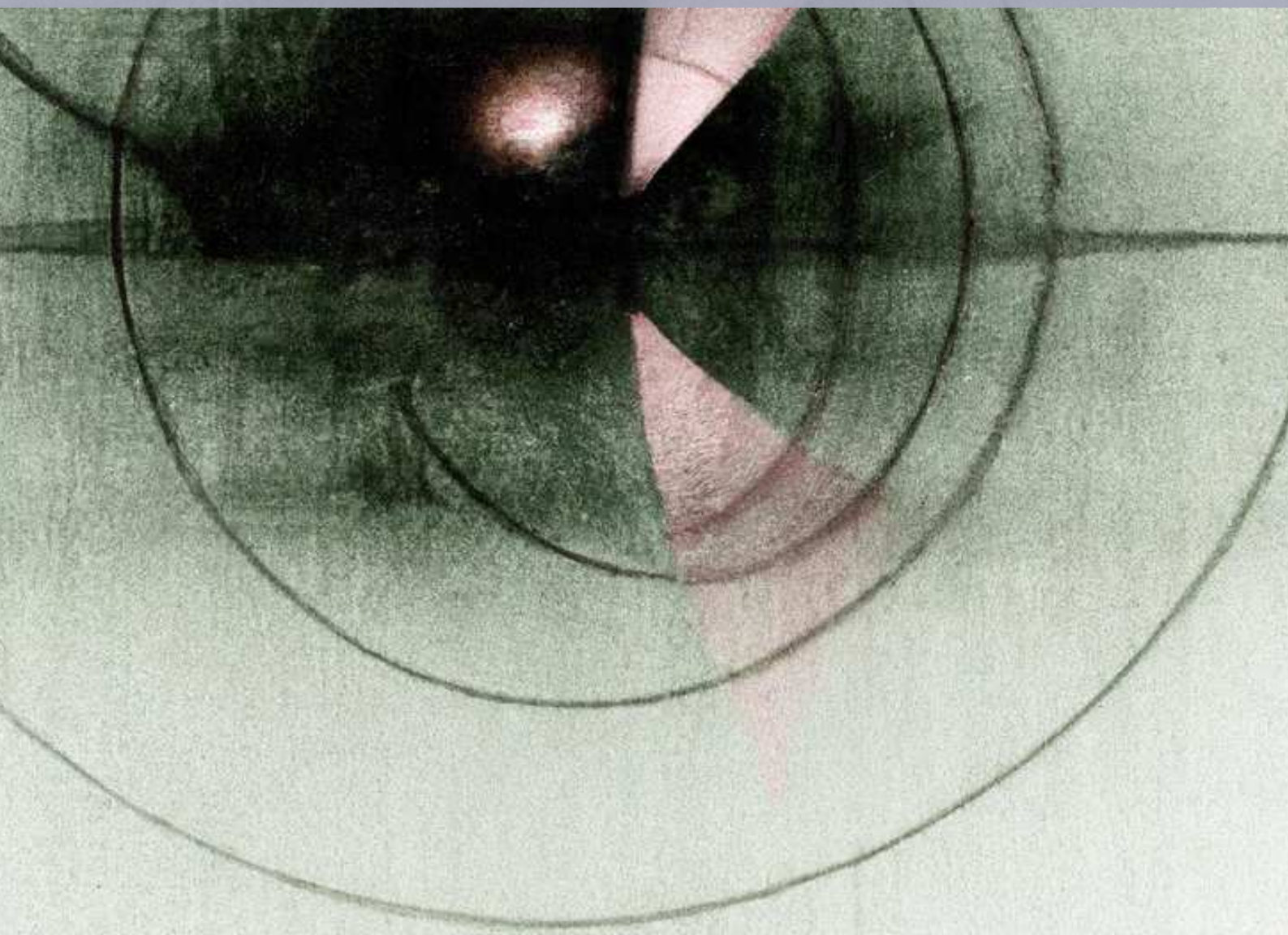
Applications of classical randomness on near term quantum computers

PhD thesis

Filip B. Maciejewski



IITiS

Applications of classical randomness on near term quantum computers

A thesis submitted in partial fulfillment
of the requirements for the degree of

PhD

in

Computer Science

Author: Filip B. Maciejewski

Supervisors: Dr Z. Puchała and Dr M. Oszmaniec

Cover image: "A noisy quantum measuring device in a style of Zdzislaw Beksinski" by DALL·E 2

Template style: Thesis style by Richelle F. van Capelleveen

IITiS PAS

INSTITUTE OF THEORETICAL AND APPLIED INFORMATICS

Baltycka 5, 44-100 Gliwice, Poland

Abstract in English

Implementing measurements on contemporary quantum computing devices is a considerable challenge. This is primarily due to pervasive noise effects and the constraints imposed by limited quantum resources such as the number of qubits and available unitary operations. However, in a wide variety of use cases, the complications engendered by these challenges can be reduced through the application of supplementary classical resources (randomness and post-processing). The principal hypothesis advanced in this dissertation contends that

> Auxiliary classical resources can be used to assess and improve the quality of the implementation of noisy quantum measurements

We offer contributions that aim to help better understand the effects of measurement noise on the performance of quantum-information protocols, as well as methods that reduce those effects. In Chapter 3 we introduce new distance measures between quantum objects (in particular, quantum measurements). We call those measures quantum average-case distances (AC distances). AC distance between a noisy measurement and its ideal model allows for quantification of the average-case performance of a noisy detector. The classical resource exploited in this context is classical *randomness*. Specifically, the AC distance between two measurements quantifies how well they can be statistically distinguished if we are given access to the random application of a certain class of random quantum circuits (circuits forming unitary 4-designs).

While the AC distance can be used to quantify the quality of the measurement's implementation, it does not provide an explicit procedure to characterize a given noisy measurement. We tackle this problem in Chapter 4, where we provide an efficient method of reconstructing certain types of local measurement noise from experimental data. The proposed method, Diagonal Detector Overlapping Tomography (DDOT), also exploits classical randomness. Moreover, we show how to use the results of DDOT to reduce the effects of readout noise on the estimation of marginal probability distributions. This is relevant, for example, in the context of the estimation of energy of local Hamiltonians. Our error-mitigation methods are classical in the sense that they are done entirely in *post-processing* of the experimental data obtained in noisy experiments.

Many of the results across the Thesis are supported by extensive numerical simulations. Furthermore, we present results of experimental implementation

of noise characterization and mitigation on superconducting quantum hardware from IBM and Rigetti (Chapter 4).

The thesis consists of the following works:

[1] *Operational Quantum Average-Case Distances*, F.B. Maciejewski, Z. Puchała, M. Oszmaniec, Quantum 7, 1106 (2023).

[2] *Exploring Quantum Average-Case Distances: Proofs, Properties, and Examples*, F.B. Maciejewski, Z. Puchała, M. Oszmaniec, IEEE Transactions on Information Theory, vol. 69, no. 7, pp. 4600-4619 (2023).

[3] *Modeling and mitigation of cross-talk effects in readout noise with applications to the Quantum Approximate Optimization Algorithm*, F.B. Maciejewski, F. Baccari, Z. Zimborás, M. Oszmaniec, Quantum 5, 464 (2021).

Abstract in Polish

Implementacja pomiarów na współczesnych urządzeniach kwantowych stanowi znaczące wyzwanie. Jest to w dużej mierze spowodowane wszechobecnym wpływem szumu oraz ograniczeniami nałożonymi przez ograniczone zasoby kwantowe, takie jak liczba qubitów i dostępne operacje unitarne. Jednakże, w szerokim zakresie sytuacji, komplikacje wywołane przez te wyzwania mogą zostać zredukowane poprzez zastosowanie dodatkowych klasycznych zasobów (losowości oraz post-processingu). Główna hipoteza przedstawiona w tej dysertacji brzmi

> Dodatkowe zasoby klasyczne mogą być użyte do oceny i poprawy jakości implementacji zaszumionych pomiarów kwantowych

Prezentujemy wyniki, które mają na celu lepsze zrozumienie wpływu szumu pomiarowego na jakość implementacji protokołów informacji kwantowej, a także metody redukcji tych efektów. W Rozdziale 3 wprowadzamy nowe miary dystansu między obiektami kwantowymi (w szczególności, pomiarami kwantowymi). Nazywamy te miary kwantowymi *average-case distances* (odległości lub dystanse AC). Dystans AC między zaszumionym pomiarem a jego idealną, teoretyczną wersją, pozwala na kwantyfikację jakości nieidalnego detektora w średnim (w sensie statystycznym) przypadku (stąd "average-case", czyli "średni przypadek/sytuacja"). Klasycznym zasobem wykorzystanym w tym kontekście jest *losowość*. W szczególności, odległość AC między dwoma pomiarami określa jak dobrze mogą być one statystycznie rozróżnione, jeśli mamy dostęp do losowej implementacji pewnej klasy losowych obwodów kwantowych (obwody tworzące tzw. "unitary 4-designs").

