathrm{diag}}(G_{ij} \mathsf{A} G_{ij}^\dagger)$ was given in [46]. The authors' method applies to a set A of any complex matrices.

Convergence of Jacobi methods

The convergence of Jacobi methods with different pivot strategies (including cyclic–by–row) has up to recently been proven only for the case when A consists of a single matrix or a pair of commuting matrices. See [47] and [48] for a review and and the most current results for real symmetric and complex Hermitian matrices.

The problem of ASD is a special case of a more general problem of Approximate Tensor Diagonalization. To understand how ASD fits into this wider context see [49]. The authors of that work have recently proven the convergence of the Jacobi method with a Riemannian gradient–based pivot strategy. Their results are summarized in [50]. The results hold for ASD of a set of Hermitian matrices as a special case. The convergence of the Jacobi method with the cyclic pivot strategy is still unknown.

It is worth noting that any $d \times d$ unitary matrix can be decomposed into a tensor product of local 1-qubit unitaries (see e.g. the Reck decomposition [51], where the decomposition is into a product of transformations acting not–trivially on only two optical modes).

2.4.2 Alternative proof for the known analytical method

The authors of [46] provide a proof of their method, which applies to any set of square complex matrices. We restate the problem in a quantum computation formalism and provide an alternative proof, which applies only to a set of Hermitian matrices but will later prove useful in extending their method to the case when the optimization is over another subset of unitary matrices.

Theorem 1. *Let A be the set of m Hermitian $d \times d$ matrices $\mathsf{A} = \{A^{(k)}\}_{k=1}^m$.*

Let $U_{ij} \in \mathrm{SU}(d)$ be a unitary $d \times d$ matrix which is the identity matrix, except for 4 entries determined by indices $i \neq j$. For fixed i, j those entries can be parametrized by two complex numbers $c, s \in \mathbb{C} : |c|^2 + |s|^2 = 1$ as follows:

$$V_{ij} \coloneqq \begin{bmatrix} v_{ii} & v_{ij} \\ v_{ji} & v_{jj} \end{bmatrix} = \begin{bmatrix} c & \bar{s} \\ -s & \bar{c} \end{bmatrix}. \tag{2.12}$$

Then such a matrix U_{ij} which minimizes the function:

$$\mathrm{Off}(U_{ij} \mathsf{A} U_{ij}^\dagger) = \sum_{k=1}^m \| U_{ij} A^{(k)} U_{ij}^\dagger - \mathrm{diag}(U_{ij} A^{(k)} U_{ij}^\dagger) \|_{\mathrm{HS}}^2 \tag{2.13}$$

is given by the parametres c, s:

$$c = \frac{\sqrt{2(z+1)}}{2}, \;\; s = \frac{x - \mathrm{i}y}{\sqrt{2(z+1)}}, \tag{2.14}$$

where $[x, y, z]^{\mathrm{T}}$ is the eigenvector normalized to 1 in euclidian norm associated with the largest eigenvalue of the matrix $N(\mathsf{A})$, defined as follows:

$$N(\mathsf{A}) = \sum_{k=1}^{m} |\mathbf{n}^{(k)}\rangle\langle\mathbf{n}^{(k)}| \in \mathbb{R}_{3\times 3}, \text{ where } |\mathbf{n}^{(k)}\rangle \coloneqq [\mathrm{Re}(a_{ji}^{(k)}), \mathrm{Im}(a_{ji}^{(k)}), \frac{1}{2}(a_{ii}^{(k)} - a_{jj}^{(k)})]^{\mathrm{T}} \in \mathbb{R}^3. \tag{2.15}$$

Here $a_{ij}^{(k)}$ denotes the matrix element of $A^{(k)}$ in row i and column j.

The solution is not unique.

Proof. Because the diagonal and off–diagonal parts of any square matrix are orthogonal with respect to the Hilbert–Schmidt inner product, the following relation holds:

$$\begin{aligned}\sum_{k=1}^{m} \|U_{ij}A^{(k)}U_{ij}^{\dagger}\|_{\mathrm{HS}}^2 &= \sum_{k=1}^{m} \left(\|\mathrm{diag}(U_{ij}A^{(k)}U_{ij}^{\dagger})\|_{\mathrm{HS}}^2 + \|U_{ij}A^{(k)}U_{ij}^{\dagger} - \mathrm{diag}(U_{ij}A^{(k)}U_{ij}^{\dagger})\|_{\mathrm{HS}}^2 \right) \\ &= \sum_{k=1}^{m} \left(\|\mathrm{diag}(U_{ij}A^{(k)}U_{ij}^{\dagger})\|_{\mathrm{HS}}^2 \right) + \mathrm{Off}(U_{ij}\mathsf{A}U_{ij}^{\dagger}).\end{aligned} \tag{2.16}$$

The left side is constant for any U_{ij} because of unitary invariance of the H–S norm. Thus the mimimization of $\mathrm{Off}(U_{ij}\mathsf{A}U_{ij}^{\dagger})$ is equivalent to the maximization of:

$$\mathrm{D}(U_{ij}\mathsf{A}U_{ij}^{\dagger}) \coloneqq \sum_{k=1}^{m} \|\mathrm{diag}(U_{ij}A^{(k)}U_{ij}^{\dagger})\|_{\mathrm{HS}}^2. \tag{2.17}$$

Let us now decompose any of the $m \times m$ matrices $\mathsf{A}^{(k)}$ as follows:

$$A^{(k)} = \sum_{g,h\notin\{i,j\}} a_{gh}^{(k)}|g\rangle\langle h| + \sum_{\substack{g\notin\{i,j\},\\ h\in\{i,j\}}} a_{gh}^{(k)}|g\rangle\langle h| + \sum_{\substack{g\in\{i,j\},\\ h\notin\{i,j\}}} a_{gh}^{(k)}|g\rangle\langle h| + \sum_{g,h\in\{i,j\}} a_{gh}^{(k)}|g\rangle\langle h|, \tag{2.18}$$

where $g, h \in \{1, ..., d\}$ label n–dimensional computational–basis states, according to convention in (1.11).

Using the same notation, U_{ij} can be decomposed as follows:

$$U_{ij} = \left(\sum_{g\notin\{i,j\}} |g\rangle\langle g| \right) + u_{ii}|i\rangle\langle i| + u_{ij}|i\rangle\langle j| + u_{ji}|j\rangle\langle i| + u_{jj}|j\rangle\langle j|. \tag{2.19}$$

Note that only the i–th and j–th rows and columns of $M^{(k)}$ are affected by rotation by U_{ij}:

$$g, h \notin \{i, j\} \implies U_{ij}|g\rangle\langle h|U_{ij}^{\dagger} = |g\rangle\langle h|. \tag{2.20}$$

Furthermore, if $g \in \{i, j\}$ or $h \in \{i, j\}$, the matrix $|g\rangle\langle h|$ rotated by U_{ij} will only have non–zero terms on the diagonal if $g \in \{i, j\}$ and $h \in \{i, j\}$. Thus:

$$\mathrm{diag}(U_{ij}A^{(k)}U_{ij}^{\dagger}) = \sum_{g\notin\{i,j\}} a_{gg}^{(k)}|g\rangle\langle g| + \mathrm{diag}(U_{ij} \sum_{g,h\in\{i,j\}} a_{gh}^{(k)}|g\rangle\langle h|U_{ij}^{\dagger}). \tag{2.21}$$

Again from orthogonality with respect to the H–S inner product of the first and second terms of this sum we have:

$$\|\mathrm{diag}(U_{ij}A^{(k)}U_{ij}^{\dagger})\|_{\mathrm{HS}}^2 = \|\sum_{g\notin\{i,j\}} a_{gg}^{(k)}|g\rangle\langle g|\|_{\mathrm{HS}}^2 + \|\mathrm{diag}(U_{ij}\sum_{g,h\in\{i,j\}} a_{gh}^{(k)}|g\rangle\langle h|U_{ij}^{\dagger})\|_{\mathrm{HS}}^2. \tag{2.22}$$

Since only the second term in the sum depends on U_{ij}, to maximize $\mathrm{D}(U_{ij}\mathsf{A}U_{ij}^{\dagger})$ (2.17), one has to maximize:

$$\sum_{k=1}^{m}\|\mathrm{diag}(U_{ij}\sum_{g,h\in\{i,j\}} a_{gh}^{(k)}|g\rangle\langle h|U_{ij}^{\dagger})\|_{\mathrm{HS}}^2. \tag{2.23}$$

The matrix $U_{ij}\sum_{g,h\in\{i,j\}} a_{gh}^{(k)}|g\rangle\langle h|U_{ij}^{\dagger}$ has non–zero terms only on a subspace of 2×2 matrices and thus the problem can be translated into the Bloch ball formalism.

Let us denote:

$$A^{(k,i,j)} = \begin{bmatrix} a_{ii}^{(k)} & a_{ij}^{(k)} \\ a_{ji}^{(k)} & a_{jj}^{(k)} \end{bmatrix}. \tag{2.24}$$

Then the maximization of (2.23) is equivalent to the maximization of:

$$\mathrm{F}(V_{ij}\mathsf{A}V_{ij}^{\dagger}) \coloneqq \sum_{k=1}^{m}\|\mathrm{diag}(V_{ij}A^{(k,i,j)}V_{ij}^{\dagger})\|_{\mathrm{HS}}^2, \tag{2.25}$$

where V_{ij} is the projection of U_{ij} onto its 2×2 nontrivial subspace, defined in (2.12).

From (1.19) we have:

$$A^{(k,i,j)} = \frac{1}{2}\mathrm{Tr}(M^{(k)})\cdot\mathbb{1}_2 + \langle\boldsymbol{\sigma}|\mathbf{n}^{(k)}\rangle,\ |\mathbf{n}^{(k)}\rangle \coloneqq [\mathrm{Re}(a_{ji}^{(k)}),\mathrm{Im}(a_{ji}^{(k)}),\frac{1}{2}(a_{ii}^{(k)}-a_{jj}^{(k)})]^{\mathrm{T}}\in\mathbb{R}^3. \tag{2.26}$$

Recall that $\langle\boldsymbol{\sigma}|\mathbf{n}\rangle = \sigma_x n_x+\sigma_y n_y+\sigma_z n_z$. Then, as in (1.21), the rotation of matrix $A^{(k,i,j)}$ by V_{ij} is equivalent to the rotation of the vector $|\mathbf{n}^{(k)}\rangle$ by a matrix $O_{V_{ij}}\in\mathrm{SO}(3)$:

$$O_{V_{ij}} = \begin{bmatrix} \mathrm{Re}(c^2-\bar{s}^2) & \mathrm{Im}(c^2+\bar{s}^2) & -2\mathrm{Re}(cs) \\ -\mathrm{Im}(c^2-\bar{s}^2) & \mathrm{Re}(c^2+\bar{s}^2) & 2\mathrm{Im}(cs) \\ 2\mathrm{Re}(cs) & 2\mathrm{Im}(cs) & |c|^2-|s|^2 \end{bmatrix}. \tag{2.27}$$

So we get:

$$\begin{aligned}\mathrm{F}(V_{ij}\mathsf{A}V_{ij}^{\dagger}) &= \sum_{k=1}^{m}\|\mathrm{diag}(\frac{1}{2}\mathrm{Tr}(A^{(k)})\cdot\mathbb{1}_2 + \langle\boldsymbol{\sigma}|O_{V_{ij}}|\mathbf{n}^{(k)}\rangle)\|_{\mathrm{HS}}^2 \\ &= \sum_{k=1}^{m}\|\frac{1}{2}\mathrm{Tr}(A^{(k)})\cdot\mathbb{1}_2 + (O_{V_{ij}}|\mathbf{n}^{(k)}\rangle)_z\cdot\sigma_z\|_{\mathrm{HS}}^2 \\ &= \sum_{k=1}^{m}\|\frac{1}{2}\mathrm{Tr}(A^{(k)})\cdot\mathbb{1}_2 + \langle\mathbf{v}_O|\mathbf{n}^{(k)}\rangle\cdot\sigma_z\|_{\mathrm{HS}}^2\end{aligned} \tag{2.28}$$

where we defined:

$$\langle\mathbf{v}_O| \coloneqq \langle\mathbf{e}_z|O_{V_{ij}},\ \ \langle\mathbf{e}_z| \coloneqq [0,0,1]. \tag{2.29}$$

Because $\mathbb{1}_2$ and σ_z are orthogonal with respect to the Hilbert–Schmidt inner product, we obtain:

$$\begin{aligned}\mathrm{F}(U_{ij}\mathsf{A}U_{ij}^{\dagger}) &= \sum_{k=1}^{m}\|\frac{1}{2}\mathrm{Tr}(A^{(k)})\cdot\mathbb{1}_2\|_{\mathrm{HS}}^2 + \|\langle\mathbf{v}_O|\mathbf{n}^{(k)}\rangle\cdot\sigma_z\|_{\mathrm{HS}}^2 \\ &= \sum_{k=1}^{m}\frac{1}{2}\mathrm{Tr}^2(A^{(k)}) + 2\langle\mathbf{v}_O|\mathbf{n}^{(\mathbf{k})}\rangle\langle\mathbf{n}^{(k)}|\mathbf{v}_O\rangle.\end{aligned} \tag{2.30}$$

Since only the second term depends on U, the maximization of the expression finally comes down to finding the normalized eigenvector $|\mathbf{v}_O\rangle_{\max}$ corresponding to the largest eigenvalue of the matrix:

$$N(\mathsf{A}) = \sum_{k=1}^{m} |\mathbf{n}^{(k)}\rangle\langle\mathbf{n}^{(k)}| \in \mathbb{R}_{3\times 3}. \tag{2.31}$$

From the definition of $|\mathbf{v}_O\rangle$ (2.29) and $O_{V_{ij}}$ (2.27) the dependence of $|\mathbf{v}_O\rangle$ on U is given by:

$$\begin{aligned} |\mathbf{v}_O\rangle = O_{V_{ij}}^{\mathrm{T}}|e_z\rangle &= [2\mathrm{Re}(cs),\, 2\mathrm{Im}(cs),\, |c|^2 - |s|^2]^{\mathrm{T}} \\ &= [cs + \overline{cs},\, \mathrm{i}(cs - \overline{cs}),\, |c|^2 - |s|^2]^{\mathrm{T}}. \end{aligned} \tag{2.32}$$

After solving the eigenproblem of $N(\mathsf{A})$ and finding the vector $|\mathbf{v}_O\rangle_{\max} = [x, y, z]^T$, we can use the relation above to find $c(x, y, z)$ and $s(x, y, z)$. It is easy to verify that the normalization $\sqrt{|x|^2 + |y|^2 + |z|^2} = 1$ implies the condition $|c|^2 + |s|^2 = 1$.

In order to find the explicit form of c, s it suffices to notice:

$$\begin{aligned} z + 1 &= 2|c|^2, \\ x - \mathrm{i}y &= 2cs, \\ \sqrt{2(z+1)} &= 2|c|. \end{aligned} \tag{2.33}$$

Then we obtain:

$$c = \frac{\sqrt{2(z+1)}}{2} e^{\mathrm{i}\phi},\ s = \frac{x - \mathrm{i}y}{\sqrt{2(z+1)}} e^{-\mathrm{i}\phi}, \text{ for some } \phi \in \mathbb{R}. \tag{2.34}$$

Since the solution is not unique, for simplicity we can take c real positive, as do the authors in [46].

□

Ambiguity of solution

The ambiguity appearing in the solution from Theorem 1 can be intuited from the fact mentioned in Section 2.4.1: if P simultaneously diagonalizes a matrix set, so does ΠDP, where by Π we denote a permutation matrix and by D – a diagonal invertible matrix. Here we are dealing with $V_{ij} \in \mathrm{SU}(2)$ so we require $\Pi \in \mathrm{SU}(2)$, $D = \begin{bmatrix} e^{\mathrm{i}\psi} & 0 \\ 0 & e^{-\mathrm{i}\psi} \end{bmatrix}$ for some $\psi \in \mathbb{R}$. The only nontrivial Π is σ_z.

Note that for such Π, D and any matrix $B \in \mathbb{C}_{2\times 2}$: $\|\mathrm{diag}(B)\|_{\mathrm{HS}} = \|\mathrm{diag}(\Pi B \Pi^\dagger)\|_{\mathrm{HS}} = \|\mathrm{diag}(DBD^\dagger)\|_{\mathrm{HS}}$.