Mimo że odległość AC może być użyta do kwantyfikacji jakości implementacji pomiaru, nie dostarcza ona konkretnego protokołu jak scharakteryzować dany nieidealny pomiar. Częściowe rozwiązanie tego problemu proponujemy w Rozdziale 4, gdzie przedstawiamy wydajną metodę rekonstrukcji pewnych typów lokalnego szumu pomiarowego z danych eksperymentalnych. Proponowana metoda, *Diagonal Detector Overlapping Tomography* (DDOT), wykorzystuje klasyczną losowość. Co więcej, pokazujemy jak użyć wyników DDOT do redukcji efektów szumu pomiarowego na estymację marginalnych rozkładów prawdopodobieństwa. Jest to istotne, na przykład, w kontekście estymacji energii lokalnych Hamiltonianów. Nasze metody redukcji błędów są klasyczne w tym sensie, że są

wykonane całkowicie w *post-processingu* danych eksperymentalnych uzyskanych w nieidealnych eksperymentach.

Wiele wyników w całej dysertacji jest wspartych przez obszerne symulacje numeryczne. Ponadto, prezentujemy wyniki eksperymentalnej implementacji charakteryzacji i mitygacji szumów na nadprzewodzących urządzeniach kwantowych od IBM i Rigetti (Chapter 4).

Dysertacja składa się z następujących prac:

[1] *Operational Quantum Average-Case Distances*, F.B. Maciejewski, Z. Puchała, M. Oszmaniec, Quantum 7, 1106 (2023).

[2] *Exploring Quantum Average-Case Distances: Proofs, Properties, and Examples*, F.B. Maciejewski, Z. Puchała, M. Oszmaniec, IEEE Transactions on Information Theory, vol. 69, no. 7, pp. 4600-4619 (2023).

[3] *Modeling and mitigation of cross-talk effects in readout noise with applications to the Quantum Approximate Optimization Algorithm*, F.B. Maciejewski, F. Baccari, Z. Zimborás, M. Oszmaniec, Quantum 5, 464 (2021).

Contents

1. Introduction

Since the dawn of civilization, humanity has been driven by a never-ending quest to simplify complex calculations. Whether it was the early merchants of Mesopotamia employing the abacus to keep track of trade, or astronomers in Ancient Greece using geometric methods to predict celestial movements, the necessity to facilitate easier and more efficient computation has been a recurrent theme throughout history [4; 5]. In the 20th century, the invention of the electronic computer marked a monumental leap in our computational capabilities. Gone were the days of laborious manual calculations. Scientists and engineers harnessed the unprecedented power of classical computers to tackle increasingly complex problems, from simulating weather patterns to decoding the human genome.

It's not an exaggeration to state that classical computing has, to a large degree, shaped the world we inhabit today. A glance at modern society reveals the ubiquitous influence of computing technology. Computers have transcended the realm of specialized scientific computation to become an omnipresent commonality in daily life. Our personal computers and smartphones are descendants of those early electronic machines, albeit miniaturized and exponentially more powerful. They serve as gateways to an interconnected world, enabling instantaneous communication, access to vast repositories of human knowledge, and the capacity for social interaction on an unprecedented scale.

Indeed, the advent of classical computing has propelled us into the realm of what once seemed like science fiction. Perhaps one of the most important inventions stemming from classical computing technology was announced a little bit over a year two years prior to (finishing) the writing of this thesis – ChatGPT, a chat-bot technology based on large language models technology [6]. For the first time since the beginning of the human race, we can have a meaningful conversation with an entity that is not a member of our species. Despite multiple pitfalls, including the fact that the conversations are not always really meaningful, it is, arguably, a mark of the new era in the history of technology and perhaps, civilization in general.

While classical computing is undoubtedly a very interesting topic, this thesis is concerned with a conceptually distinct sub-field of information technologies – quantum information science. Quantum computing, which can be considered a subfield of general quantum information most relevant to this Thesis, is an en-

deavor with essentially the same overarching goal as classical computing – that is, arguably, to make our lives better in the long term. However, it aims to do so by using a set of very different (perhaps richer and more fascinating) set of tools. The general hope of the scientists and engineers working in this field is that some of the problems that are in practice intractable using known classical methods might become feasible with sufficiently powerful quantum computers. While there has been no convincing demonstration of useful quantum computation performed in experiments as of the writing of this Thesis, it is not impossible that it will happen in our lifetimes (which we certainly hope for). At the same time, even if it never happens, it would certainly still be worth pursuing. Indeed, quantum information research has already been very fruitful, improving our understanding of fundamental problems, such as the physics of information processing [7; 8; 9; 10], the hardness of finding ground states of local Hamiltonians [11; 12; 13], the philosophy of quantum mechanics [14; 15; 16], and more [17; 5]. In what follows we briefly outline the basics of classical and quantum computing, and situate the contents of this Thesis within the broader context of quantum information science.

1.1 Overview of classical and quantum computing

The fundamental concept of classical computing is a bit (short for binary digit) – an abstraction for a basic unit of information that can exist in one of two states conventionally denoted by 0 and 1. On a very high level, classical computing can be viewed as a study of how to manipulate bits to perform calculations. The textbook model for classical computation is a Turing machine [17]. A Turing machine consists of an infinite tape with cells containing symbols; a tape head that writes and reads from the tape, as well as has its own internal state; and a set of rules specifying the head's actions based on the current state of the tape's head and state on the tape to which the head is pointing. The original motivation of Turing was to come up with a notion of a "definite procedure" that can represent, in a rigorous manner, a set of steps to arrive at a solution of a given problem [18]. While its historical relevance cannot be overstated (as it laid the foundations for modern computer science), it is also a fairly academic model that can be hard to operate with in practice.

An equivalent, useful representation of classical computation makes use of classical circuits. A classical circuit is a diagram representation where multiple wires represent bits, and various connectors between wires represent *gates* that transform input bits into output bits. In particular, a (classical) logic gate is a function from $\{0,1\}^k$ to $\{0,1\}^l$ for $k \geq 1$, $l \geq 1$. Consider the NAND logic gate that returns 0 only if both inputs are 1, and returns 1 otherwise. It turns out that the NAND gate is universal for classical computation, meaning any $\{0,1\}^k \to \{0,1\}^l$ function can be implemented using circuits containing only that gate (naturally, different universal sets of logic gates exist). Importantly, an implicit assumption here is that we can actually produce a description of such circuit in reasonable (polynomial) time. Here, one typically considers so-called uniform circuit fam-

ilies, meaning, roughly speaking, circuits that are an output of some algorithm evaluated, e.g., on a Turing machine (and then focus on algorithms that produce the circuit description in polynomial time). While exact details of defining/constructing such families are not important for us, it is worth noting that bounding the size of circuits needed to compute some Boolean function can be used to prove separation of computational complexity classes [5].

At the same time, the general idea of encoding a computation into a circuit model is of particular relevance in the context of this thesis, as it was inherited by the field of quantum information. The basic representation of the so-called "gate-based" quantum computing paradigm uses *quantum* circuits [17]. Many differences between classical and quantum computing are well illustrated by studying the differences between the circuit representations of the two. The diagram representing a quantum circuit has the same structure as a classical circuit, but now wires correspond to qubits. A qubit (short for a quantum bit) is a fundamental concept of quantum information. Formally, in a noiseless computation model, it is represented by a unit vector living in a two-dimensional vector space equipped with an inner product (a Hilbert space). We therefore see that quantum computing makes a significant conceptual jump by replacing binary digits with continuously parametrized vectors. The input-output operations between qubits are now quantum gates that are represented with unitary operators – bounded operators on the Hilbert space that preserve the inner product; physically, this represents a time evolution of a quantum system (generated by a Hamiltonian describing interactions within the system and between systems via the Schrödinger equation) Finally, of particular importance in the context of this Thesis is the fact that the quantum circuit ends in *quantum measurement*. Conceptually, a quantum measurement can be thought of as a mapping from quantum states to classical outcomes (formally, this particular concept is represented by so-called quantum instruments, and we forget here about the post-measurement states). The classical outcomes are represented, as in the case of classical computing, via strings of symbols, usually bitstrings.

Indeed, while the processing of quantum information (the "computing" part) happens in the Hilbert space by means of the transformation of many-qubit states via quantum gates, the actual outcome of the computation is classical, and accessed via a quantum measurement (the "readout" part). The famous difference between classical and quantum readout is that, in general, the quantum measurement outcomes are probabilistic. This hints at the general difficulty of designing quantum algorithms, where one needs to find ways of transforming *probabilities* (or more strictly, probability amplitudes) in a way that will (most likely) lead to an answer to the problem.

Quantum computers are thus machines comprised of qubits, which facilitate the application of controllable unitary operations among these qubits, and enable the extraction of classical information through quantum measurements. From this perspective, quantum computers are devices that allow for sampling from probability distributions across (typically exponentially large) sets of possible outcomes. The overarching task of a quantum algorithm (or of the individuals

developing such algorithms) is, broadly speaking, to engineer ways of obtaining a probability distribution that is, vaguely speaking, concentrated around the solution to the problem at hand.

An important obstacle that prevents the existing prototypes of quantum computers (or, as we will often refer to them, the quantum devices) from performing practically useful calculations is noise. In the case of classical computing, when the bit-encoded information is physically transformed and stored, errors may occur. The only possible states of bits are 0 and 1, thus the classical error to consider is generally just a bitflip (a NOT gate; and we are disregarding here the potential loss of bits). The task of preventing such classical errors from destroying the computation is the subject of classical error correction [17; 19]. Roughly speaking, classical error correction uses *redundancy* to encode a single *logical* bit into multiple *physical* bits, by means of error-correcting *codes*. Quantum error correction (QEC) uses similar notions (replacing bits with qubits), but now the errors (and error-correcting codes) are more general because qubits are more complex objects than bits. Fortunately, it turns out that the errors are not infinitely more complicated (as might naively be expected when considering the transition from classical binary digits to continuous quantum qubits) due to the digitization of errors, i.e., the possibility to represent general quantum errors by insertions of simple types of errors [20; 19]. One of the most important results of quantum information theory is the threshold theorem that, roughly speaking, asserts that if we can keep the physical error rate below some threshold value, we can effectively suppress the logical errors arbitrarily by means of QEC.

1.2 Thesis context

While there are many possible definitions, let us say that the task of "quantum advantage" corresponds to implementing a useful computation on quantum hardware faster than any existing classical algorithm. While there have been reports suggesting the attainment of quantum advantage in specialized tasks – most notably in the domain of random quantum circuit sampling (see, e.g., [21]) – the field is yet to witness a compelling demonstration of quantum computing's utility in addressing problems of practical significance.

This is the context in which the research for this thesis was performed – quantum devices have indeed transitioned from theoretical curiosity to tangible prototypes. Yet those prototypes are moderate in size, the operations are very noisy, and currently, no one has implemented a useful quantum computing task. At the same time, existing quantum devices offer an incredible level of control for quantum systems. What is particularly striking is the unprecedented ease of access to the actual technology. Indeed, a lot of quantum computing technology was developed in the past decade – and it seems that engineers embraced the automation of classical control of those devices. For many existing architectures developed by such manufacturers as IBM [22], Google [23], Rigetti [24], IonQ [25], Quantinuum [26], all one needs to perform actual experiments is writing some Python code (and, usually, money to pay for it – with a commendable

exception of IBM's small size devices that are fully open-access for everyone who registers [22]).

While in the long run, the threshold theorem provides theoretical justification for the feasibility of fault-tolerant quantum computation, reaching the requisite error thresholds at scale remains a formidable engineering challenge [27]. Multiple estimates on quantum error correction suggest we would need millions of physical qubits to perform useful computation (there has been, however, a lot of recent progress in that field, see, e.g., [28]). Where exactly are we today? The biggest state-of-the-art quantum devices consist usually of hundreds of qubits (some going to thousands) that are often very noisy. Since the technology is currently nascent, the "quantum resources" (i.e., qubits, allowed quantum gates, etc.) tend to be limited. This stage of development is usually referred to as Noisy Intermediate-Scale Quantum (NISQ) devices [29]. However, the dawn of the (very-) early fault-tolerant era (tens of error-corrected qubits) seems to be on the horizon. Indeed, a cautious optimism is certainly warranted in light of the enormous progress in small-scale experimental demonstrations of quantum error correction subroutines in recent years – see, e.g., [30; 31; 32; 33]. Notably, in Ref. [33] (published in second half of 2024), the Google's team achieved, among other results, the milestone of extending the lifetime of a logical qubit by a factor of over 2 compared to the underlying physical qubits .

The past few years have seen a rapid development of an intermediate approach to the problem of noise – quantum error mitigation (QEM) [34]. On the one hand, the QEC usually aims to correct errors in real time to reduce (or effectively eliminate) logical error rates thanks to additional resources that include, most notably, a lot of redundant qubits. On the other hand, the QEM is typically concerned with constructing noise-reduced *estimators* of quantities of interest (weak error-mitigation) or sampling from noise-reduced probability distributions (strong error-mitigation) [35], exploiting very limited amounts of auxiliary quantum resources (and plenty of cheap classical resources). It is worth noting that even though the QEM subfield has been developing mainly in the context of near-term, noisy devices, the general insights and methods related to noise characterization and mitigation will likely prove useful also in longer timescales. Indeed, it is conceivable that the first useful quantum computation demonstrations (if ever achieved) will be possible thanks to the application of some type of hybrid QEM/QEC techniques (see, for example, Ref. [36]).

While QEM can be achieved in multiple ways [34], in this Thesis we focus mainly on the application of additional *classical* resources to reduce errors. In Chapter 4, based on Ref. [3], we present a method to efficiently characterize and reduce the effects of the measurement noise on various estimated quantities. We introduce Diagonal Detector Overlapping Tomography (DDOT), a method that allows to efficiently characterize a certain type of readout noise using $O(\log(n))$ (with n being system size) calibration circuits. The classical resource exploited in DDOT is classical *randomness* – a uniform random sampling from a set of very simple quantum circuits is shown to, with high probability, allow characterization of locally correlated measurement errors. The DDOT results can be used to cor-

rect estimators of marginal probability distributions on local subsets of qubits – in this case, the exploited classical resource is *post-processing*. Indeed, based on the DDOT calibration results, we show how to reduce noise effects on the estimators of marginals using very simple post-processing of the experimental results.

In the earlier Chapter 3, based on Refs. [1] and [2], we show that the classical *randomness* can play an important role in quantifying the quality of implementation of a given measuring device. In that Chapter, we introduce a novel distance measure between quantum measurements (we also consider quantum states and general quantum channels). This measure, the quantum average-case distance (quantum AC distance) between an ideal (theoretical) measurement model and its noisy (experimental) implementation, quantifies the average-case performance of a noisy detector. In particular, the AC distance between two measurements quantifies how well they can be statistically distinguished if we are given access to the random application of a certain class of random quantum circuits (circuits forming an approximate unitary 4-design).

Based on the above, we can phrase the principal thesis of this dissertation

> Auxiliary classical resources can be used to assess and improve the quality of the implementation of noisy quantum measurements

It is important to note that the ideas revolving around exploiting classical resources to aid quantum computation are prevalent across multiple sub-fields of quantum information science. As indicated previously, parts of the quantum error mitigation can, risking an oversimplification, be viewed as a study of how much we can do given access to very limited additional quantum resources (usually one considers gathering more samples, and rarely additional qubits), and a lot of additional classical resources (such as cheap classical randomness and usually cheap post-processing). This provides the context for Chapter 4 about readout error mitigation. On the other hand, we have Chapter 3 that introduces distance measures based, roughly speaking, on randomized measurement protocols (note that also DDOT in Chapter 4 is based on an instance of such protocols). This is in line with recent research directions that exploit randomized measurements to perform efficient estimation (see, e.g., [37]); but also with so-called quantum resource theories, where often classical operations are considered a free resource and one studies a general structure of quantum sets augmented with them (see, e.g., [38] in the context of quantum measurements).