In summary the solution from Theorem 1 is unique up to:

- multiplication of V_{ij} by $\begin{bmatrix} e^{\mathrm{i}\psi} & 0 \\ 0 & e^{-\mathrm{i}\psi} \end{bmatrix}$ for some $\psi \in \mathbb{R}$,
- multiplication of V_{ij} by σ_z.

2.4.3 A method of minimization over 1–qubit local unitaries

A similar method can be used to minimize the same cost function $\mathrm{Off}^{\mathrm{diag}}(U\mathsf{M}U^\dagger)$ but over the set $\mathrm{S}_{\mathrm{loc}}$ of 1–local unitary transformations of n– qubits:

$$\mathrm{S}_{\mathrm{loc}} = \{U \in \mathrm{SU}(2^n) :\, U = \bigotimes_{i=1}^{j} \mathbb{1}_2 \otimes V_j \bigotimes_{i=j+1}^{n} \mathbb{1}_2\}. \tag{2.35}$$

For simplicity of presentation we assume the unitary in question is local to the last qubit in an n-qubit circuit. The general case of a single–qubit rotation acting on any other qubit can be analyzed in a similar way. The the matrix representation in computational basis of such a unitary is shown in Figure 2.3 and U can be written as:

$$U = \bigotimes_{i=1}^{n-1} \mathbb{1}_2 \otimes V = \mathbb{1}_{2^{n-1}} \otimes V = \mathbb{1}_{2^{n-1}} \otimes \begin{bmatrix} c & \bar{s} \\ -s & \bar{c} \end{bmatrix}, \tag{2.36}$$

where c, s are complex parameters satisfying $|c|^2 + |s|^2 = 1$.

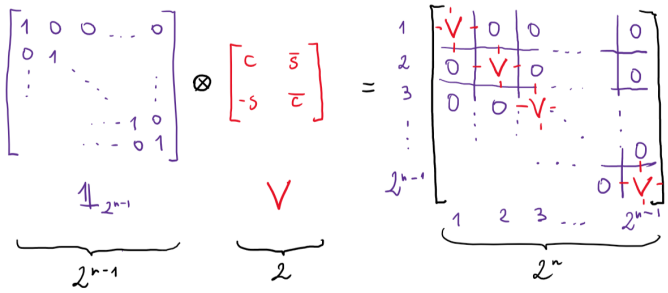


Figure 2.3: Schematic matrix representation of a unitary acting non-trivially on only last qubit

Theorem 2. *Let* M *be the set of* m *Hermitian* $2^n \times 2^n$ *matrices* $\mathsf{M} = \{M^{(k)}\}_{k=1}^m$.

Let $\mathrm{S}_{\mathrm{last}}$ *be the set:*

$$\mathrm{S}_{\mathrm{last}} = \{U \in \mathrm{SU}(2^n) : U = \mathbb{1}_{2^{n-1}} \otimes V = \mathbb{1}_{2^{n-1}} \otimes \begin{bmatrix} c & \bar{s} \\ -s & \bar{c} \end{bmatrix}\}, \tag{2.37}$$

where $c, s \in \mathbb{C} : |c|^2 + |s|^2 = 1$. *Then a matrix* $U \in \mathrm{S}_{\mathrm{last}}$ *which minimizes the function:*

$$\mathrm{Off}(U\mathsf{M}U^\dagger) = \sum_{k=1}^{m} \|UM^{(k)}U^\dagger - \mathrm{diag}(UM^{(k)}U^\dagger)\|_{\mathrm{HS}}^2 \tag{2.38}$$

is given by the parametres c, s*:*

$$c = \frac{\sqrt{2(z+1)}}{2}, \quad s = \frac{x - \mathrm{i}y}{\sqrt{2(z+1)}}, \tag{2.39}$$

where $[x, y, z]^{\mathrm{T}}$ *is the eigenvector normalized to 1 in euclidean norm associated with the largest eigenvalue of the matrix* $N(\mathsf{M})$*, defined as follows:*

$$N(\mathsf{M}) \coloneqq \sum_{k=1}^{m} \sum_{\mathbf{x} \in \{0,1\}^{n-1}} |\mathbf{n}^{(k,\mathbf{x})}\rangle\langle\mathbf{n}^{(k,\mathbf{x})}|, \tag{2.40}$$

with:

$$|\mathbf{n}^{(k,\mathbf{x})}\rangle \coloneqq [\mathrm{Re}(m_{21}^{(k,\mathbf{x},\mathbf{x})}), \mathrm{Im}(m_{21}^{(k,\mathbf{x},\mathbf{x})}), \frac{1}{2}(m_{11}^{(k,\mathbf{x},\mathbf{x})} - m_{22}^{(k,\mathbf{x},\mathbf{x})})]^{\mathrm{T}} \in \mathbb{R}^3, \tag{2.41}$$

and the elements of $|\mathbf{n}^{k,\mathbf{x}}\rangle$ *taken from the decomposition:*

$$M^{(k)} = \sum_{\substack{\mathbf{x},\mathbf{y} \in \{0,1\}^{n-1} \\ i,j \in \{0,1\}}} m_{ij}^{(k,\mathbf{x},\mathbf{y})} |\mathbf{x}\rangle\langle\mathbf{y}| \otimes |i\rangle\langle j|. \tag{2.42}$$

The solution is not unique.

Proof. When a POVM M is rotated by a local unitary $U = \mathbb{1}_{2^{n-1}} \otimes V$, its k-th effect takes form:

$$UM^{(k)}U^\dagger = \sum_{\substack{\mathbf{x},\mathbf{y}\in\{0,1\}^{n-1}\\ i,j\in\{0,1\}}} |\mathbf{x}\rangle\langle\mathbf{y}| \otimes V m_{ij}^{(k,\mathbf{x},\mathbf{y})} |i\rangle\langle j| V^\dagger = \sum_{\substack{\mathbf{x},\mathbf{y}\in\{0,1\}^{n-1}\\ i,j\in\{0,1\}}} |\mathbf{x}\rangle\langle\mathbf{y}| \otimes V m_{ij}^{(k,\mathbf{x},\mathbf{y})} |i\rangle\langle j| V^\dagger. \tag{2.43}$$

Let us define a block $M^{(k,\mathbf{x})}$ on the diagonal of the effect, corresponding to the last qubit:

$$M^{(k,\mathbf{x})} = \sum_{i,j\in\{0,1\}} m_{ij}^{(k,\mathbf{x},\mathbf{x})} |i\rangle\langle j| = \begin{bmatrix} m_{11}^{(k,\mathbf{x},\mathbf{x})} & m_{12}^{(k,\mathbf{x},\mathbf{x})} \\ m_{21}^{(k,\mathbf{x},\mathbf{x})} & m_{22}^{(k,\mathbf{x},\mathbf{x})} \end{bmatrix}. \tag{2.44}$$

Recall from Section 1.3 that in the Bloch ball picture:

$$VM^{(k,\mathbf{x})}V^\dagger = \frac{1}{2}\mathrm{Tr}(M^{(k,\mathbf{x})}) \cdot \mathbb{1}_2 + \langle \boldsymbol{\sigma} | O_V | \mathbf{n}^{(k,\mathbf{x})} \rangle. \tag{2.45}$$

where $O_V \in SO(3)$ is can be obtained from relation (1.21) and $|\mathbf{n}^{(k,\mathbf{x})}\rangle$ (2.41) is the Bloch vector of $M^{k,\mathbf{x}}$.

We can now return to the minimized function. As in (2.16), in order to minimize (2.38), we can equivalently maximize:

$$\mathrm{F}(U\mathsf{M}U^\dagger) \coloneqq \sum_{k=1}^{m} \|\mathrm{diag}(UM^{(k)}U^\dagger)\|_{\mathrm{HS}}^2. \tag{2.46}$$

The diagonal part of $UM^{(k)}U^\dagger$ is:

$$\begin{aligned} \mathrm{diag}(UM^{(k)}U^\dagger) &= \mathrm{diag}\left(\sum_{\mathbf{x}} |\mathbf{x}\rangle\langle\mathbf{x}| \otimes \left(\frac{1}{2}\mathrm{Tr}(M^{(k,\mathbf{x})}) \cdot \mathbb{1}_2 + \langle \boldsymbol{\sigma} | O_V | \mathbf{n}^{(k,\mathbf{x})} \rangle \right) \right) \\ &= \sum_{\mathbf{x}} |\mathbf{x}\rangle\langle\mathbf{x}| \otimes \left(\frac{1}{2}\mathrm{Tr}(M^{(k,\mathbf{x})}) \cdot \mathbb{1}_2 + \langle \mathbf{v}_O | \mathbf{n}^{(k,\mathbf{x})} \rangle \sigma_z \right). \end{aligned} \tag{2.47}$$

Here, as in (2.29), we defined:

$$\langle \mathbf{v}_O | \coloneqq \langle \mathbf{e}_z | O_V, \quad \langle \mathbf{e}_z | \coloneqq [0,0,1]. \tag{2.48}$$

Because $\mathbb{1}_2$ and σ_z are orthogonal with regards to the Hilbert–Schmidt inner product, we obtain:

$$\begin{aligned} \|\mathrm{diag}(UM^{(k)}U^\dagger)\|_{\mathrm{HS}}^2 &= \| \sum_{\mathbf{x}} |\mathbf{x}\rangle\langle\mathbf{x}| \otimes \frac{1}{2}\mathrm{Tr}(M^{(k,\mathbf{x})}) \cdot \mathbb{1}_2 \|_{\mathrm{HS}}^2 + \| \sum_{\mathbf{x}} |\mathbf{x}\rangle\langle\mathbf{x}| \otimes \langle \mathbf{v}_O | \mathbf{n}^{(k,\mathbf{x})} \rangle \sigma_z \|_{\mathrm{HS}}^2 \\ &= \sum_{\mathbf{x}} 2^{n-2} \mathrm{Tr}^2(M^{(k,\mathbf{x})}) + \sum_{\mathbf{x}} 2^n |\langle \mathbf{v}_O | \mathbf{n}^{(k,\mathbf{x})} \rangle|^2. \end{aligned} \tag{2.49}$$

Then the whole sum to be maximized (2.46) is given by:

$$\sum_{k=1}^{m} \|\mathrm{diag}(UM^{(k)}U^\dagger)\|_{\mathrm{HS}}^2 = 2^{n-2} \sum_{k=1}^{m} \sum_{\mathbf{x}} \mathrm{Tr}^2(M^{(k,\mathbf{x})}) + 2^n \sum_{\mathbf{x}} \langle \mathbf{n}^{(k,\mathbf{x})} | \mathbf{v}_O \rangle \langle \mathbf{v}_O | \mathbf{n}^{(k,\mathbf{x})} \rangle. \tag{2.50}$$

The first term doesn't depend on U, so only the second term has to be maximized. In order to maximize the expression, we have to find the ℓ_2 normalized eigenvector $|\mathbf{v}_O\rangle_{max} = [x,y,z]^{\mathrm{T}} \in \mathbb{R}^3$ corresponding to the largest eigenvalue of matrix $N(\mathsf{M})$, defined as:

$$N(\mathsf{M}) \coloneqq \sum_{k=1}^{m} \sum_{\mathbf{x}} |\mathbf{n}^{(k,\mathbf{x})}\rangle\langle\mathbf{n}^{(k,\mathbf{x})}|. \tag{2.51}$$

We can then find the dependence of c, s on x, y, z exactly as in (2.33) and obtain the same result (2.34).

□

Ambiguity

For reasons analogous to 2.4.2, the solution provided in Theorem 2 is unique up to:

- multiplication of V by $\begin{bmatrix} e^{i\psi} & 0 \\ 0 & e^{-i\psi} \end{bmatrix}$ for some $\psi \in \mathbb{R}$,
- multiplication of V by σ_z.

Example

It is worth noting that this method of optimization comes down to solving the eigenproblem for a 3×3 real matrix, for any n-qubit POVM M. The size of the system only comes up because the matrix $N(\mathsf{M})$ is obtained by summing 2^{n-1} 3×3 matrices, each obtained from $3 \cdot 2^{n-1}$ terms a POVM's effect from relation (2.41). This is illustrated in Figure 2.5 for a 3-qubit POVM.

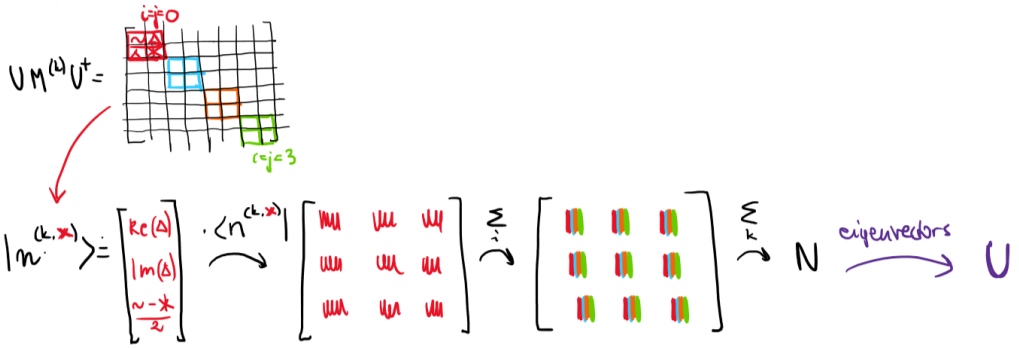


Figure 2.4: Schematic illustration of the optimization method of $\mathrm{Off}^{\mathrm{diag}}(U\mathsf{M}U^\dagger)$ on an example of a 3-qubit POVM.

2.5 Maximizing the function $\mathrm{Ovl}^{\mathrm{comp}}(U\mathsf{M}U^\dagger)$

In this section the motivation for the cost function $\mathrm{Ovl}^{\mathrm{comp}}(U\mathsf{M}U^\dagger)$ (2.10) and the explicit form of a local unitary which maximizes it will be given.

2.5.1 Motivation

Recall from Section 1.2.3 that the computational basis POVM's effects can be labelled by bit–strings corresponding to states onto which they project: $\mathsf{P} = \{P^{(\mathbf{x})}\}_{\mathbf{x}\in\{0,1\}^n}$. We will analogously label the effects of the POVM M which is a noisy implementation of P: $\mathsf{M} = \{M^{(\mathbf{x})}\}_{\mathbf{x}\in\{0,1\}^n}$. Then that the average–case distance between n–qubit POVMs M and P on a d–dimensional system ($d = 2^n$) is given by:

$$\mathrm{D}_{\mathrm{avg}}(\mathsf{M},\mathsf{P}) = \frac{1}{2d}\sum_{\mathbf{x}\in\{0,1\}^n}\sqrt{\mathrm{Tr}[(M^{(\mathbf{x})}-P^{(\mathbf{x})})^2]+\mathrm{Tr}^2[M^{(\mathbf{x})}-P^{(\mathbf{x})}]}. \tag{2.52}$$

If M gets rotated by P, its average-case distance from P is given by:

$$\begin{aligned}
\mathrm{D}_{\mathrm{avg}}(U\mathsf{M}U^\dagger,\mathsf{P}) &= \frac{1}{2d}\sum_{\mathbf{x}\in\{0,1\}^n}\sqrt{(\mathrm{Tr}[(UM^{(\mathbf{x})}U^\dagger-P^{(\mathbf{x})})^2]+\mathrm{Tr}^2(UM^{(\mathbf{x})}U^\dagger-P^{(\mathbf{x})})} \\
&= \frac{1}{2d}\sum_{\mathbf{x}\in\{0,1\}^n}\sqrt{\mathrm{Tr}[(UM^{(\mathbf{x})}U^\dagger-|\mathbf{x}\rangle\langle\mathbf{x}|)^2]+\mathrm{Tr}^2(UM^{(\mathbf{x})}U^\dagger-|\mathbf{x}\rangle\langle\mathbf{x}|)} \\
&= \frac{1}{2d}\sum_{\mathbf{x}\in\{0,1\}^n}\sqrt{\mathrm{Tr}((M^{(k)})^2)+\mathrm{Tr}^2(M^{(\mathbf{x})})-2\mathrm{Tr}(M^{(\mathbf{x})})+2-2\langle\mathbf{x}|UM^{(\mathbf{x})}U^\dagger|\mathbf{x}\rangle},
\end{aligned} \tag{2.53}$$

where the last step could be made thanks to the unitary invariance of trace.
The last terms under the square roots in the expression above, $\langle\mathbf{x}|UM^{(\mathbf{x})}U^\dagger|\mathbf{x}\rangle$, are the only ones which depend on U. Intuitively speaking, they are the overlaps of the $\mathbf{x}$–labelled effects of the considered POVMs. So the bigger the overlaps, the smaller the distance between the POVMs.

2.5.2 A method of maximization over 1–qubit local unitaries

We will now provide an explicit expression for a single–qubit unitary which maximizes the considered cost function. As in Section 2.4.3 we focus on unitaries acting on the last qubit. The general case of a single–qubit unitary acting on any other qubit can be analyzed in a similar way.

Theorem 3. *Let* M *be the set of* 2^n *Hermitian* $2^n\times 2^n$ *matrices* $\mathsf{M}=\{M^{(\mathbf{x})}\}_{\mathbf{x}\in\{0,1\}^n}$.

Let $\mathrm{S}_{\mathrm{last}}$ *be:*

$$\mathrm{S}_{\mathrm{last}}=\{U\in\mathrm{SU}(2^n):\ U=\mathbb{1}_{2^{n-1}}\otimes V=\mathbb{1}_{2^{n-1}}\otimes\begin{bmatrix}c & \bar{s}\\ -s & \bar{c}\end{bmatrix}\}, \tag{2.54}$$

where $c,s\in\mathbb{C}:|c|^2+|s|^2=1$. *Then a matrix* $U\in\mathrm{S}_{\mathrm{last}}$ *which maximizes the function:*

$$\mathrm{Ovl}^{\mathrm{comp}}(U\mathsf{M}U^\dagger)=\sum_{\mathbf{x}\in\{0,1\}^n}\langle\mathbf{x}|UM^{(\mathbf{x})}U^\dagger|\mathbf{x}\rangle \tag{2.55}$$

is given by the parameters c,s*:*

$$c=\sqrt{\frac{z+1}{2}},\ s=\frac{x-\mathrm{i}y}{\sqrt{2(z+1)}}, \tag{2.56}$$

where $|w\rangle=[x,y,z]^{\mathrm{T}}$ *is the vector:*

$$|w\rangle=\frac{|v_O\rangle}{||v_O\rangle|_2},\ \textit{with}\ |v_O\rangle=\sum_{\mathbf{y}\in\{0,1\}^{n-1}}|\mathbf{n}^{(\mathbf{y}.[0],\mathbf{y})}\rangle-|\mathbf{n}^{(\mathbf{y}.[1],\mathbf{y})}\rangle, \tag{2.57}$$

where $|\mathbf{n}^{(\mathbf{x},\mathbf{y})}\rangle$ *is defined in* (2.41) *and* $\mathbf{y}.[i]$ *denotes the bit–string with* i *appended at the end.*

The solution is not unique.

Proof. Note that for any $\mathbf{x}$ only the diagonal part of $UM^{(\mathbf{x})}U^\dagger$ will contribute to $\langle \mathbf{x}|UM^{(\mathbf{x})}U^\dagger|\mathbf{x}\rangle$. So using the expression for the diagonal part of $UM^{(\mathbf{x})}U^\dagger$ from (2.47), we can write:

$$\langle \mathbf{x}|UM^{(\mathbf{x})}U^\dagger|\mathbf{x}\rangle = \langle \mathbf{x}|\mathrm{diag}(UM^{(\mathbf{x})}U^\dagger)|\mathbf{x}\rangle = \langle \mathbf{x}| \sum_{\mathbf{y}\in\{0,1\}^{n-1}} |\mathbf{y}\rangle\langle\mathbf{y}| \otimes \left(\frac{1}{2}\mathrm{Tr}(M^{(\mathbf{x},\mathbf{y})})\cdot \mathbb{1}_2 + \langle \mathbf{v}_O|\mathbf{n}^{(\mathbf{x},\mathbf{y})}\rangle \sigma_z\right)|\mathbf{x}\rangle, \tag{2.58}$$

where $M^{(\mathbf{x},\mathbf{y})}$ is a 2×2 block on the diagonal of $M^{(\mathbf{x})}$ defined in (2.44) and $|\mathbf{n}^{\mathbf{x},\mathbf{y}}\rangle$ is its Bloch vector.

Since only the last terms in the sum depend on U, in the end the expression to be maximized is:

$$\begin{aligned}\sum_{\mathbf{x}\in\{0,1\}^n} \langle \mathbf{x}| \sum_{\mathbf{y}\in\{0,1\}^{n-1}} |\mathbf{y}\rangle\langle\mathbf{y}| \otimes \langle \mathbf{v}_O|\mathbf{n}^{(\mathbf{x},\mathbf{y})}\rangle\sigma_z|\mathbf{x}\rangle &= \sum_{\mathbf{x}\in\{0,1\}^n} \langle \mathbf{x}| \sum_{\mathbf{y}\in\{0,1\}^{n-1}} \langle \mathbf{v}_O|\mathbf{n}^{(\mathbf{x},\mathbf{y})}\rangle |\mathbf{y}\rangle\langle\mathbf{y}| \otimes \sigma_z|\mathbf{x}\rangle \\ &= \langle \mathbf{v}_O| \sum_{\mathbf{y}\in\{0,1\}^{n-1}} |\mathbf{n}^{(\mathbf{y}.[0],\mathbf{y})}\rangle - |\mathbf{n}^{(\mathbf{y}.[1],\mathbf{y})}\rangle.\end{aligned} \tag{2.59}$$

The last step could be made because:

$$\langle \mathbf{x}|(|\mathbf{y}\rangle\langle\mathbf{y}|\otimes\sigma_z)|\mathbf{x}\rangle = \begin{cases} 1 \text{ if } \mathbf{x} = \mathbf{y}.[0], \\ -1, \text{ if } \mathbf{x} = \mathbf{y}.[1], \\ 0 \text{ otherwise.}\end{cases} \tag{2.60}$$

The expression above is maximized if $|v_O\rangle$ is equal to the vector obtained in the sum on the right. Then from the relation between the matrix O and U the explicit form of the parameters c, s of the matrix U can be obtained, as it was done in (2.34). It can be seen there that the solution is only unique up to multiplication of the parameters by opposite phases. □

Example

An example of this optimization is illustrated in Figure 2.5 for a 2-qubit POVM.

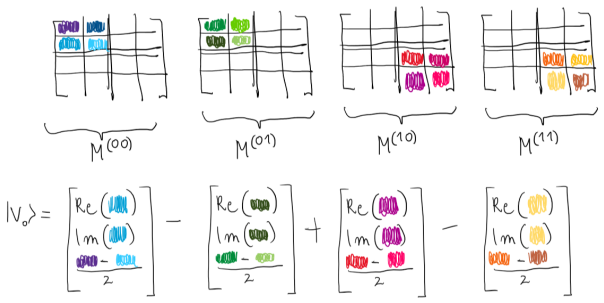


Figure 2.5: Schematic illustration of the optimization method of $\mathrm{Ovl}^{\mathrm{comp}}(U\mathsf{M}U^\dagger)$ on an example of a 2-qubit POVM.

2.6 Physical assumptions

In the sections above we implicitly assumed that we can represent POVMs by matrices exactly. However, in experiments we only have access to estimates of the true values of those objects, due to physical and statistical noise in detector tomography.

Because we only have access to a finite number of experimental results, the form of POVM obtained experimentally from tomography is burdened with statistical errors.

Also, in a physical implementation of a unitary gate, some error will always be present, which will affect the accuracy of QDT and the effect of the unitary which we propose to add to the circuit.

How big this accumulated noise is will determine whether it is feasible to use the method described above to mitigate coherent error with gates before measurement in a given experimental setup. Further discussion follows our tests of the method in Section 3.3.1.

Chapter 3

Tests of the mitigation method on experimental data

In this chapter the mitigation methods introduced before are tested on experimental data. The data is obtained from characterization of two devices: ibm_washington, which is the 127–qubit IBM Quantum Eagle r1 Processor and Rigetti's 80–qubit ASPEN–M–1.

3.1 Data description

3.1.1 Method of experimentally estimating matrix elements of POVMs

Diagonal Overlapping Tomography (DOT) was used to estimate matrix elements of 2–qubit reduced POVMs' effects. The choice of this procedure allowed for an efficient tomography of the largest devices currently available. The method employed to obtain the data we use is outlined below. For details see [21].

Let us first define the **reduced POVM** M_A on an l–qubit subset A of all n qubits in the device. Let $\bar{A}$ denote the subset of the $n-l$ remaining qubits. Let $\mathbf{x}_A, \mathbf{x}_{\bar{A}}$ denote the bit–strings corresponding to computational basis states on the qubit subsets A and $\bar{A}$ respectively. Then the effect $M_A^{(\mathbf{x}_A)}$ of the reduced POVM M_A, corresponding to a projection onto computational basis state $\mathbf{x}_A$, is defined as:

$$M_A^{(\mathbf{x}_A)} := \frac{1}{2^{N-k}} \sum_{\mathbf{x}_{\bar{A}}} \mathrm{Tr}_{\bar{A}}(M^{(\mathbf{x}_A,\mathbf{x}_{\bar{A}})}), \tag{3.1}$$

where $M^{(\mathbf{x}_A,\mathbf{x}_{\bar{A}})}$ denotes effects of the non–reduced POVM M corresponding to projections onto computational basis states labeled by bitstrings $\mathbf{x}_A, \mathbf{x}_{\bar{A}}$. The reduced POVM defined in this way has a clear operational meaning. It corresponds to a measurement on the whole device when the qubits in subset $\bar{A}$ are in the maximally mixed state, the outcome of the measurement on $\bar{A}$ is forgotten and only the outcome on A is remembered (as illustrated in Figure 3.1).

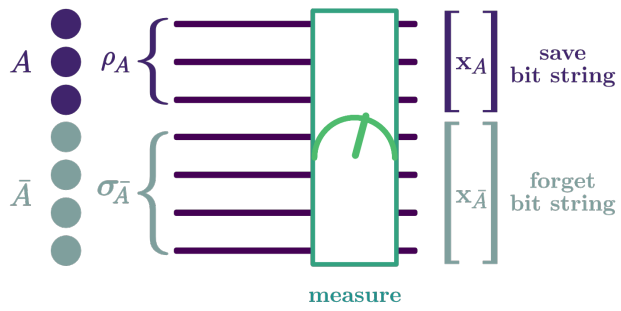


Figure 3.1: A schematic illustration of a reduced POVM M_A.

The procedure of estimating the matrix elements of the reduced POVM's effects for all l–qubit subsets in the device can be outlined in the following steps:

1. Generate a set $\mathbf{Y}^C$ of C strings $\mathbf{Y}^i$ of length n, each string representing a random n–qubit circuit with a state on each qubit chosen in i.i.d. manner from Pauli eigenstates.
2. For each string $\mathbf{Y}^i \in \mathbf{Y}^C$ prepare the corresponding circuit and measure it s times, to obtain the experimental probability bitstring $\mathbf{x}^i$. The set of all C such bitstrings is $\mathbf{x}^C$.
3. For each l–qubit subset A in the device estimate the frequencies of measuring $\mathbf{x}_A$ when $\mathbf{y}_A$ was prepared, based on $\mathbf{Y}^C$ and $\mathbf{x}^C$. From those empirical frequencies reconstruct the reduced POVM M^A effects using the PLS method.

Here PLS means projected least–squares. First, matrix elements are reconstructed using least–sqares. Then the resulting matrices are projected onto positive semi-definite matrices. s denotes the number of shots, meaning how many times a specific circuit was executed to obtain the experimental probability vector.

3.1.2 Details of experiments

The details of the data which is analyzed below are presented in Table 3.1. The data is awaiting publication in [21]. On both devices only subsets of all qubits were characterized because of low fidelity of single–qubit gates on the excluded qubits, which would reduce the accuracy of the tomography. 2–qubit POVMs on all the pairs in the subsets of characterized qubits were reconstructed.

device name	number of qubits characterized (all)	experiment date	number of circuits C	number of shots s
ibm_washington	109 (127)	26 April 2022	1200	10 000
ASPEN–M–1	56 (80)	01 March 2022	300	10 000

Table 3.1: Details of DOT experiments on IBM's and Rigetti's devices.

3.1.3 Errors of estimation

In [21] theoretical bounds on the error of estimation of POVM effects' terms are given. However, they are not tight and here we attempted to estimate those errors heuristically. After a given DOT experiment, the C circuits were randomly divided into 10 batches. Then for each batch the experimental data was treated as data from a separate DOT procedure with $C_{batch} = \frac{1}{10}C$ circuits and used to reconstruct the 2–qubit reduced POVMs. In what follows we will call these results POVMs from batches. The POVMs reconstructed from all C circuits will be called POVMs from full data.

Then an attempt could be made to determine the error of estimating the value of a function for a given POVM. The lowest and highest value of the considered function was found among all 10 batches. They were treated as the lower– and upper– limit of the interval within which the value of the function would fall for the true POVM.

3.2 Results of optimization

For all of the 2–qubit POVMs the mitigation methods were tested by verifying how much a considered cost function changes after a POVM is rotated by a unitary resulting from optimization. To conduct the tests, a program in Python was written. The code, experimental data and optimization results are available in the online repository [52].

3.2.1 Numerical optimization of operationally meaningful functions over local unitaries

For each of the operationally motivated functions (2.3)–(2.4), (2.6)–(2.9) a numerical optimization was conducted. For a given 2–qubit POVM, 2 local unitaries were found, which minimized the considered cost function over $\mathrm{SU}(2) \otimes \mathrm{SU}(2)$. Each local unitary was parameterized with 2 complex parameters c, s with 1 nonlinear constraint: $|c|^2 + |s|^2$, so effectively the optimization was carried out over 6 real parameters. The solver trust-constr available in the scipy.optimize library was used.

The results for the ASPEN–M–1 device are presented in Figures 3.3–3.5. The results for the ibm_washington device are presented in Figures A.1–A.3 in Appendix A.2. For each cost function the results for 20 pairs for which the biggest reduction was obtained are plotted. The plots present both the change in the cost function which was optimized and the change in other cost functions after that optimization. The legend for all the plots is shown in Figure 3.2.

3.2.2 Half–analytical optimization over local unitaries

For the two heuristically motivated cost functions 2.5, 2.10, whose analytical optimization over one local unitary was presented in the previous chapter, a half–analytical method was implemented. For a given 2–qubit POVM in one iteration a qubit was chosen and on that quibt a local unitary which optimizes a cost function was found analytically. Then the POVM was rotated by that local unitary. In the next iteration the optimization was conducted for the other qubit and for the rotated POVM and so on. 10 iterations were used for each qubit pair. The results were identical up to numerical precision, regardless of the qubit which was first in the optimization. For all of the considered data the method converged after a few iterations. The results of these optimizations are presented for 20 pairs with greatest reductions of cost functions in Figure 3.6 for the ASPEN–M–1 device and in Figure A.4 in Appendix A.2 for the ibm_washington device.

To compare the execution time of this half–analytical method with a numerical one, the numerical optimizations of these two cost functions were also conducted for a sample of 100 pairs, using the method described in the previous section. The half–analytical method was a few hundred times faster than the numerical one, yielding results identical up to numerical precision. For example for all 5886 of the considered pairs of the ibm_washington device, the half–analytical method took around 25 s, while numerically it would take around 1 h (it took around 65 s for 100 pairs).

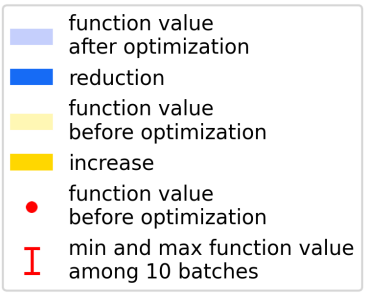


Figure 3.2: Legend for plots in Figures 3.3–3.6 and A.1–A.8 in Appendix A.2, A.3. If the considered function was reduced for a given pair, the reduction is marked with dark blue and the value of the function after optimization is marked with light blue. If the function was increased, the increase is marked with dark yellow and the value of the function after optimization is marked with light yellow. The red dot marks the value of the function before optimization for the POVM reconstructed from the whole data set. The red error–bars signify the heuristically determined error of estimating the value of the function before optimization, described in Section 3.1.3.