1.3 Structure of the work

The research contained in this thesis revolves around the topic of making use of the limited available quantum resources by augmenting them with cheap classical resources. We focus most of our attention on a particular part of every quantum algorithm – quantum measurements. In Chapter 2, we provide a brief description of the necessary theoretical background used later. The main body of the thesis consists of three published works [1; 2; 3] that are presented in

two main chapters. In Chapter 3, based on Refs. [1] and [2], we fill the void in the distance measures' literature to quantify the average-case performance of measurements (but also states and channels) instead of the worst-case that was typically considered. The average-case distances that we introduce correspond to scenarios that exploit classical randomness in the implementation of quantum algorithms. Then, in Chapter 4, we continue with studying noisy measurements and introduce a new, scalable method for characterizing readout noise and reducing its effects on the estimation of local quantum observables. The method is tested both numerically and experimentally, and we report significant improvements in the quality of the results. That chapter is based on Ref. [3]. We conclude the Thesis with a summary and future research outlook in Chapter 5.

2. Theoretical background

2.1 Structure of the Chapter

In this Chapter, we present a very brief overview of the most relevant, in the context of the main body of the Thesis, concepts from quantum information science. This includes discussing the basics of a quantum-mechanical description of physical systems in Section 2.2, followed in Section 2.3 by an overview of noise modeling and noise mitigation. The emphasis is put on noisy measurements, especially relevant in the context of Chapter 4 that present methods for characterization and reduction of readout noise. In Section 2.4, we discuss the commonly used distance measures between states, measurements, and channels. The worst-case nature of those measures is in contrast to the average-case distances that are introduced in Chapter 3. Various proofs in Chapter 3 use the concept of unitary designs, which are briefly described in Section 2.5. The Chapter contains major excerpts from works [1; 2; 3].

2.2 Basics of quantum mechanics

We start by recalling basic quantum-mechanical concepts used throughout the thesis [39; 17]. We will be interested in finite-dimensional physical systems. According to postulates of quantum mechanics, a full description of any d-dimensional system can be obtained by associating with it a d-dimensional Hilbert space $\mathcal{H}_d \approx \mathbb{C}^d$. The state of the system is then described by a quantum state ρ defined as a positive-semi-definite operator with unit trace. The evolution of a closed quantum system is described by a unitary U generated by system Hamiltonian $\mathcal{H}$ via $\exp(-it\mathcal{H})$ (where $t \geq 0$). A general quantum measurement $\mathbf{M}$ is represented by a Positive Operator-Valued Measure (POVM) [39]. A POVM $\mathbf{M}$ with n outcomes on a $\mathcal{H}_d$ is a set of positive-semidefinite operators that sum up to identity, i.e.,

$$\mathbf{M} = \{M_i\}_i^n \,, \qquad \forall i \; M_i \geqslant 0 \,, \qquad \sum_{i=1}^{n} M_i = \mathbb{I}, \tag{2.1}$$

where $\mathbb{I}$ is the identity operator.

In a quantum computing context, a perfect measurement is often modeled as a *projective* measurement $\mathbf{P} = \{P_i\}_i^r$ for which the measurement operators, in addition to the requirements from Eq. (2.1), are also projectors, i.e., $\forall_i P_i^2 = P_i$. An important example of a measurement that will be useful throughout the thesis is a computational-basis measurement defined as $\mathbf{M}^{\text{comp}} = (|1\rangle\langle 1|, \ldots, |\mathrm{d}\rangle\langle \mathrm{d}|)$ (sometimes denoted simply as $\mathbf{P}$ if it is clear from the context what projective measurement we mean).

2.3 Noise modeling and mitigation

2.3.1 Channels

When a physical system of interest is not isolated from the environment, its evolution will, in general, not be governed by unitary evolution. Instead, it can be described by a quantum channel Λ, defined as a linear CPTP (Completely-Positive Trace-Preserving) map [17]. Trace-preserving condition means that for any quantum state $\rho, \Lambda(\rho) \in \mathsf{D}(\mathcal{H}_\mathrm{d})$, we have $\mathrm{tr}(\Lambda(\rho)) = \mathrm{tr}(\rho)$. Complete-positivity means that $(\Lambda \otimes \mathcal{I}_{\mathrm{d}'})\,\tilde{\rho} \geq 0$ for any d' and any $\tilde{\rho} \in \mathsf{D}(\mathcal{H}_{\mathrm{dd}'})$, where $\mathcal{I}_{\mathrm{d}'}$ denotes identity channel on $\mathcal{H}_{\mathrm{d}'}$. Quantum channel Λ is in one-to-one correspondence to a Choi-Jamiołkowski state defined as $\mathcal{J}_\Lambda := (\mathcal{I}_\mathrm{d} \otimes \Lambda)(|\Phi^+\rangle\langle\Phi^+|)$, where we extend Hilbert space by its copy and act with channel Λ on a half of the maximally entangled state $|\Phi^+\rangle := \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |ii\rangle$. We denote the set of all quantum channels from $\mathcal{H}$ to itself as $\mathsf{CPTP}(\mathcal{H}_\mathrm{d})$. A *unital* quantum channels is a channel $\Phi \in \mathsf{CPTP}(\mathcal{H}_\mathrm{d})$ such that $\Phi(\tau_\mathrm{d}) = \tau_\mathrm{d}$, where τ_d is the maximally mixed state in $\mathcal{H}_\mathrm{d}$.

Finally, when a quantum state $\rho \in \mathsf{D}(\mathcal{H}_\mathrm{d})$ undergoes process $\Lambda \in \mathsf{CPTP}(\mathcal{H}_\mathrm{d})$ followed by measurement described by POVM $\mathbf{M} \in \mathsf{P}(\mathcal{H}_\mathrm{d}, n)$, the probability of outcome labeled as "i" is given by Born's rule

$$\Pr(i|\rho, \Lambda, \mathbf{M}) = \mathrm{tr}(\Lambda(\rho) M_i) \ . \tag{2.2}$$

Throughout the thesis, if Λ is an identity channel, we will omit it in the above notation.

2.3.2 Readout noise

Quantum measurements can be represented as a special-case of a quantum channel (quantum instruments), thus the relationship between an ideal measurement $\mathbf{P}$ and its noisy implementation $\mathbf{M}$, in general requires considering transformations between quantum channels (or superchannels/supermaps [40]). However, recently it has been experimentally demonstrated for superconducting quantum devices that in practice this relationship, to a good approximation, can be represented as a stochastic map [41; 42], which we will refer to as 'classical measurement noise'. In such a model, the relation between $\mathbf{M}$ and $\mathbf{P}$ is given by some stochastic transformation T. Namely, we have $\mathbf{M} = \mathsf{T}\mathbf{P}$, i.e., $M_i = \sum_j \mathsf{T}_{ij} P_j$. Due to the linearity of Born's rule, it follows that probabilities $\mathbf{p}^{noisy}$ from which

noisy detector samples are related to the noiseless probabilities $\mathbf{p}^{ideal}$ via the same stochastic map, hence [42]

$$\mathbf{p}^{noisy} = \mathsf{T}\mathbf{p}^{ideal} . \tag{2.3}$$

Specifically, in the convention where probability vectors are columns, the noise matrix T is left-stochastic, meaning that each of its columns contains non-negative numbers that sum up to 1. Such noise is thus equivalent to a stochastic process, in which an outcome from a perfect device probabilistically changes to another (possibly erroneous) one.

As mentioned above, Eq. (2.3) does not present the most general model of quantum measurement noise. Specifically, *coherent* errors might occur, and reduce the effectiveness of error-mitigation (a detailed analysis of this effect was presented in [42]). This type of noise will not be analyzed in this thesis.

Note that in the above discussion, we denoted ideal measurement as $\mathbf{P}$, indicating that in many quantum-information protocols the model for ideal measurement is indeed some type of projective measurement.