3.3 Discussion

3.3.1 Observations from results

IBM vs Rigetti

The measurement errors in IBM's device were significantly lower than in Rigetti's. We don't consider the estimation of coherent errors in this device sufficiently accurate to test our method. For completeness the results are presented in Appendix A.2. In what follows we base our observations on the results for the Rigetti ASPEN–M–1 device.

Worst–case vs average–case

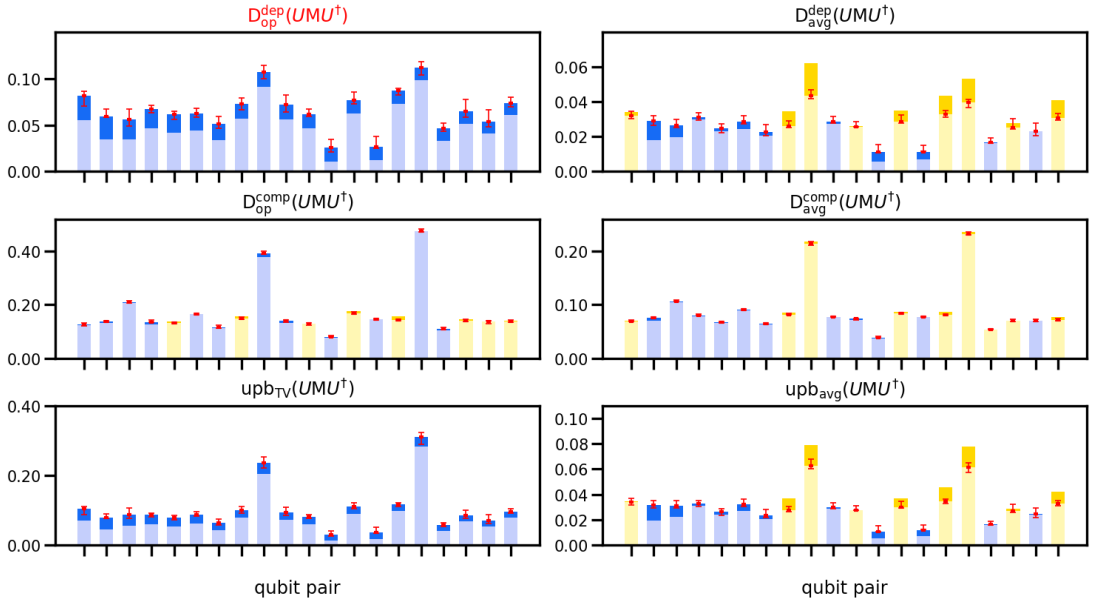
The minimization over worst–case distances lead to a significant increase of the other distances for multiple qubit pairs. This includes the average–case distance from the same POVM, e.g. when the worst–case distance from a dephased POVM was minimized, the average–case distance from dephased POVM was increased for almost half of the pairs.

On the other hand, the minimization over average–case distances also reduced the other cost functions for most pairs. It reduced the other average–case distances for all the plotted pairs. For 150 pairs with the biggest reductions (see Appendix A.3) optimizing over $\mathrm{D}^{\mathrm{dep}}_{\mathrm{avg}}(U\mathsf{M}U^\dagger)$ (see Eq. (2.4)) and $\mathrm{upb}_{\mathrm{avg}}(U\mathsf{M}U^\dagger)$ (see Eq.(2.7)) always reduced the other average-case distances and reduced the worst-case distances for the vast majority of pairs. If a worst–case distance increased as a result of such an optimization, the relative increase was small.

Distance from computational basis vs dephased POVM

The distances from computational basis POVMs are significantly higher that from dephased POVMs. This is understandable, since classical errors dominate over coherent errors (see Section 1.7). The relative reductions of distances are significantly higher when optimizing over distances from dephased POVMs rather than distances from computational basis POVMs. It shows that this method is not promising as a way of mitigating classical errors.

(a) Optimization of $\mathrm{D}_{\mathrm{op}}^{\mathrm{dep}}(U\mathsf{M}U^{\dagger})$:



(b) Optimization of $\mathrm{D}_{\mathrm{avg}}^{\mathrm{dep}}(U\mathsf{M}U^{\dagger})$:

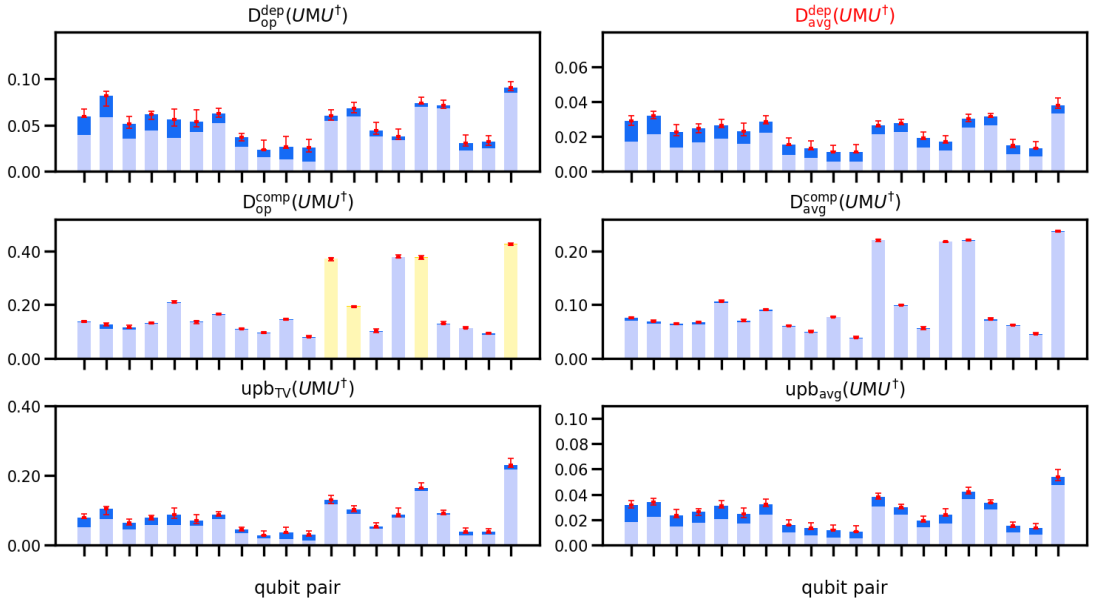
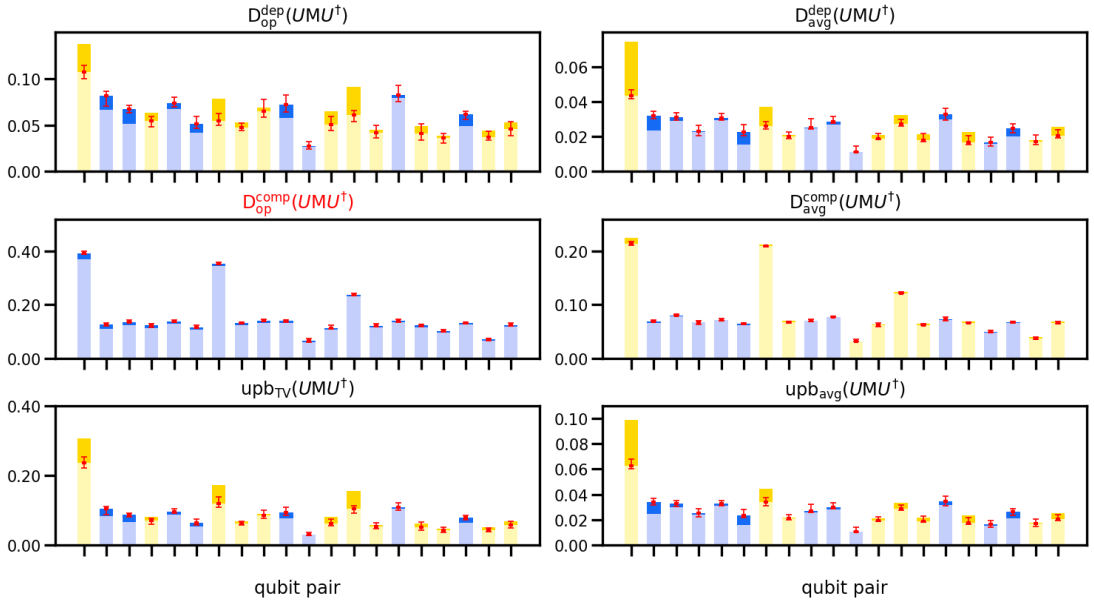


Figure 3.3: **ASPEN–M–1**. Changes in values of cost functions (2.3)–(2.4), (2.6)–(2.9) after rotations of 2–qubit POVMs from the device by 2 local unitaries optimizing the POVM's a) worst–case (2.3), b) average–case (2.4) distance from its **dephased** version. 20 pairs with biggest reductions are plotted. The qubit pairs in the left and right column are the same. A blue bar signifies reduction of the function and yellow its increase. Detailed legend is shown in Figure 3.2.

(a) Optimization of $\mathrm{D}_{\mathrm{op}}^{\mathrm{comp}}(U\mathrm{M}U^{\dagger})$:



(b) Optimization of $\mathrm{D}_{\mathrm{avg}}^{\mathrm{comp}}(U\mathrm{M}U^{\dagger})$:

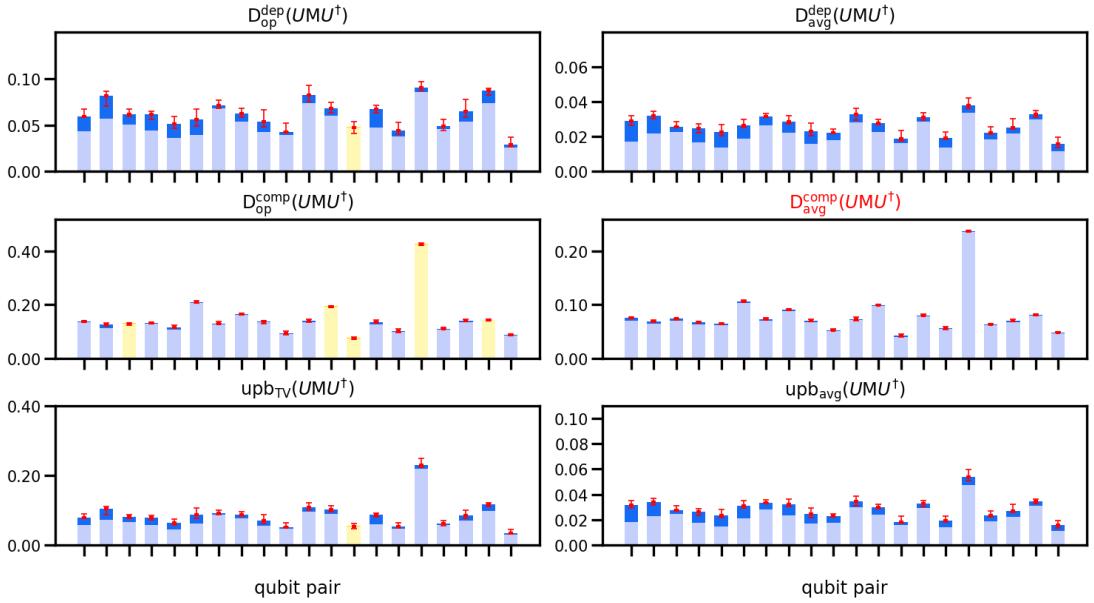
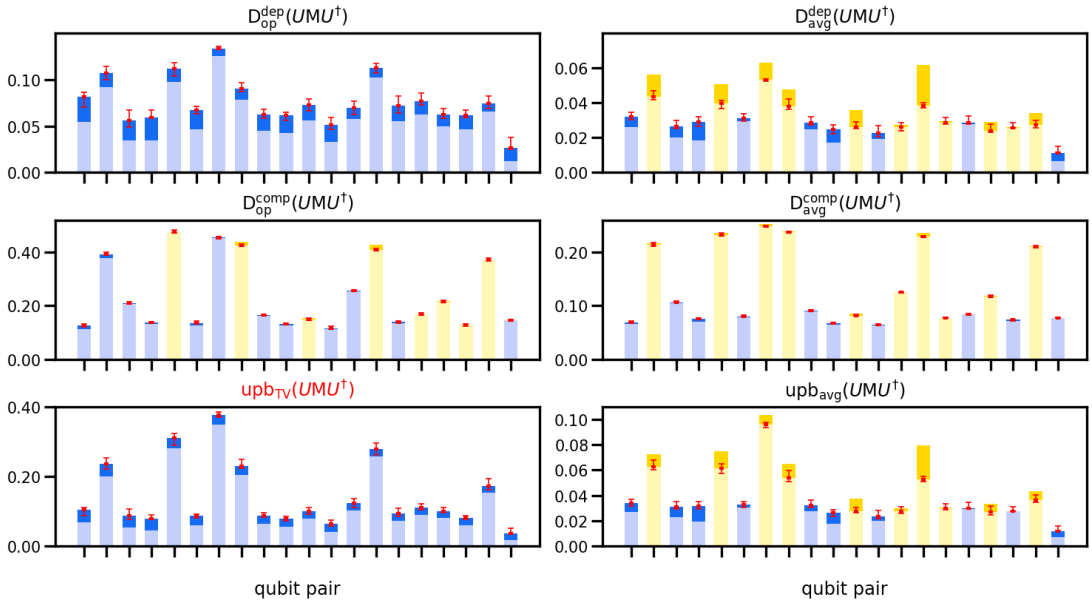


Figure 3.4: **ASPEN–M–1**. Changes in values of cost functions (2.3)–(2.4), (2.6)–(2.9) after rotations of 2–qubit POVMs from the device by 2 local unitaries optimizing the POVM's a) worst–case (2.8), b) average–case (2.9) distance from **computational basis** POVM. 20 pairs with biggest reductions are plotted. The qubit pairs in the left and right column are the same. A blue bar signifies reduction of the function and yellow its increase. Detailed legend is shown in Figure 3.2.