2.3.3 Noise mitigation

We now note that the equation (2.3) suggests a simple way to mitigate errors on the noisy device – via left-multiplying the estimated statistics by the inverse of noise matrix T^{-1} [41; 42]. From the stochasticity of T it follows that its inverse does preserve the sum of the elements of probability vectors, however it may introduce some unphysical (i.e., lower than 0 or higher than 1) terms in the corrected vector. A common practice in such a scenario is to solve an optimization problem

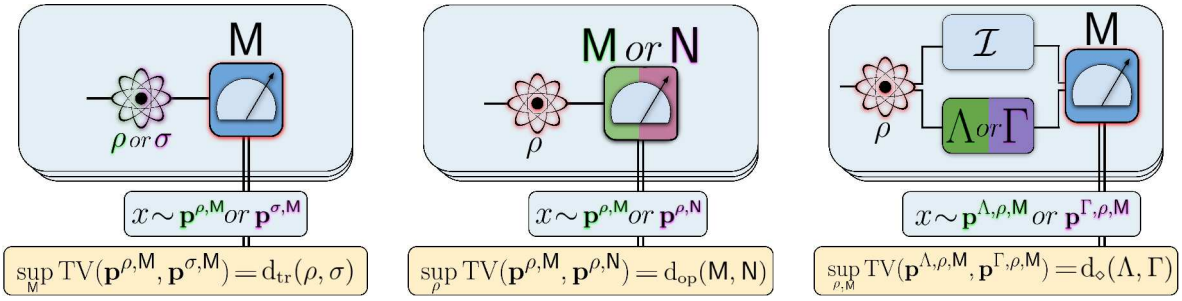
$$\mathbf{p} = \mathrm{argmin}_{\mathbf{q}} ||\mathbf{q} - \mathsf{T}^{-1}\mathbf{p}^{\mathrm{noisy}}||_2^2 , \tag{2.4}$$

where minimization goes over all proper probability distributions. This introduces additional errors in the final estimations which can be upper-bounded [42].

We note that the task of error mitigation is more general than just correcting noisy measurements and there exist a plethora of methods to obtain either samples from corrected probability distributions ("strong error-mitigation" [35]), or construct error-mitigated estimators of observables ("weak error-mitigation") – see, e.g., [34] for an overview of various techniques. In the context of this thesis, the above simple readout error mitigation is of interest and will be expanded upon in Chapter 4.

2.4 Worst-case distance measures

Having two quantum objects – for example, ideal implementation of a quantum measurement, and its noisy, experimental version – it is useful to have a way to quantify their (dis)similarity. Having this in mind, we will now discuss commonly used distance measures.



(a) *Trace distance between quantum states.*

(b) *Operational distance between POVMs.*

(c) *Diamond distance between quantum channels.*

Figure 2.1: *Depiction of measures of distance between quantum objects based on optimal statistical distinguishability – which can be also interpreted as "worst-case" distance. For quantum states (2.1a), we optimize over all POVMs, while for measurements (2.1b) we optimize over all states. For quantum channels (2.1c) we optimize over both states and measurements on the extended Hilbert space.*

The distances of interest will be induced by the following norms. Denote by $L(\mathcal{H}_\mathrm{d})$ a space of linear operators on $\mathcal{H}_\mathrm{d}$. Then for $A \in L(\mathcal{H}_\mathrm{d})$, the trace norm is defined as

$$||A||_1 = \mathrm{tr}\left(\sqrt{AA^\dagger}\right) . \tag{2.5}$$

For a channel $\Lambda \in \mathrm{CPTP}(\mathcal{H}_\mathrm{d})$, the diamond norm is defined through optimization of a trace norm as

$$||\Lambda||_\diamond = \max_{A \in L(\mathcal{H}_{\mathrm{d}^2}),\ ||A||_1 \leq 1} ||\left(\Lambda \otimes \mathcal{I}_\mathrm{d}\right) A||_1 . \tag{2.6}$$

Total-Variation Distance between two probability distributions $\mathbf{p} = \{p_i\}_{i=1}^n$ and $\mathbf{q} = \{q_i\}_{i=1}^n$ is defined by

$$\mathrm{TV}\left(\mathbf{p}, \mathbf{q}\right) = \frac{1}{2} \sum_{i=1}^{n} |p_i - q_i| . \tag{2.7}$$

The TV distance quantifies the maximal statistical distinguishability of $\mathbf{p}$ and $\mathbf{q}$. Specifically, in a task when we are asked to decide whether the provided samples come from $\mathbf{p}$ or $\mathbf{q}$ (where both are promised to be given with probability $\frac{1}{2}$), the optimal success probability (i.e., probability of correctly guessing using the best possible strategy) is $\frac{1}{2}\left(1 + \mathrm{TV}\left(\mathbf{p}, \mathbf{q}\right)\right)$ [17]. In quantum mechanics, the analogous task is to distinguish between two quantum objects, which can be either states, measurements, or channels (and, again, both are promised to be given with probability $\frac{1}{2}$), provided samples from the probability distributions that the objects of interest generate (via Born's rule, Eq. (2.2)). In all cases, the optimal success probability of performing this task is related to the optimal (maximized) TV distance between relevant probability distributions. This success probability

is given by similar formula $\frac{1}{2}(1+\mathrm{d}(\alpha_1,\alpha_2))$, where α_1 and α_2 denote two objects to be distinguished, and the distance $\mathrm{d}(.\,,\,.)$ depends on the scenario. In Fig. 2.1, we pictorially represent the most important distances based on optimal statistical distinguishability.

In the task where we want to distinguish between quantum states ρ and σ, we optimize over measurements (POVMs) performed on them, and the relevant distance is *trace distance* defined as [17]

$$\mathrm{d}_{\mathrm{tr}}(\rho,\sigma) = \sup_{\mathbf{M}\in\mathrm{P}(\mathcal{H})} \mathrm{TV}\left(\mathbf{p}^{\rho,\mathbf{M}},\mathbf{p}^{\sigma,\mathbf{M}}\right) = \frac{1}{2}||\rho-\sigma||_1\,, \tag{2.8}$$

where by $\mathbf{p}^{\rho,\mathbf{M}}$ we denote probability distribution obtained via Born's rule when measurement $\mathbf{M}$ is performed on state ρ. In this case, the optimal measurement, known as Helstrom's measurement, is projective with 2 outcomes [43].

In the case of quantum measurements, we want to decide whether the measurement performed is a POVM $\mathbf{M}$ or $\mathbf{N}$, and we are optimizing over input states. The relevant distance is so called *operational distance* defined as [44; 45; 46]

$$\mathrm{d}_{\mathrm{op}}(\mathbf{M},\mathbf{N}) = \sup_{\rho\in\mathrm{D}(\mathcal{H})} \mathrm{TV}\left(\mathbf{p}^{\rho,\mathbf{M}},\mathbf{p}^{\sigma,\mathbf{N}}\right). \tag{2.9}$$

Finally, for distinguishing between two quantum channels Λ and Γ, we are optimizing over both input states (with ancillae) and measurements. In this case, the relevant distance is known as *diamond distance* defined as [17]

$$\mathrm{d}_{\diamond}(\Lambda,\Gamma) = \sup_{\rho\in\mathrm{D}\left(\mathcal{H}^{\otimes 2}\right),\ \mathbf{M}\in\mathrm{P}\left(\mathcal{H}^{\otimes 2}\right)} \mathrm{TV}\left(\mathbf{p}^{\rho,\Lambda,\mathbf{M}},\mathbf{p}^{\sigma,\Gamma,\mathbf{M}}\right), \tag{2.10}$$

where we extended channel $\Lambda\otimes\mathcal{I}_{\mathrm{d}}$ via identity channel $\mathcal{I}_{\mathrm{d}}$ acting on ancillary system. While for the above distance, we do not have a simple expression as a function of underlying objects, its calculation can be formulated as an SDP program that can be efficiently computed for moderate system sizes [47].

Note that in each case, the statistical distinguishably of two objects of interest is *maximized*. Thus, in the use-case where one of the objects is target implementation, and the second is its noisy version, the distances above correspond to worst-case performance of the given object. For example, in the case of a noisy measurement, the worst-case operational distance corresponds to an input quantum state that is the worst possible in terms of distance from ideal, noiseless statistics. In Chapter 3, we will propose and study new types of distances that, in contrast, quantify the average-case performance.

2.5 Unitary designs

In Chapter 3, we will be interested in expected values (integrals) $\mathop{\mathbb{E}}_{\beta\sim\nu} f(\beta) = \int_{\mathrm{U}(\mathcal{H})} d\nu(\beta) f(\beta)$ of a random variable f with respect to measure ν defined on

unitary group $\mathsf{U}(\mathcal{H})$. The measure ν on unitary group induces measure on the set of pure quantum states in the following way – choose arbitrary fixed state ψ_0 and apply to it unitary $U \sim \nu$ drawn from measure ν, obtaining random state $\psi = U\psi_0 U^\dagger$. In short, we denote $\psi \sim \nu_{\mathcal{S}}$. The unique left-and right-invariant probability measure on $\mathsf{U}(\mathcal{H})$ is known as the Haar measure [48] and it will be denoted as μ. Random states obtained from the induced measure on states are called Haar-random states and the corresponding measure will be denoted by $\mu_{\mathcal{S}}$.

Instrumental in our considerations, will be the notion of (approximate) unitary 4-designs. Unitary k-designs are, by definition, measures on $\mathsf{U}(\mathcal{H})$ that reproduce averages of Haar measure μ on balanced polynomials of degree k in entries of U [49]. For approximate k-designs these averages agree only approximately, and the quantitative notion of approximation can be defined differently (see, e.g., [50]). Here we adapt the notion of approximation based on the diamond norm. We say that a measure ν on $\mathsf{U}(\mathcal{H})$ is δ-approximate k-design if

$$\left\| \mathcal{T}_{k,\nu} - \mathcal{T}_{k,\mu} \right\|_\diamond \leq \delta \,, \tag{2.11}$$

where $\mathcal{T}_{k,\nu}$ is the quantum channel acting on $\mathcal{H}^{\otimes k}$ defined as

$$\mathcal{T}_{k,\nu}(\rho) = \int_{\mathsf{U}(\mathcal{H})} d\nu(U) U^{\otimes k} \rho (U^\dagger)^{\otimes k} \,. \tag{2.12}$$

An important example of approximate k-designs are the 1D architecture random quantum circuits formed from *arbitrary* universal gates that randomly couple neighboring qubits. These easy-to-implement circuits approximate k-designs efficiently with the number of qubits N [51; 52; 53**?**]. In particular, it has been recently shown [54] that δ-approximate 4-designs are generated by circuits of depth $O\left(\log\left(\frac{N}{\delta}\right)\right)$ with architecture as simple as 1D line, with moderate numerical constants (see Appendix B of [54] for the exact scaling).

3. Operational Quantum Average-Case Distances

3.1 Chapter overview

3.1.1 Summary in the context of the thesis

In the quest to facilitate robust execution of quantum measurements, it is imperative to assess the quality of implementation itself. To that end, reliable measures of quality are needed. In this Chapter, we introduce novel metrics for the distance (quantum Average-Case distances, or simply AC distances) between quantum objects (states, measurements, and channels) that have a sound operational interpretation - they quantify the average-case statistical distinguishability of two objects, where the average is taken over random quantum circuits. From the point of view of assessing the measuring device's quality, the AC distance between the ideal measurement and its noisy (experimental) version may be interpreted as quantifying an average-case performance of the noisy measurement. From a fundamental perspective, it allows us to quantitatively assess how protocols involving random quantum circuits can perform in the task of distinguishing between two measurements.

The chapter is based on works [1; 2] co-authored by the author of this Thesis. The chapter contains large excerpts from both publications, with edits necessary to adjust them for the format of a Ph.D. thesis.

3.1.2 Technical abstract

We introduce distance measures between quantum states, measurements, and channels based on their statistical distinguishability in generic experiments. Specifically, we analyze the average Total Variation Distance (TVD) between output statistics of protocols in which quantum objects are intertwined with random circuits and measured in a standard basis.

We prove that once a family of random circuits forms an δ-approximate 4-design, with $\delta = o(d^{-8})$, then the average-case distances can be approximated by simple explicit functions that can be expressed via simple degree two polynomials in objects of interest. For systems of moderate dimension, they can be easily explicitly computed – no optimization is needed as opposed to diamond norm distance between channels [47] or operational distance between measurements [45]. We prove that those functions, which we call quantum Average-Case Dis-

tances, have a plethora of desirable properties, such as subadditivity w.r.t. tensor products, joint convexity, and they respect (a certain form of) data-processing inequalities. Notably, all distances utilize the Hilbert-Schmidt (HS) norm, which provides this norm with a new operational interpretation. We also provide upper bounds on the maximal ratio between worst-case and average-case distances, and for each of them, we provide an example that saturates the bound. Specifically, we show that for each dimension d this ratio is at most $d^{\frac{1}{2}}$, d, $d^{\frac{3}{2}}$ for states, measurements, and channels, respectively.

To support the practical usefulness of our findings, we apply them to analyze the effects of noise in quantum advantage experiments and for efficient discrimination of high-dimensional states and channels without quantum memory. We argue that AC distances are better suited for assessing the quality of NISQ devices than common distance measures such as trace distance or the diamond norm. We support those claims by performing numerical simulations that compare worst-case and average-case distances in practically-inspired scenarios.

3.2 Introduction

The question of how far away are two quantum objects (states, measurements, or channels) is of both fundamental and practical importance. That question is often phrased in terms of the statistical distinguishability of probability distributions corresponding to two objects in question (which is a problem of classical hypothesis testing [17]). Indeed, the most common distances, such as trace distance or diamond norm distance, are based on optimal protocols for such statistical discrimination. However, those protocols have limitations. In general, they might require a lot of resources (e.g., high-depth circuits) [55], thus they are not necessarily practical. Another perspective on the limitations of common distance measures comes from a study of noise on quantum devices. A distance between a theoretical (ideal) model of an object in question, and a model for its experimental (noisy) implementation, can be used to study the potential effects of experimental imperfections on the protocol one wishes to implement. When that is the case, the distances based on optimal state/measurement/channel discrimination inform about the worst-case performance of a protocol. However, in practice, one does not necessarily expect the worst-case to be representative of a typical device's performance.

With this motivation in mind, we propose distance measures of distance based on *average statistical distinguishability* using random quantum circuits. Operationally, if the average-case distance between a pair of quantum objects is significant, this implies that they can be (statistically) distinguished almost perfectly using just a few implementations of random circuits. This provides a natural interpretation analogous to conventional distances, but we consider averages over random circuits instead of optimal scenarios. Such quantifiers can be more suitable for studying the performance of NISQ devices' performance than the above-mentioned conventional distances quantifying worst-case performance. In particular, one of the most promising near-term applications of quantum computing

are hybrid quantum-classical variational algorithms [56], such as Quantum Approximate Optimization Algorithm (QAOA) [57; 58; 59] and Variational Quantum Eigensolver (VQE) [60; 61; 62]. Since NISQ devices are expected to suffer from a significant amount of noise, it is instrumental to understand how it can affect such algorithms (see, e.g., [63; 64; 65; 66]). Our distance measures might prove particularly useful in this context because, as explained later, the random circuits we consider form unitary designs. Recently it was realized that circuits appearing in variational algorithms are expected to have, on average, design-like properties. Thus we expect quantum average-case distances to be a good metric to quantify the average performance of such algorithms [67].

3.2.1 Summary of results

In this chapter, we study the average Total-Variation (TV) distance between measurement outputs (statistics) of two quantum processes, in which quantum objects of interest are intertwined with random quantum circuits. TV distance is well known to quantify the statistical distinguishability of two probability distributions. However, in general, since TV distance is not a polynomial function of underlying probability distributions, the relevant averages are hard to calculate. To tackle that problem, we derive lower and upper bounds for average TV distance and show that both bounds differ only by dimension-independent *constants*. The derivation of upper bounds requires the calculation of 2nd moments of quantities of interest, and lower bounds are derived using 2nd and 4th moments. Formally, this means that to get both upper and lower bounds, the random circuits must form an approximate 4-design. Importantly, our results are valid also for any (approximate) k-design with $k \geq 4$. We note that a particular choice of 4-designs is of purely technical origin – as remarked above, our proof techniques require 4th-degree polynomials to get lower bounds on average TV distance, while for upper bounds already 2-designs suffice.

The existence of the above bounds implies that for a broad family of random quantum circuits, the average TV distance is *approximated*, up to the known relative error, by a simple explicit function (2nd-degree polynomial) of the objects that we wish to compare (states, measurements, or channels). These functions, the quantum average-case distances, define bona fide distance measures with multiple desired properties, such as subadditivity w.r.p. to tensor products, joint convexity, or (restricted) data processing inequalities. Notably, all of the distances utilize the Hilbert-Schmidt (HS) norm in some way which provides it with a new operational interpretation.