(a) Optimization of $\mathrm{upb}_{\mathrm{TV}}(U\mathsf{M}U^{\dagger})$:



(b) Optimization of $\mathrm{upb}_{\mathrm{avg}}(U\mathsf{M}U^{\dagger})$:

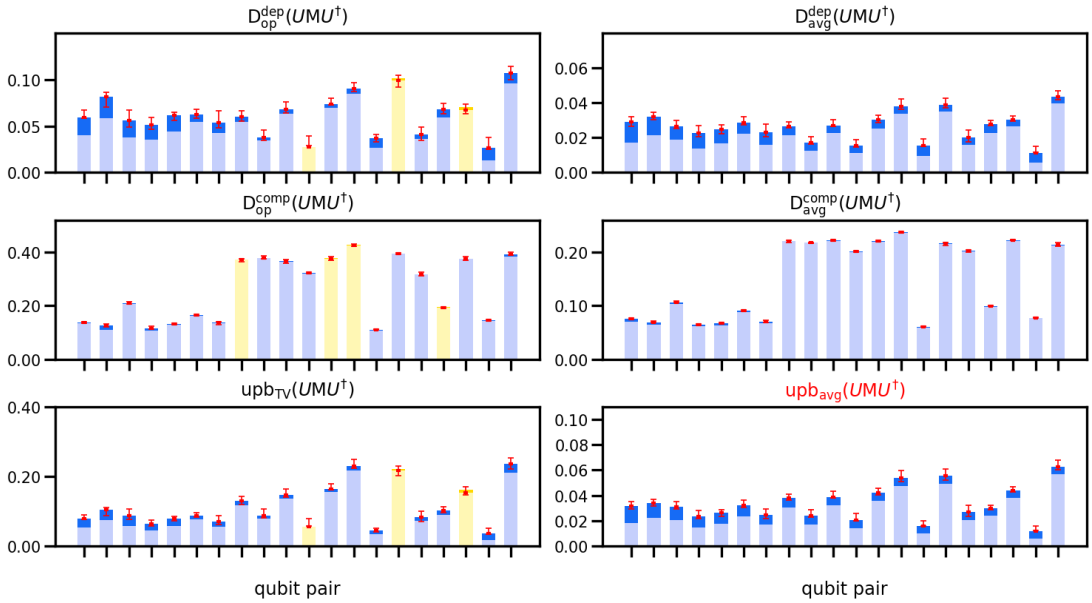
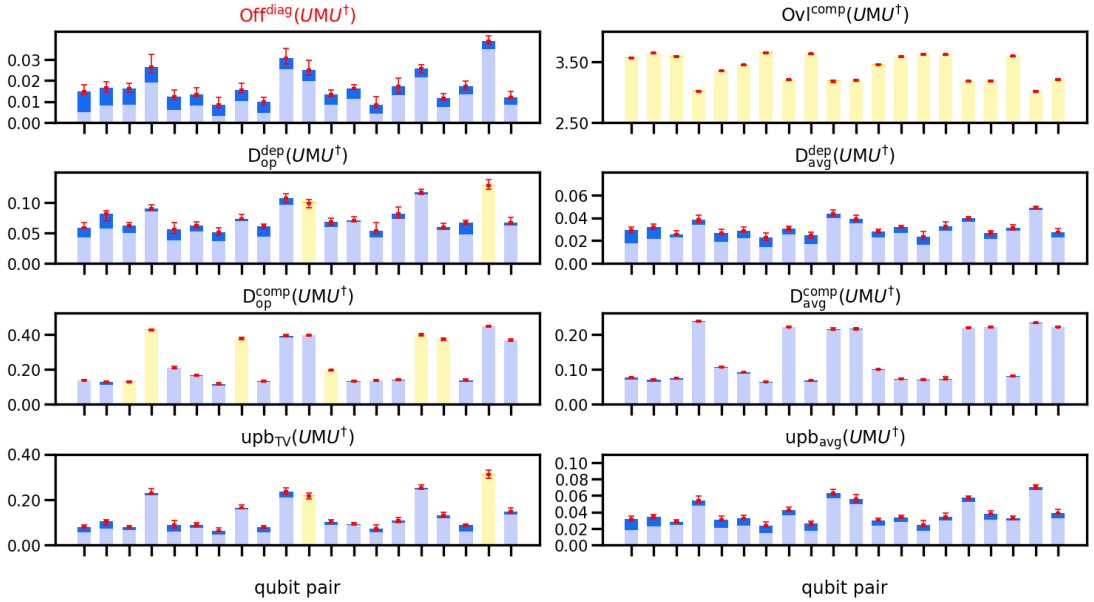


Figure 3.5: **ASPEN–M–1**. Changes in values of cost functions (2.3)–(2.4), (2.6)–(2.9) after rotations of 2–qubit POVMs from the device by 2 local unitaries optimizing the upper–bound on a) worst–case (2.6), b) average–case (2.7) TV distance between ideal probability vector and that resulting from mitigation using method from Section 1.6.2. 20 pairs with biggest reductions are plotted. The qubit pairs in the left and right column are the same. A blue bar signifies reduction of the function and yellow its increase. Detailed legend is shown in Figure 3.2.

(a) Optimization of $\mathrm{Off}^{\mathrm{diag}}(U\mathsf{M}U^\dagger)$:



(b) Optimization of $\mathrm{Ovl}^{\mathrm{comp}}(U\mathsf{M}U^\dagger)$:

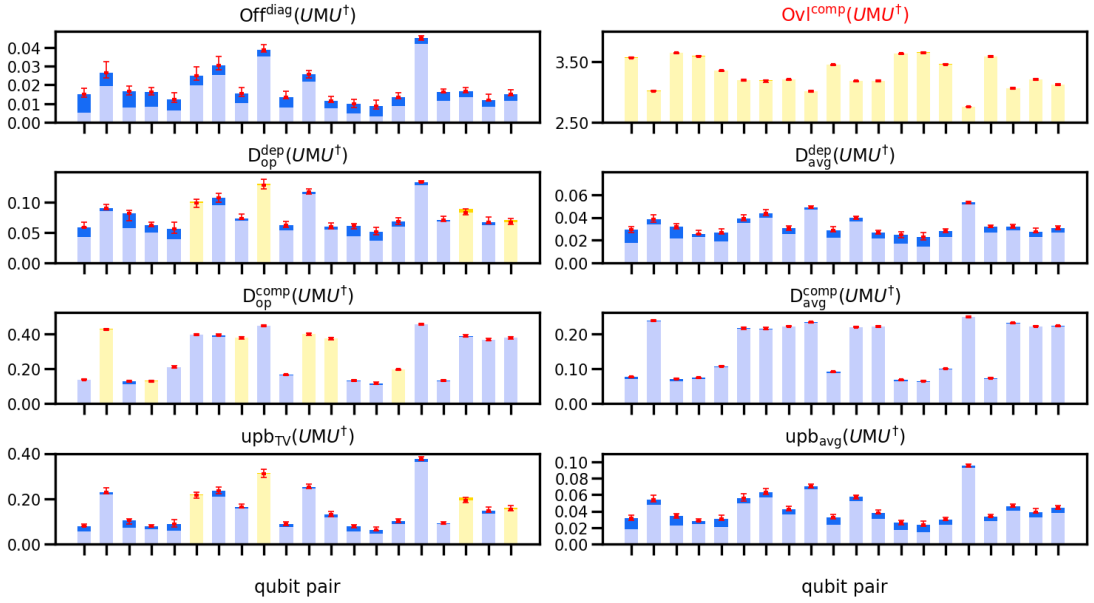


Figure 3.6: **ASPEN–M–1**. Changes in values of cost functions (2.3)–(2.10) after rotations of 2–qubit POVMs from the device by 2 local unitaries optimizing a) $\mathrm{Off}^{\mathrm{diag}}(U\mathsf{M}U^\dagger)$ (2.5), b) $\mathrm{Ovl}^{\mathrm{comp}}(U\mathsf{M}U^\dagger)$ (2.10). 20 pairs with biggest reductions are plotted. The qubit pairs in the left and right column are the same. A blue bar signifies reduction of the function and yellow its increase. Detailed legend is shown in Figure 3.2. Note that the vertical axis for $\mathrm{Ovl}^{\mathrm{comp}}(U\mathsf{M}U^\dagger)$ starts from 2.5.

Operationally vs heuristically motivated functions

The optimization over the distance function $\mathrm{Off}^{\mathrm{diag}}(U\mathsf{M}U^{\dagger})$ (see Eq. (2.5)), quantifying the magnitude of off-diagonal terms in POVMs' effects, also reduces most of the other cost functions (especially the average–case ones) for most qubit pairs. What is more, the set of qubit pairs with the biggest reductions of this cost function has a big overlap with such respective sets for other functions (see data in the online repository [52]). For example for 50 pairs with the biggest reduction this set has over half of elements in common with the respective set for the average distance of a POVM from its dephased version.

Somewhat surprisingly, optimization over distance function $\mathrm{Ovl}^{\mathrm{comp}}(U\mathsf{M}U^{\dagger})$ (see Eq. (2.10)) lead to similar behaviour of other cost functions. While the heuristic motivation for this function was for POVMs to approach computational basis, there was no significant difference in the behaviour of the distances from computational basis between results from optimization of $\mathrm{Off}^{\mathrm{diag}}(U\mathsf{M}U^{\dagger})$ and $\mathrm{Ovl}^{\mathrm{comp}}(U\mathsf{M}U^{\dagger})$.

Overlapping pairs

For a given optimized function some pairs with biggest reductions have a qubit in common (see data in the online repository [52]). Adding to a circuit a unitary which reduces a cost function for one of those pairs might increase it for another. Dividing the qubits into disjoint sets would be one way to avoid it, e.g. using the method of clustering described in [21].

Assessing applicability of the mitigation method

The results suggest that in the ASPEN–M–1 device there exist 2–qubit POVMs with coherent error of order of a few percent (quantified by the average distance from the dephased POVM) for which the proposed method could reduce the error by around 40%.

However, it is crucial to stress that reliable bounds on the errors present in the experimental setup, mentioned in 2.6, are necessary to conclusively assess whether the mitigation method is applicable on a given device. If those bounds are available, a criterion can be developed to assess if applying to a given POVM the unitary found in optimization would indeed always optimize the cost function, taking into account the errors of QDT and gate errors.

While the gate fidelity is out of our control, an important next step would be to obtain analytically or estimate experimentally reliable bounds on the estimation errors of a QDT procedure on which our method relies.

Chapter 4

Summary and outlook

4.1 Results summary

In this thesis we have proposed a method of mitigation of measurement errors in quantum devices with the use of gates applied to a circuit directly before measurement. The method first requires Quantum Detector Tomography to be conducted on a device. Based on its results, a unitary transformation is found, such that a cost function is minimized when this unitary is added to a circuit, effectively modifying the POVM. We have proposed 6 operationally meaningful and 2 heuristically motivated cost functions (see Table 2.1). They quantify the distance of the POVM from its dephased version or from computational basis POVM.

We have provided analytical methods of optimizing the two heuristically motivated cost functions over local unitary gates, inspired by the method of Approximate Simultaneous Diagonalization of matrices presented in [46].

We have tested our method on data obtained from Diagonal Detector Tomography of state-of-the art quantum devices. The accuracy of QDT of IBM's ibm_washington device was considered insufficient to conclusively assess the method's efficacy. For Rigetti's ASPEN–M–1 device the results might suggest a possible usefulness of our method for mitigation of coherent errors. However, a better understanding of statistical errors in tomography is necessary for a reliable conclusion.

4.2 Outlook

Several directions for future research expanding on our work can be outlined.

- Reliable bounds on the errors of estimation of POVM are needed, either derived analytically or estimated experimentally.
- A criterion (see 3.3.1) to assess if the method is applicable for a given device and POVM is needed before implementation of the method on real devices.
- The effect of minimising measurement errors on one qubit subset on those errors on another, overlapping, subset (see Section 3.3.1) needs to be studied. Possibly dividing all qubits into disjoint subsets (see e.g. [21]) might solve this problem.
- If those problems are solved, other devices and cost functions can be tested.
- The proposed closed–form solution for an optimal 1–qubit unitary for ASD may find its application in other problems within quantum information.

Bibliography

[1]Y. Manin, *Computable and uncomputable (in Russian)*, Kibernetika (Moscow, 1980).

[2]R. P. Feynman, "Simulating physics with computers", International Journal of Theoretical Physics **21**, 467–488 (1982).

[3]G. H. Low and I. L. Chuang, "Hamiltonian simulation by qubitization", Quantum **3**, 163 (2019).

[4]P. Shor, "Algorithms for quantum computation: discrete logarithms and factoring", in Proceedings 35th annual symposium on foundations of computer science (Nov. 1994), pp. 124–134.

[5]S. Lee et al., "Is there evidence for exponential quantum advantage in quantum chemistry?", `10.48550/arXiv.2208.02199` (2022).

[6]P. W. Shor, "Scheme for reducing decoherence in quantum computer memory", Physical Review A **52**, R2493–R2496 (1995).

[7]R. Acharya et al., "Suppressing quantum errors by scaling a surface code logical qubit", `10.48550/arXiv.2207.06431` (2022).

[8]J. Preskill, "Quantum computing in the NISQ era and beyond", Quantum **2**, 79 (2018).

[9]K. Temme, S. Bravyi, and J. M. Gambetta, "Error mitigation for short-depth quantum circuits", Physical Review Letters **119**, 180509 (2017).

[10]W. J. Huggins et al., "Virtual distillation for quantum error mitigation", Physical Review X **11**, 041036 (2021).

[11]C. Song, J. Cui, H. Wang, J. Hao, H. Feng, and Y. Li, "Quantum computation with universal error mitigation on a superconducting quantum processor", Science Advances **5**, eaaw5686 (2019).

[12]A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, "Error mitigation extends the computational reach of a noisy quantum processor", Nature **567**, 491–495 (2019).

[13]E. v. d. Berg, Z. K. Minev, A. Kandala, and K. Temme, "Probabilistic error cancellation with sparse pauli-lindblad models on noisy quantum processors", `10.48550/arXiv.2201.09866` (2022).

[14]Y. Chen, M. Farahzad, S. Yoo, and T.-C. Wei, "Detector tomography on IBM quantum computers and mitigation of an imperfect measurement", Physical Review A **100**, 052315 (2019).

[15]G. J. Mooney, G. A. L. White, C. D. Hill, and L. C. L. Hollenberg, "Generation and verification of 27-qubit greenberger-horne-zeilinger states in a superconducting quantum computer", Journal of Physics Communications **5**, 095004 (2021).

[16]F. B. Maciejewski, Z. Zimborás, and M. Oszmaniec, "Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography", Quantum **4**, 257 (2020).

[17]S. Zhang et al., "Error-mitigated quantum gates exceeding physical fidelities in a trapped-ion system", Nature Communications **11**, 587 (2020).

[18]F. Arute et al., "Quantum supremacy using a programmable superconducting processor", Nature **574**, 505–510 (2019).

[19]H.-Y. Huang et al., "Quantum advantage in learning from experiments", Science **376**, 1182–1186 (2022).

[20]M. R. Geller and M. Sun, "Toward efficient correction of multiqubit measurement errors: pair correlation method", Quantum Science and Technology **6**, 025009 (2021).

[21]J. Tuziemski, F. B. Maciejewski, J. Majsak, O. Słowik, and M. Oszmaniec, "Learning, benchmarking and validation of cross-talk models in readout noise in near-term quantum devices", awaiting publication.

[22] *Measurement error mitigation*, https://community.qiskit.org/textbook/ch-quantum-hardware/measurement-error-mitigation.html (visited on 08/09/2022).
[23] F. B. Maciejewski, F. Baccari, Z. Zimborás, and M. Oszmaniec, "Modeling and mitigation of cross-talk effects in readout noise with applications to the quantum approximate optimization algorithm", Quantum **5**, 464 (2021).
[24] S. Bravyi, S. Sheldon, A. Kandala, D. C. Mckay, and J. M. Gambetta, "Mitigating measurement errors in multi-qubit experiments", Physical Review A **103**, 042605 (2021).
[25] S. Seo, J. Seong, and J. Bae, "Mitigation of crosstalk errors in a quantum measurement and its applications", 10.48550/arXiv.2112.10651 (2021).
[26] E. v. d. Berg, Z. K. Minev, and K. Temme, "Model-free readout-error mitigation for quantum expectation values", Physical Review A **105**, 032620 (2022).
[27] S. Tang, C. Zheng, and K. Wang, "Detecting and eliminating quantum noise of quantum measurements", 10.48550/arXiv.2206.13743 (2022).
[28] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information: 10th anniversary edition*, 1st ed. (Cambridge University Press, June 5, 2012).
[29] L. Tisza, *The Pauli algebra*, Mathematics LibreTexts, (June 28, 2020) https://math.libretexts.org/Bookshelves/Abstract_and_Geometric_Algebra/Applied_Geometric_Algebra_(Tisza) (visited on 09/02/2022).
[30] J. McMillan, "SU(2): a primer", 10.13140/RG.2.2.20216.06405 (2019).
[31] F. B. Maciejewski, Z. Puchała, and M. Oszmaniec, "Exploring quantum average-case distances: proofs, properties, and examples", 10.48550/arXiv.2112.14284 (2022).
[32] E. Magesan, R. Blume-Kohout, and J. Emerson, "Gate fidelity fluctuations and quantum process invariants", Physical Review A **84**, 012309 (2011).
[33] E. Knill et al., "Randomized benchmarking of quantum gates", Physical Review A **77**, 012307 (2008).
[34] J. S. Lundeen et al., "Tomography of quantum detectors", Nature Physics **5**, 27–30 (2009).
[35] L. Zhang et al., "Mapping coherence in measurement via full quantum tomography of a hybrid optical detector", Nature Photonics **6**, 364–368 (2012).
[36] J. Z. Blumoff et al., "Implementing and characterizing precise multiqubit measurements", Physical Review X **6**, 031041 (2016).
[37] J. Cotler and F. Wilczek, "Quantum overlapping tomography", Physical Review Letters **124**, 100401 (2020).
[38] *IBM Quantum*, https://quantum-computing.ibm.com/services/resources (visited on 08/09/2022).
[39] R. Akema, M. Yamagishi, and I. Yamada, "Approximate simultaneous diagonalization of matrices via structured low-rank approximation", IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences **104**, 680–690 (2021).
[40] R. Dawes and T. Carrington, "Using simultaneous diagonalization and trace minimization to make an efficient and simple multidimensional basis for solving the vibrational schrödinger equation", The Journal of Chemical Physics **124**, 054102 (2006).
[41] D. T. Pham, "Joint approximate diagonalization of positive definite hermitian matrices", SIAM Journal on Matrix Analysis and Applications **22**, 1136–1152 (2001).
[42] P. A. Ablin, J.-F. Cardoso, and A. Gramfort, "Beyond pham's algorithm for joint diagonalization", in ESSAN 2019 - 27th european symposium on artificial neural networks (Apr. 2019).
[43] B. Afsari, "Simple LU and QR based non-orthogonal matrix joint diagonalization", in Proceedings of the 6th international conference on independent component analysis and blind signal separation, ICA'06 (Mar. 5, 2006), pp. 1–7.
[44] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals", IEE Proceedings F (Radar and Signal Processing) **140**, 362–370 (1993).
[45] V. Maurandi, E. Moreau, and C. De Luigi, "Jacobi like algorithm for non-orthogonal joint diagonalization of hermitian matrices", in 2014 IEEE international conference on acoustics, speech and signal processing (ICASSP) (May 2014), pp. 6196–6200.
[46] J.-F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization", SIAM Journal on Matrix Analysis and Applications **17**, 161–164 (1996).