Finally, so-defined quantum average-case distances have sound operational interpretation. Namely, if a TV distance is bounded from below by a constant c (here proportional to quantum average-case distance), then there exists a strategy that uses random circuits which distinguishes between two objects with probability at least $\frac{1}{2}(1 + c)$ in *single-shot* scenario (we note that we do not attempt to provide constructions of such strategies, thus it's purely an existence statement). Thus from Hoeffding bound, it follows that having access to multiple copies (samples) allows one to exponentially quickly approach the success

probability of discrimination equal to 1 using a simple majority vote.

3.2.2 Related works

Let us now comment on some of the commonly used distances. The study of similarity measures between quantum objects has a long history [17; 68], and thus there are a lot of different metrics currently used in the field. Some of the most popular distances are based on the *optimal* statistical distinguishability of quantum objects – this includes trace distance between states [69], the operational distance between measurements [45], as well as diamond norm distance between channels [69]. While in those distances the optimization is done over all possible operations, there has been an interest also in distinguishability under restricted sets of operations – such as local POVMs for discrimination of quantum states [70; 71]. Recently, a quantum Wasserstein distance of order 1 was proposed as a measure of distance between quantum states. It generalizes a classical Wasserstein distance based on the Hamming weight and captures the notion of similarity of quantum states based on differences between their marginals [72].

For quantum states, the other very common similarity measure is quantum fidelity, which induces distance between states known as Bures distance [73; 74]. When one wants to compare unitary channel (quantum gate) with a general channel (noisy implementation of a gate), the relevant notions are worst-case [17] and average-case gate fidelity [75; 76; 77; 78; 79]. In both cases, the relevant optimization/averaging is over all quantum states. For distance measures between measurements, one of the natural choices is to treat measurement as a quantum-classical channel and compute diamond norm distance [45; 46]. In the context of detector tomography sometimes fidelities between theoretical and experimental POVM's elements were considered [80; 81; 82]. When the target measurement is a computational basis, it is customary to use single-qubit error probabilities as a simplified quantifier of measurement's quality [21]. See [68] for an extensive overview of distinguishability measures between quantum objects.

The distance measures introduced by us rely on random quantum circuits which have many applications in the context of practical quantum computing. A notable example is shadow tomography, where random circuits are exploited to estimate multiple properties of quantum states with relatively low sample complexity [83; 84; 85; 86; 87]. Another example are generalizations of the classical randomized-benchmarking scheme [88; 89; 90; 91] that use random circuits to estimate averaged quality metrics of quantum gates [92; 93; 94].

In Ref. [49] the authors prove that two states distant in Hilbert-Schmidt norm can be distinguished by a POVM constructed from approximate 4-design. Our proofs concerning average TV distances for quantum states and measurements were inspired by the proofs therein. In Ref. [95] the authors derived lower bounds (also containing HS distance) for TV distance in the same scenario for Haar-random POVMs, investigating applications for hidden subgroup problems. In Ref. [96] the "total operational distance" between states was introduced. It is

based on the differences in obtained statistics when one performs mutually complementary projective measurements (known more commonly as mutually unbiased bases [97]). Importantly, the authors show that such distance is equivalent to HS distance between states of interest.

3.2.3 Main applications overview

Let us now briefly review the two main practical applications of our results that we study in detail in Section 3.5.

Noise in quantum advantage based on random circuits sampling

Multiple recent quantum advantage proposals are based on random circuits sampling [21; 98]. We apply our findings to understand the effects of noise on such protocols. We approach the problem from two sides. First, the AC distances allow to easily *lower* bound the average-case TV distance between the noisy distribution and the ideal distribution, thus giving insight into how well separated, on average, are noisy distributions from target distributions. Second, AC distances allow to *upper* bound the average-case TV distance between a noisy distribution and a (trivial) uniform distribution. This allows us to study how quickly the noise makes the average distribution useless. For example, we show that even in the absence of gate and state-preparation noise, a local, symmetric bitflip error in measurements causes the noisy distribution to approach trivial one exponentially quickly in system size.

Randomized protocols for distinguishing quantum objects

Recently there has been a lot of interest in algorithms that use randomized quantum circuits, such as shadow tomography [83; 84; 85; 86; 87] and randomized-benchmarking [88; 89; 90; 91; 92; 93]. Our results can be employed to quantify the performance of randomized algorithms in the task of statistical distinguishability of quantum objects. Namely, if the average-case distance between a pair of quantum objects on N qubit systems is large, then they can be (statistically) distinguished almost perfectly using a randomized protocol with just a few implementations of local random circuits of depth $O(N)$. We observe that such behavior takes place in two scenarios related to those recently analyzed in the context of so-called Quantum Algorithmic Measurement [99] and complexity growth of quantum circuits [55]: (i) distinguishing Haar random N qubit pure state from maximally mixed state and (ii) distinguishing N qubit Haar random unitary from maximally depolarizing channel. This shows that protocols employing random circuits can be used to efficiently discriminate quantum objects. Since they do not depend on the objects to be distinguished, randomized measurement schemes can be interpreted as "universal discriminators", analogous to the SWAP test but not requiring the usage of entanglement or coherent access to copies of quantum systems.

3.2.4 Structure of the chapter

In Section 3.3 we define the average Total-Variation distance between two states, measurements, and channels. We also outline the general methodology of the proofs presented in the main section of this chapter – Section 3.4, in which we prove that the average Total-Variation distances between quantum objects can be approximated by explicit functions of the objects in question – quantum states in Theorem 1 (Section 3.4.1), quantum measurements in Theorem 2 (Section 3.4.2), and quantum channels in Theorem 3 (Section 3.4.3). Those functions are what we call quantum average-case distances. The main section is followed by Section 3.5 in which we discuss in detail the most interesting applications of AC distances (summarized already above in Section 3.2.3). This includes results of extensive numerical simulations in Section 3.5.4. In Section 3.6 we prove that quantum average-case distances possess a variety of desired properties, such as subadditivity, joint convexity, and restricted data-processing inequalities – summarized in Table 3.1 for states, Table 3.2 for measurements, and Table 3.3 for channels. In this section, we also prove asymptotic separations between average-case and worst-case distances, together with examples that saturate derived bounds. In Section 3.7 we study exemplary scenarios where quantum average-case distances can be calculated analytically. We also show that average-case distances can be used to study the average convergence of noisy distribution to uniform (trivial) distribution, a property that was crucial to studying some of the practical applications in Section 3.5. We conclude the Chapter with Section 3.8 where we summarize our contributions.

3.3 Methodology

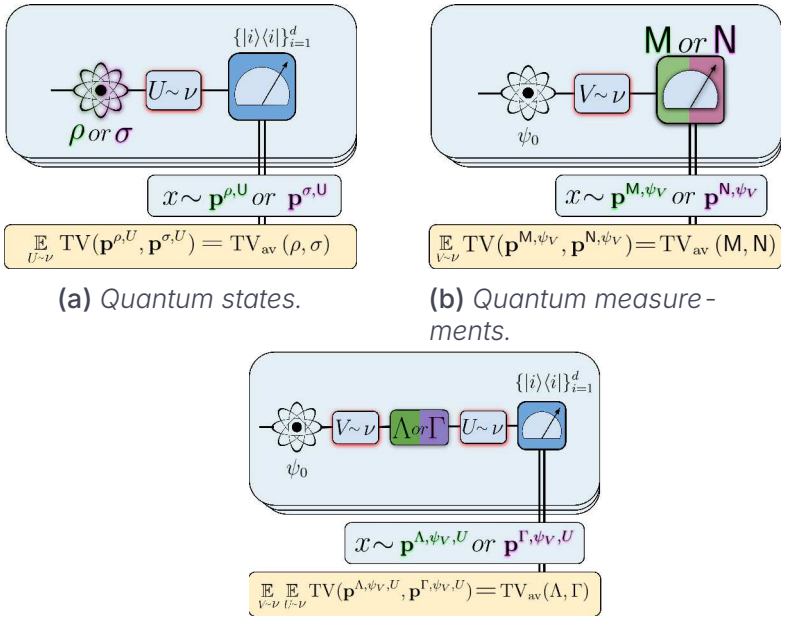
The goal of this section is to provide an overview of the general setup and a proof methodology for the main results of this work that are presented in Section 3.4.