[47]V. Hari and E. B. Kovac, "Convergence of the cyclic and quasi-cyclic block jacobi methods", Electronic Transactions on Numerical Analysis **46**, 107–148 (2017).
[48]V. Hari and E. B. Kovač, "On the convergence of complex Jacobi methods", Linear and Multilinear Algebra **69**, 489–514 (2021).
[49]K. Usevich, J. Li, and P. Comon, "Approximate matrix and tensor diagonalization by unitary transformations: convergence of jacobi-type algorithms", Feb. 2020.
[50]J. Li, K. Usevich, and P. Comon, "On the convergence of Jacobi-type algorithms for independent component analysis", in 2020 IEEE 11th sensor array and multichannel signal processing workshop (SAM) (June 2020), pp. 1–5.
[51]M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, "Experimental realization of any discrete unitary operator", Physical Review Letters **73**, 58–61 (1994).
[52]jo-majs, *POVM distances optimization*, `https://gitfront.io/r/jo-majs/iWwTtWkFcT7j/POVM-distances-optimization/`.

Appendix A

A.1 Upper–bound of average TV distance between ideal and mitigated probability vectors

Fact 1. *Let* $\mathbf{p}_{\psi_V,\mathsf{M}}$ *denote the probability vector obtained from measuring state* $\psi_V = V|0\rangle_d\langle 0|V^\dagger$ *with POVM* M*. Let* $\mathbf{p}^{\text{ideal}}$ *denote the probability vector which would be obtained in an ideal experiment and* $\mathbf{p}^{\text{mit}}$ *the probability vector obtained from the mitigation procedure 1.6.2. Let* μ *be the Haar measure on* $\mathrm{SU}(d)$*. Then the average over all pure states* ψ_V *of the TV distance between the probability distributions* $\mathbf{p}^{\text{mit}}_{\psi_V,\mathsf{M}}$ *and* $\mathbf{p}^{\text{ideal}}_{\psi_V,\mathsf{M}}$ *is upper–bounded by:*

$$\mathbb{E}_{V\sim\mu}\mathrm{D}_{\mathrm{TV}}(\mathbf{p}^{\text{mit}}_{\psi_V,\mathsf{M}},\mathbf{p}^{\text{ideal}}_{\psi_V,\mathsf{M}}) \leq \frac{1}{2\sqrt{d(d+1)}}\sum_{j=1}^{d}\sqrt{\mathrm{Tr}(\theta^{(j)2})+\mathrm{Tr}^2(\theta^{(j)})}, \tag{A.1}$$

where:

$$\theta^{(j)} = \sum_{k=1}^{d}\left[\Lambda^{-1}_{jk}\cdot[M^{(k)}-\mathrm{diag}(M^{(k)})]\right] \tag{A.2}$$

and Λ *denotes the stochastic noise matrix used in the mitigation method, defined in* (1.30)*.*

Proof. Recall from (1.33) that a noisy, experimentally available POVM M can be written as a sum of the parts with classical and coherent noise as:

$$\mathsf{M} = \Lambda\mathsf{P} + \Delta, \tag{A.3}$$

where Δ is the coherent part of the POVM: $\Delta^{(k)} = M^{(k)} - \mathrm{diag}(M^{(k)})$. The k–th effect of M is:

$$M^{(k)} = \sum_{l=1}^{d}\Lambda_{kl}P^{(l)} + \Delta^{(k)}. \tag{A.4}$$

Then from Born's rule we can get the k–th element of the probability vector obtained experimentally (assuming no statistical errors) from measuring state ψ with POVM M:

$$p^{\exp}(k|\psi) = \mathrm{Tr}(M^{(k)}\psi) = \sum_{l=1}^{d}\Lambda_{kl}\,p^{\text{ideal}}(l|\psi) + \mathrm{Tr}(\Delta^{(k)}\psi). \tag{A.5}$$

The j–th element of the vector obtained from the mitigation procedure is given by:

$$p^{\text{mit}}(j|\psi) = \sum_{k=1}^{d} \Lambda_{jk}^{-1} p^{\text{exp}}(k|\psi) = p^{\text{ideal}}(j|\psi) + \text{Tr}(\sum_{k=1}^{d} \Lambda_{jk}^{-1} \Delta^{(k)} \psi). \tag{A.6}$$

Let us now denote:

$$\theta^{(j)} := \sum_{k=1}^{d} \Lambda_{jk}^{-1} \Delta^{(k)}. \tag{A.7}$$

Then the TV distance between the ideal probability vector and that resulting from the mitigation procedure is given by:

$$\text{D}_{\text{TV}}(\mathbf{p}_{\psi,\mathsf{M}}^{\text{mit}}, \mathbf{p}_{\psi,\mathsf{M}}^{\text{ideal}}) = \frac{1}{2} \sum_{k=j}^{d} |p^{\text{mit}}(j|\psi) - p^{\text{ideal}}(j|\psi)| = \frac{1}{2} \sum_{k=j}^{d} |\text{Tr}(\theta^{(j)} \psi)|. \tag{A.8}$$

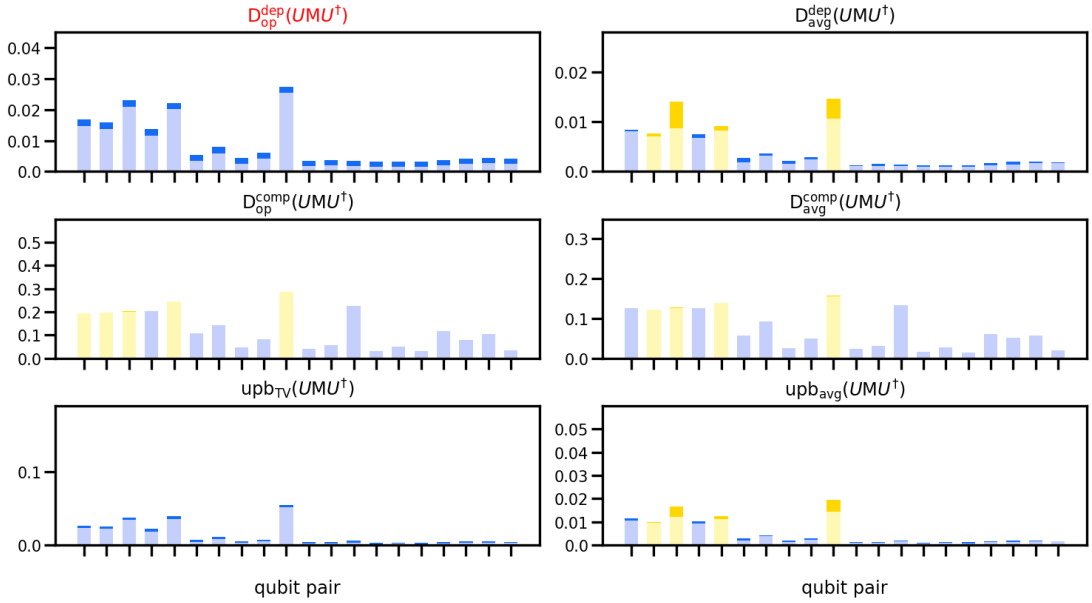
Now we can average this distance over all pure states ψ_V:

$$\begin{aligned} \mathbb{E}_{V\sim\mu} \text{D}_{\text{TV}}(\mathbf{p}_{\psi_V,\mathsf{M}}^{\text{mit}}, \mathbf{p}_{\psi_V,\mathsf{M}}^{\text{ideal}}) &= \frac{1}{2} \sum_{j=1}^{d} \mathbb{E}_{V\sim\mu} |\text{Tr}(\theta^{(j)} \psi_V)| \\ &\leq \frac{1}{2} \sum_{j=1}^{r} \sqrt{\mathbb{E}_{V\sim\mu} \text{Tr}^2(\theta^{(j)} \psi_V)} \\ &= \frac{1}{2} \sum_{j=1}^{r} \frac{1}{\sqrt{d(d+1)}} \sqrt{\text{Tr}(\theta^{(j)2}) + \text{Tr}^2(\theta^{(j)})}. \end{aligned} \tag{A.9}$$

The inequality is due to Jensen's inequality for a random variable A: $\langle A \rangle \leq \sqrt{\langle A^2 \rangle}$. The last equality results directly from Corollary 1 in [31]. □

A.2 DOT and optimization results for IBM's device

(a) Optimization of $\mathrm{D}_{\mathrm{op}}^{\mathrm{dep}}(U\mathsf{M}U^{\dagger})$:



(b) Optimization of $\mathrm{D}_{\mathrm{avg}}^{\mathrm{dep}}(U\mathsf{M}U^{\dagger})$:

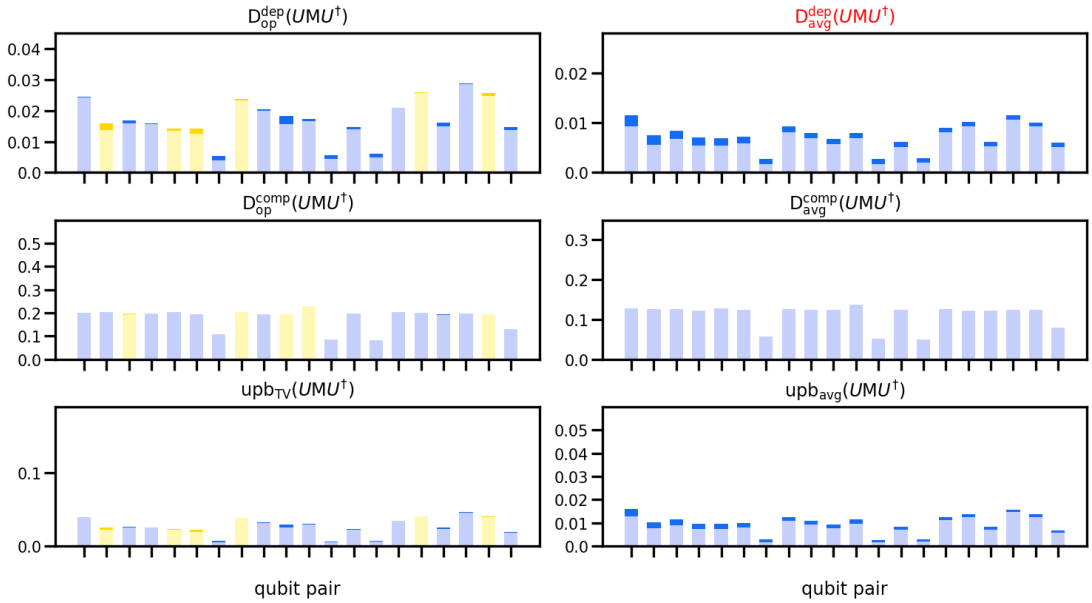
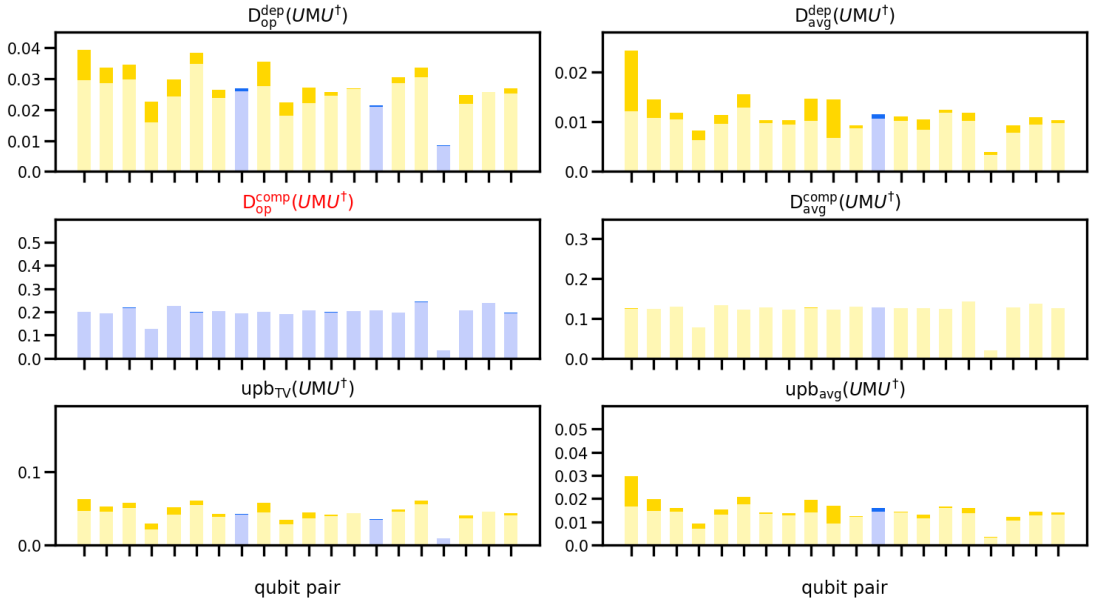


Figure A.1: **ibm_washington**. Changes in values of cost functions (2.3)–(2.4), (2.6)–(2.9) after rotations of 2–qubit POVMs from the device by 2 local unitaries optimizing the POVM's a) worst–case (2.3), b) average–case (2.4) distance from its **dephased** version. 20 pairs with biggest reductions are plotted. The qubit pairs in the left and right column are the same. A blue bar signifies reduction of the function and yellow its increase. Detailed legend is shown in Figure 3.2.

(a) Optimization of $\mathrm{D}_{\mathrm{op}}^{\mathrm{comp}}(U\mathrm{M}U^{\dagger})$:



(b) Optimization of $\mathrm{D}_{\mathrm{avg}}^{\mathrm{comp}}(U\mathrm{M}U^{\dagger})$:

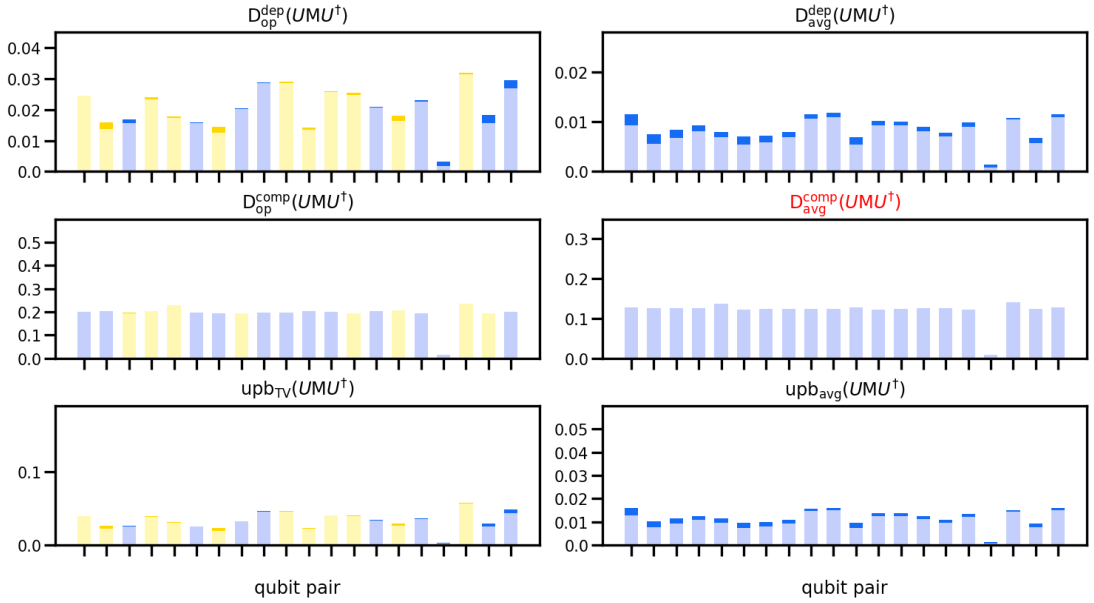
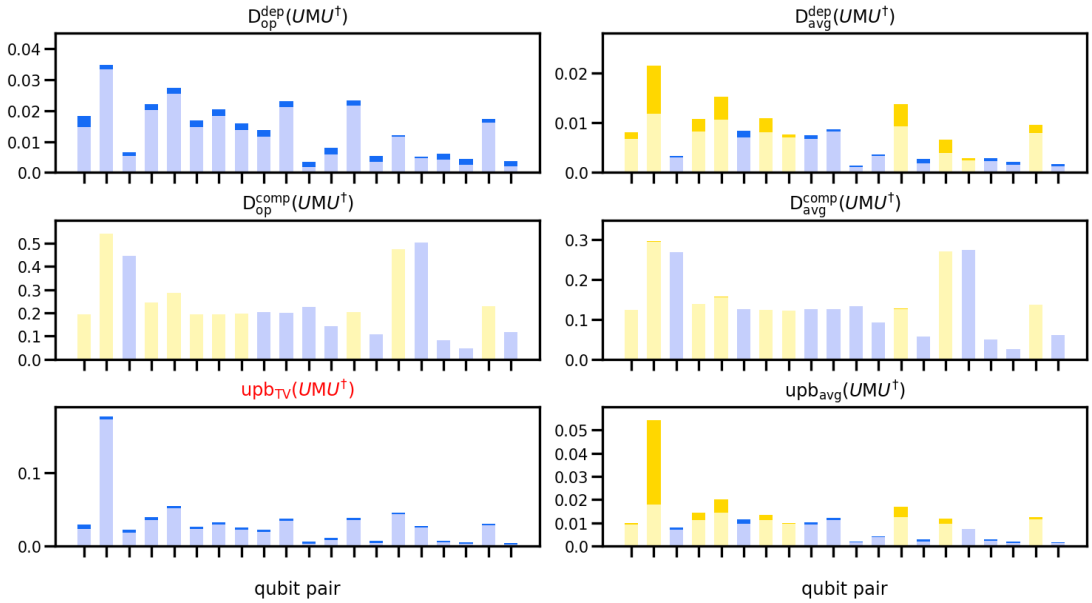


Figure A.2: **ibm_washington**. Changes in values of cost functions (2.3)–(2.4), (2.6)–(2.9) after rotations of 2–qubit POVMs from the device by 2 local unitaries optimizing the POVM's a) worst–case (2.8), b) average–case (2.9) distance from **computational basis** POVM. 20 pairs with biggest reductions are plotted. The qubit pairs in the left and right column are the same. A blue bar signifies reduction of the function and yellow its increase. Detailed legend is shown in Figure 3.2.

(a) Optimization of $\mathrm{upb}_{\mathrm{TV}}(U\mathsf{M}U^\dagger)$:



(b) Optimization of $\mathrm{upb}_{\mathrm{avg}}(U\mathsf{M}U^\dagger)$:

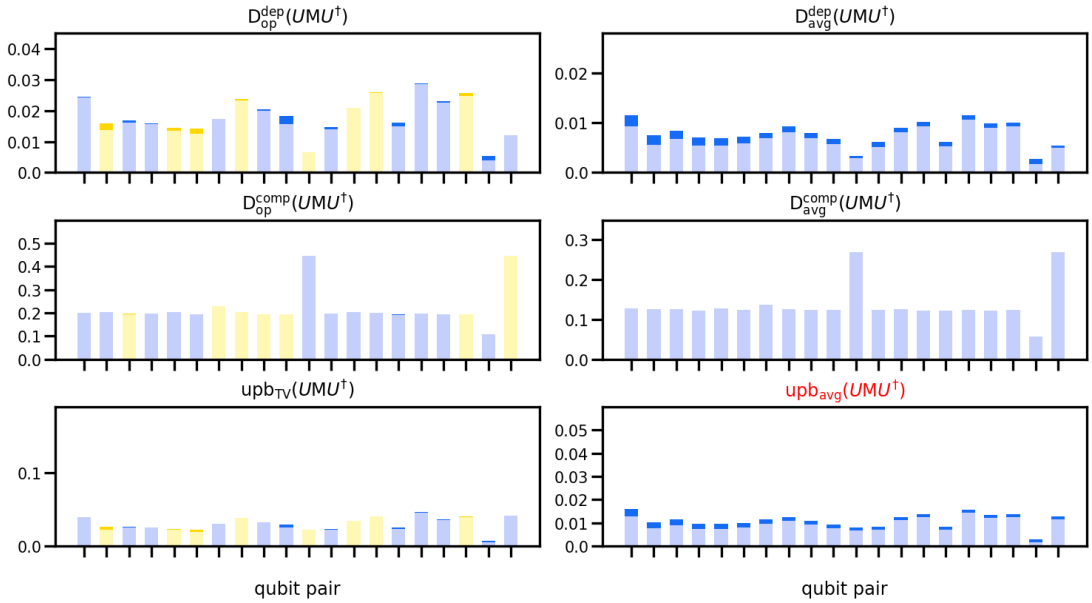
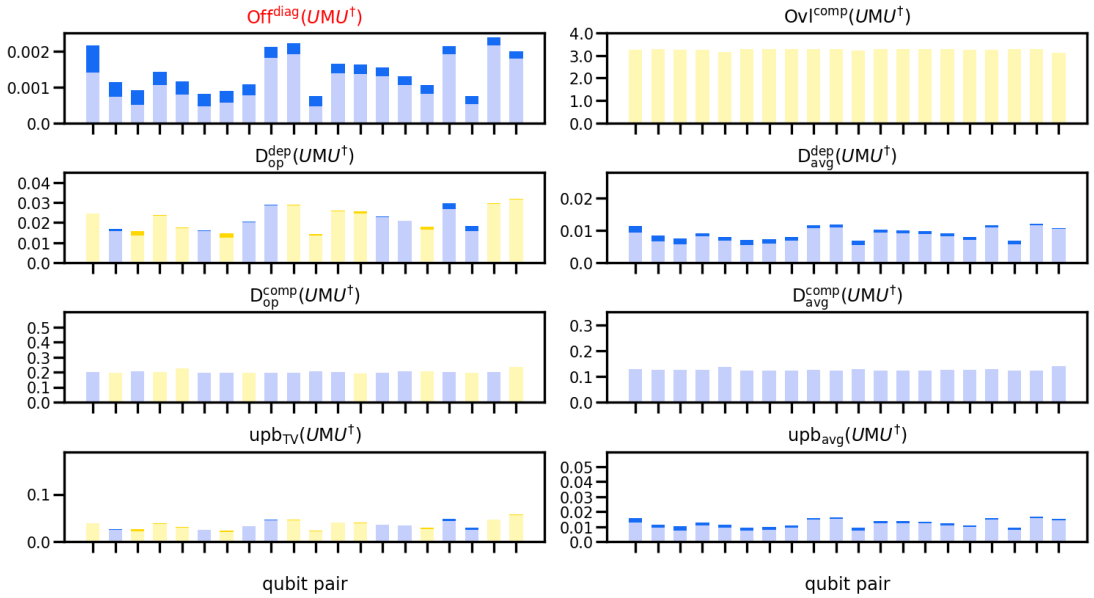


Figure A.3: **ibm_washington**. Changes in values of cost functions (2.3)–(2.4), (2.6)–(2.9) after rotations of 2–qubit POVMs from the device by 2 local unitaries optimizing the upper–bound on a) worst–case (2.6), b) average–case (2.7) TV distance between ideal probability vector and that resulting from mitigation using method from Section 1.6.2. 20 pairs with biggest reductions are plotted. The qubit pairs in the left and right column are the same. A blue bar signifies reduction of the function and yellow its increase. Detailed legend is shown in Figure 3.2.

(a) Optimization of $\mathrm{Off}^{\mathrm{diag}}(U\mathsf{M}U^\dagger)$:



(b) Optimization of $\mathrm{Ovl}^{\mathrm{comp}}(U\mathsf{M}U^\dagger)$:

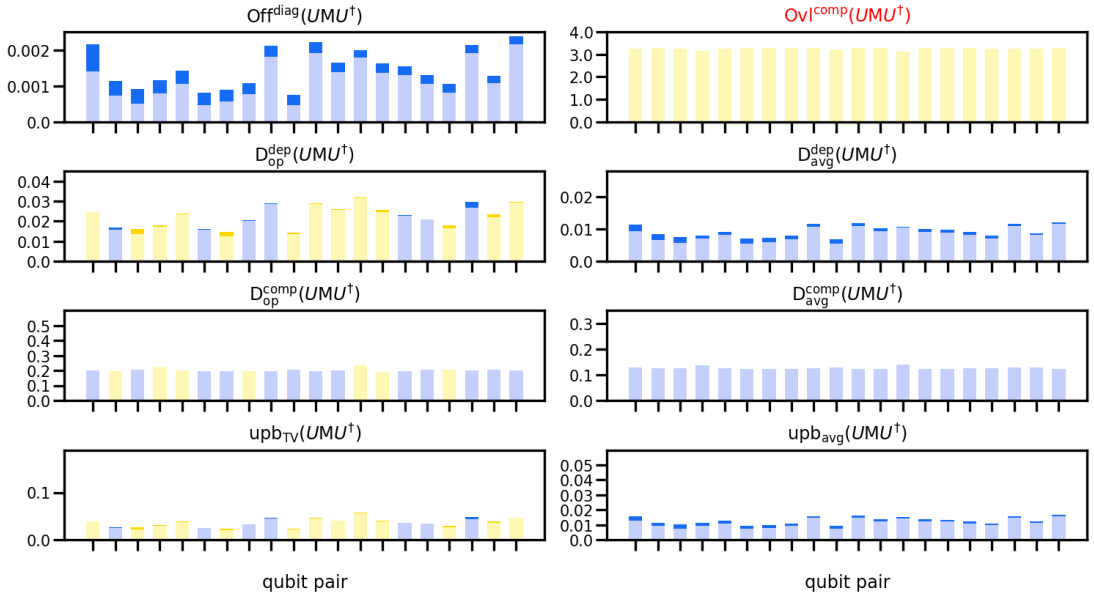
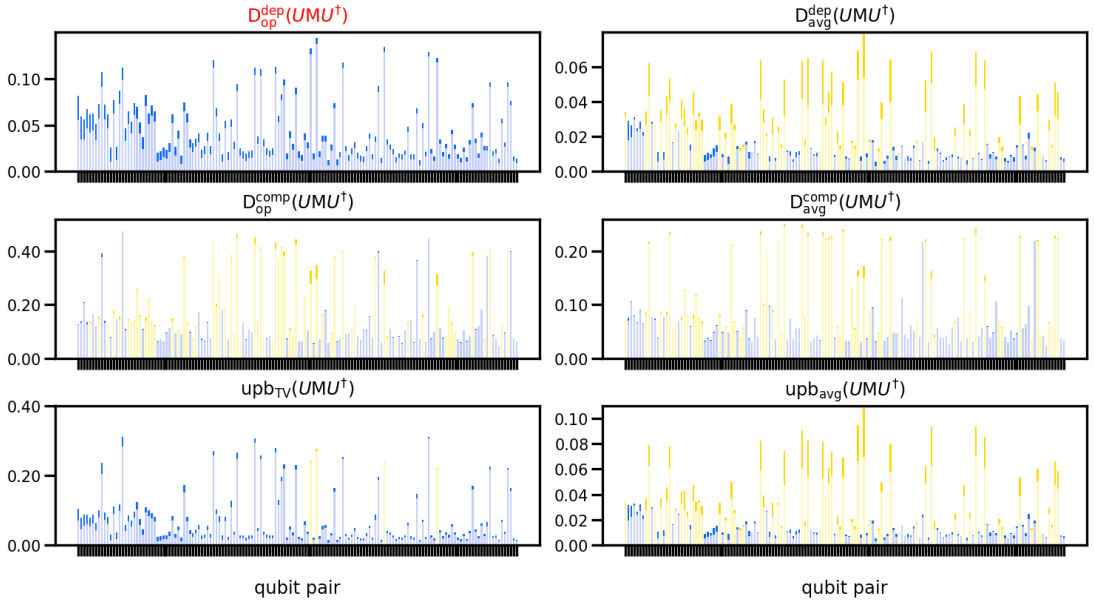


Figure A.4: **ibm_washington**. Changes in values of cost functions (2.3)–(2.10) after rotations of 2–qubit POVMs from the device by 2 local unitaries optimizing a) $\mathrm{Off}^{\mathrm{diag}}(U\mathsf{M}U^\dagger)$ (2.5), b) $\mathrm{Ovl}^{\mathrm{comp}}(U\mathsf{M}U^\dagger)$ (2.10). 20 pairs with biggest reductions are plotted. The qubit pairs in the left and right column are the same. A blue bar signifies reduction of the function and yellow its increase. Detailed legend is shown in Figure 3.2. Note that the vertical axis for $\mathrm{Ovl}^{\mathrm{comp}}(U\mathsf{M}U^\dagger)$ starts from 2.5.

A.3 Supplementary DOT and optimization results for Rigetti's device

(a) Optimization of $\mathrm{D}_{\mathrm{op}}^{\mathrm{dep}}(U\mathsf{M}U^{\dagger})$:



(b) Optimization of $\mathrm{D}_{\mathrm{avg}}^{\mathrm{dep}}(U\mathsf{M}U^{\dagger})$:

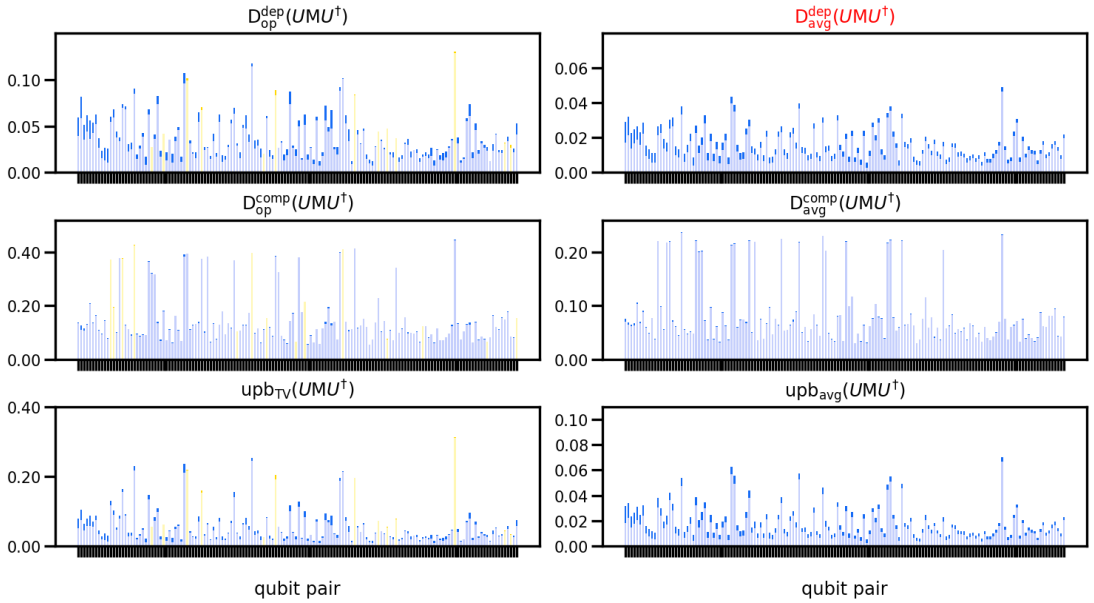
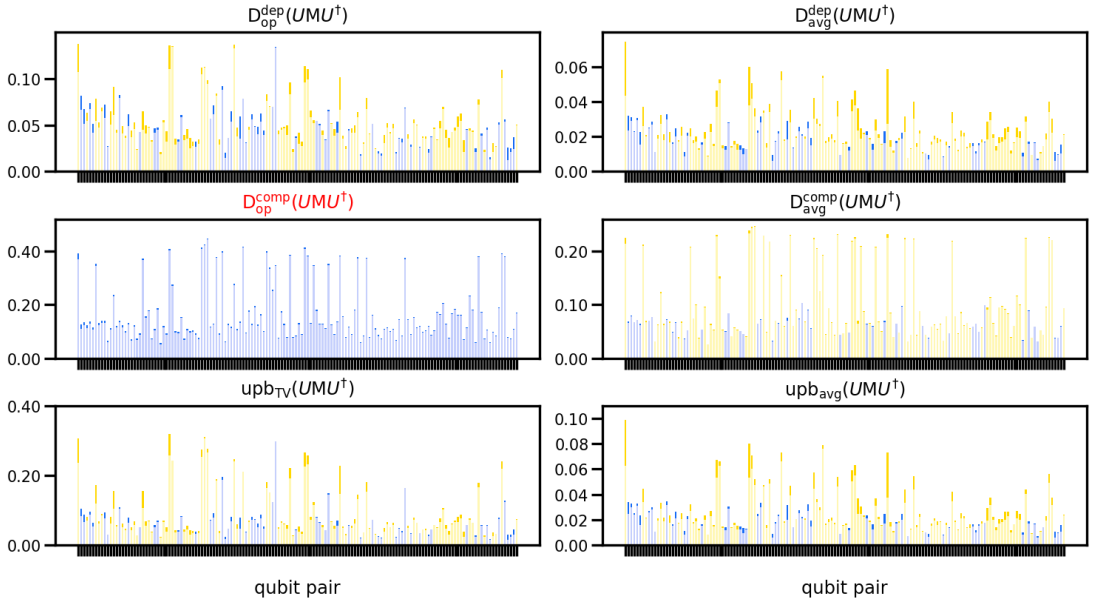


Figure A.5: **ASPEN–M–1**. Changes in values of cost functions (2.3)–(2.4), (2.6)–(2.9) after rotations of 2–qubit POVMs from the device by 2 local unitaries optimizing the POVM's a) worst–case (2.3), b) average–case (2.4) distance from its **dephased** version. 150 pairs with biggest reductions are plotted. The qubit pairs in the left and right column are the same. A blue bar signifies reduction of the function and yellow its increase. Detailed legend is shown in Figure 3.2.

(a) Optimization of $\mathrm{D}_{\mathrm{op}}^{\mathrm{comp}}(U\mathsf{M}U^{\dagger})$:



(b) Optimization of $\mathrm{D}_{\mathrm{avg}}^{\mathrm{comp}}(U\mathsf{M}U^{\dagger})$:

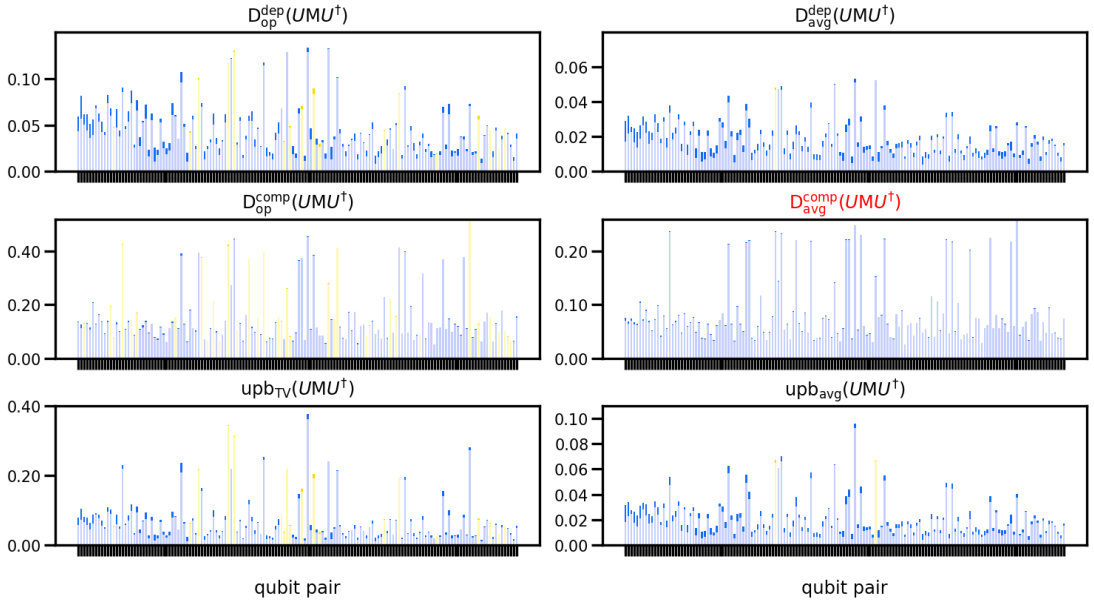
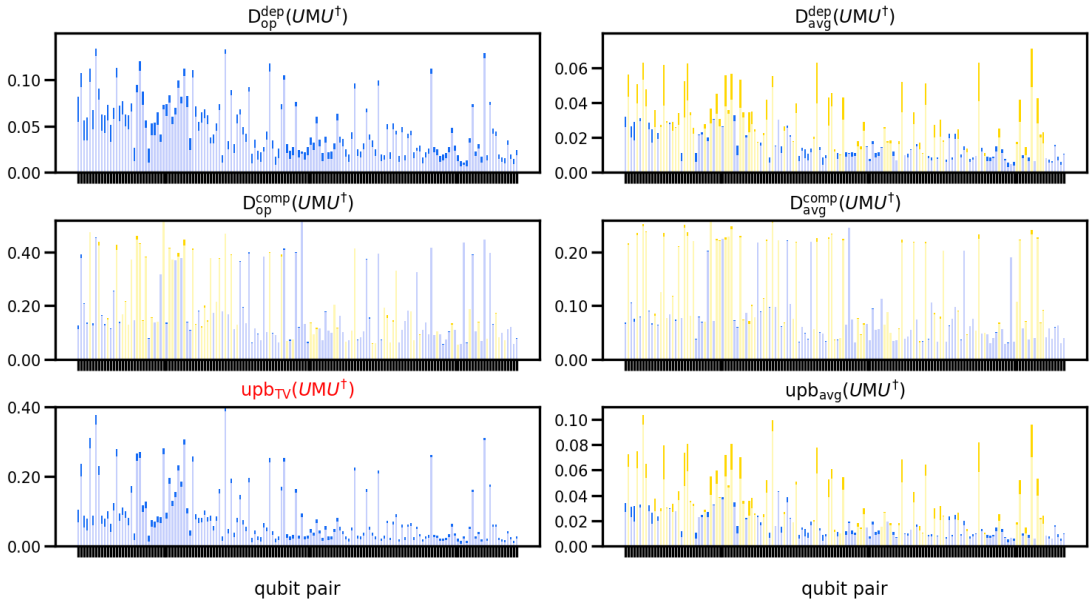


Figure A.6: **ASPEN–M–1**. Changes in values of cost functions (2.3)–(2.4), (2.6)–(2.9) after rotations of 2–qubit POVMs from the device by 2 local unitaries optimizing the POVM's a) worst–case (2.8), b) average–case (2.9) distance from **computational basis** POVM. 150 pairs with biggest reductions are plotted. The qubit pairs in the left and right column are the same. A blue bar signifies reduction of the function and yellow its increase. Detailed legend is shown in Figure 3.2.

(a) Optimization of $\mathrm{upb}_{\mathrm{TV}}(U\mathsf{M}U^{\dagger})$:



(b) Optimization of $\mathrm{upb}_{\mathrm{avg}}(U\mathsf{M}U^{\dagger})$:

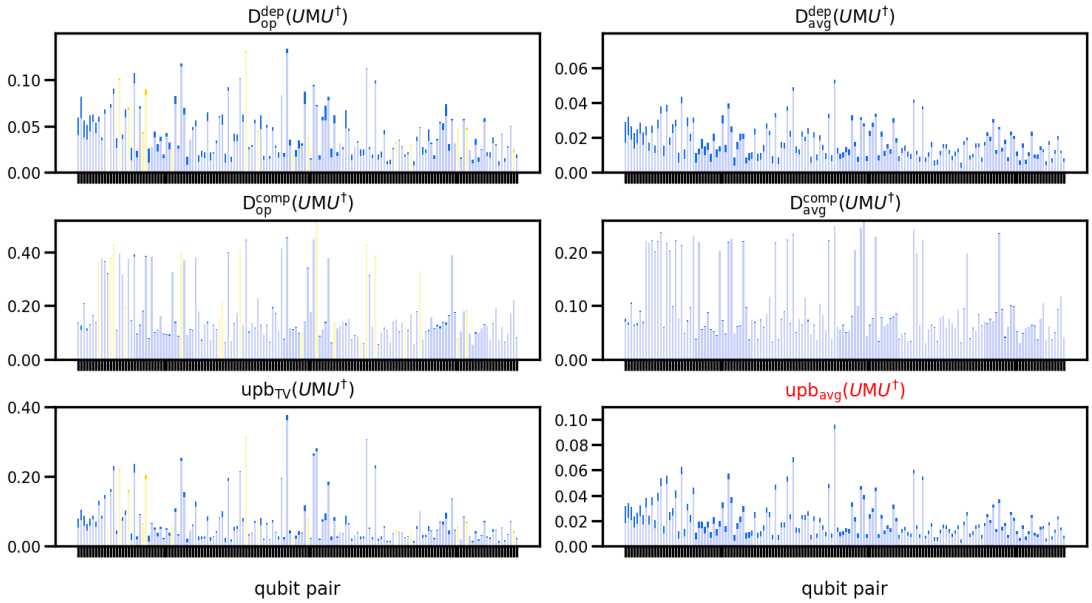
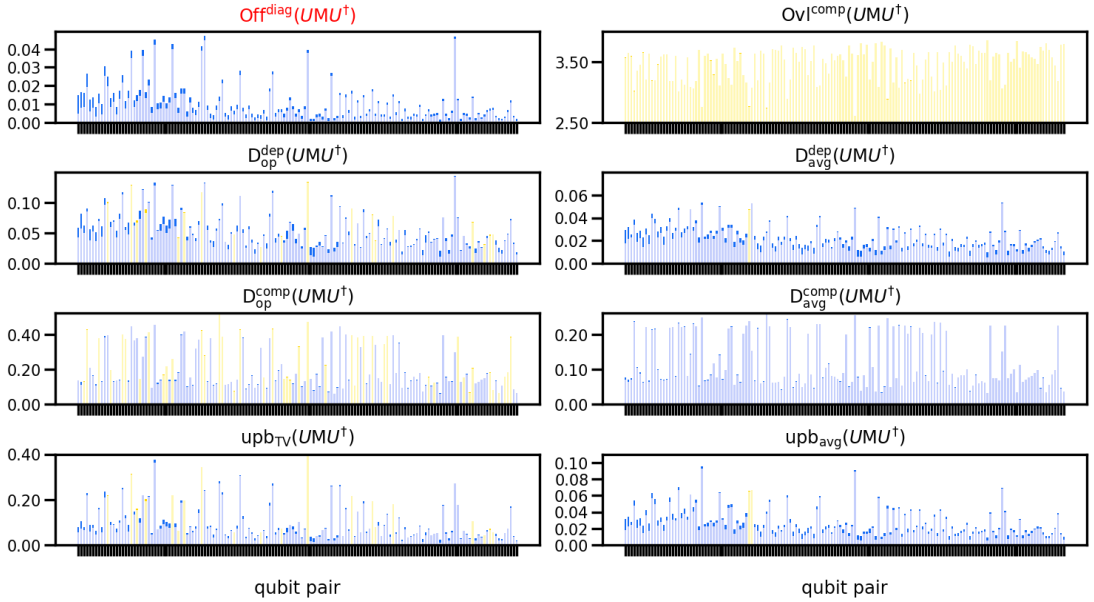


Figure A.7: **ASPEN–M–1**. Changes in values of cost functions (2.3)–(2.4), (2.6)–(2.9) after rotations of 2–qubit POVMs from the device by 2 local unitaries optimizing the upper–bound on a) worst–case (2.6), b) average–case (2.7) TV distance between ideal probability vector and that resulting from mitigation using method from Section 1.6.2. 150 pairs with biggest reductions are plotted. The qubit pairs in the left and right column are the same. A blue bar signifies reduction of the function and yellow its increase. Detailed legend is shown in Figure 3.2.

(a) Optimization of $\mathrm{Off}^{\mathrm{diag}}(U\mathsf{M}U^\dagger)$:



(b) Optimization of $\mathrm{Ovl}^{\mathrm{comp}}(U\mathsf{M}U^\dagger)$:

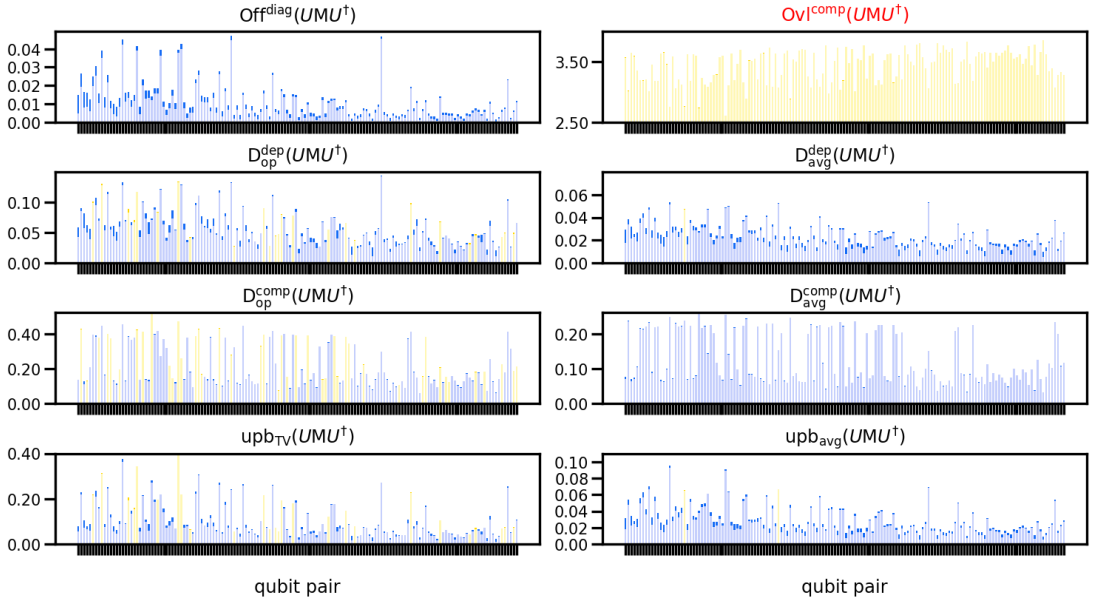


Figure A.8: **ASPEN–M–1**. Changes in values of cost functions (2.3)–(2.10) after rotations of 2–qubit POVMs from the device by 2 local unitaries optimizing a) $\mathrm{Off}^{\mathrm{diag}}(U\mathsf{M}U^\dagger)$ (2.5), b) $\mathrm{Ovl}^{\mathrm{comp}}(U\mathsf{M}U^\dagger)$ (2.10). 150 pairs with biggest reductions are plotted. The qubit pairs in the left and right column are the same. A blue bar signifies reduction of the function and yellow its increase. Detailed legend is shown in Figure 3.2. Note that the vertical axis for $\mathrm{Ovl}^{\mathrm{comp}}(U\mathsf{M}U^\dagger)$ starts from 2.5.