3.3.1 Average Total Variation distances

We will be interested in establishing bounds for *average Total-Variation distance* between probability distributions generated by two quantum objects (states, measurements, or general channels). The average will be taken over an ensemble of *random* circuits. These notions are represented pictorially in Fig. 3.1, and we will now formally define them.

Consider a general quantum protocol that consists of a state preparation, an evolution of the system, and a quantum measurement. Now we consider *average Total-Variation distance* between two quantum objects:

1. (*States*) Two quantum states $\rho, \sigma \in \mathsf{D}(\mathcal{H})$ are fixed, rotated by a random unitary, and measured in the computational basis. Let us denote by $\mathbf{p}^{\rho,U}$ probability distribution obtained in this process, i.e., $p_i^{\rho,U} = \mathrm{tr}\left(|i\rangle\langle i| U\rho U^\dagger\right)$.



(a) *Quantum states.*

(b) *Quantum measurements.*

(c) *Quantum channels.*

Figure 3.1: *Measures of distance between quantum objects based on* average *statistical distinguishability. For quantum states (3.1a), we take the average over random unitaries applied to the state, followed by measurement in a standard basis. For quantum measurements (3.1a), we take the average over random pure states measured on the detector. Finally, for quantum channels (2.1c) we take the average over random input states, and random unitaries applied* after *the action of the channel. Note the difference with Fig. 2.1, where for common distance measures the* optimal *protocol is chosen, while here we consider random protocols.*

The average TV distance between ρ and σ is

$$\mathrm{TV}_{\mathrm{av}}(\rho,\sigma) \coloneqq \mathop{\mathbb{E}}_{U\sim\nu} \mathrm{TV}\left(\mathbf{p}^{\rho,U},\mathbf{p}^{\sigma,U}\right). \tag{3.1}$$

2. (*Measurements*) Two n-outcome quantum measurements $\mathbf{M},\mathbf{N}\in\mathrm{P}(\mathcal{H},n)$ are fixed, while states are taken to be random. Let us denote by $\mathbf{p}^{\mathbf{M},\psi_V}$ probability distribution obtained in this process, i.e., $p_i^{\mathbf{M},\psi_V}=\mathrm{tr}\left(M_i V\psi_0 V^\dagger\right)$, where ψ_0 is a fixed pure state. The average TV distance between $\mathbf{M}$ and $\mathbf{N}$ is

$$\mathrm{TV}_{\mathrm{av}}(\mathbf{M},\mathbf{N}) \coloneqq \mathop{\mathbb{E}}_{V\sim\nu} \mathrm{TV}\left(\mathbf{p}^{\mathbf{M},\psi_V},\mathbf{p}^{\mathbf{N},\psi_V}\right). \tag{3.2}$$

3. (*Channels*) Two quantum channels $\Lambda,\Gamma\in\mathrm{CPTP}(\mathcal{H})$ are fixed. The input state is taken to be a random pure state $V\psi_0V^\dagger$ for fixed ψ_0. The output state is rotated by independent random unitary U (hence we have random unitary rotations before and after the application of a channel), followed by

measurement in a standard basis. Let us denote by $\mathbf{p}^{\Lambda,\psi_V,U}$ probability distribution obtained in this process, i.e., $p_i^{\Lambda,\psi_V,U} = \mathrm{tr}\left(|i\rangle\langle i| U\Lambda\left(V\psi_0 V^\dagger\right)U^\dagger\right)$ The average TV distance between Λ and Γ is

$$\mathrm{TV}_{\mathrm{av}}(\Lambda,\Gamma) := \mathop{\mathbb{E}}_{U\sim\nu}\mathop{\mathbb{E}}_{V\sim\nu}\mathrm{TV}\left(\mathbf{p}^{\Lambda,\psi_V,U},\mathbf{p}^{\Gamma,\psi_V,U}\right). \tag{3.3}$$

Remark 1. *If the average TV distance is bounded from below by a constant c, then there exists a strategy that uses random circuits which distinguishes between two objects with probability at least $\frac{1}{2}(1+c)$ in single-shot scenario. Thus, by the virtue of Hoeffding's inequality, having access to s copies (samples) gives an error probability of the majority vote strategy dropping as $2\exp(-\frac{c^2}{2}s)$. We note that while the lower bound implies existence of such strategy, it does not tell what is the exact protocol for realizing this success probability.*

Remark 2. *The value of the average* TV *distance for quantum states can be reinterpreted as* TV*-distance of output statistics resulting from a measurement of a* single *POVM with effects $M_{i,U_j} = \nu_j U_j^\dagger |i\rangle\langle i| U_j$, where ν_j is the probability of occurrence of circuit U_j in the ensemble ν (for simplicity of presentation we assume that ensemble ν is discrete). This POVM can be interpreted as a convex combination [100] of projective measurements $\mathbf{M}^{U_j}$ with effects $M_i^{U_j} = U_j^\dagger |i\rangle\langle i| U_j$. Analogous interpretation holds also for the average* TV *distances for quantum measurements and channels – they can be interpreted as* TV *distances between output statistics of the corresponding randomized protocols [49]. Recall from Remark 1 that a lower bound on average TV distance implies that such randomized protocol distinguishes between quantum states with high probability. We note that it immediately follows that there also exists a deterministic (i.e., requiring implementation of a single quantum measurement, for example, via Naimark's dilation [101], as opposed to randomized implementation of multiple measurements)* optimal *distinguishability protocol that achieves the same success probability (the value of which depends on the states in question).*

3.3.2 Auxiliary lemmas

We will later be interested in bounding from below and from above the expected values of some random variables. In bounding from above, we will use the following

Lemma 1. *(Jensen's inequality [102]) Let f be a concave function, and X be a random variable. Then we have*

$$f\left(\mathbb{E}X\right) \geq \mathbb{E}f(X).$$

On the other hand, in bounding from below, we will use the following

Lemma 2. *(Berger's inequality [103]) Let X be a random variable with well-defined second and fourth moments. Then we have*

$$\frac{(\mathbb{E}[X^2])^{\frac{3}{2}}}{(\mathbb{E}[X^4])^{\frac{1}{2}}} \leq \mathbb{E}|X| \,. \tag{3.4}$$

We will also make use of the following auxiliary lemmas.

Lemma 3 (Auxiliary integral involving k-th moment [104, Prop. 6])**.** *Let $X \in \mathrm{Herm}((\mathcal{H}_d)^{\otimes k})$ and μ be a Haar measure. Then we have*

$$\mathop{\mathbb{E}}_{U\sim\mu} \mathrm{tr}\left(U^{\otimes k}\,|i\rangle\langle i|^{\otimes k}\,(U^\dagger)^{\otimes k} X\right) = \frac{1}{\binom{d+k-1}{k}}\,\mathrm{tr}\left(\mathbb{P}^{(k)}_{\mathrm{sym}} X\right)\,, \tag{3.5}$$

where $\mathbb{P}^{(k)}_{\mathrm{sym}}$ is the projector onto k-fold symmetric subspace $\mathcal{H}^{(k)}_{\mathrm{sym}} \subset \mathcal{H}_d^{\otimes k}$.

Corollary 1 (Auxiliary integral for 2nd moment)**.** *Let $X \in \mathrm{Herm}(\mathcal{H}_d)$. Then we have*

$$\mathop{\mathbb{E}}_{U\sim\mu} \mathrm{tr}(|i\rangle\langle i|\,UXU^\dagger)^2 = \frac{1}{d(d+1)}\left(\mathrm{tr}(X^2) + \mathrm{tr}(X)^2\right)\,. \tag{3.6}$$

Proof. The above identity follows from Lemma 3. We use the identities $\mathbb{P}^{(2)}_{\mathrm{sym}} = \frac{1}{2}(\mathbb{I}\otimes\mathbb{I} + \mathbb{S})$ and $\mathrm{tr}(\mathbb{S}\rho\otimes\rho) = \mathrm{tr}(\rho^2)$, where $\mathbb{S}$ denotes the swap operator acting on $\mathcal{H}^{\otimes 2}$. □

Lemma 4 (Lemma 2 from [105])**.** *Let $X \in \mathrm{Herm}(\mathcal{H})$. Let $\mathbb{P}^{(k)}_{\mathrm{sym}}$ denotes orthogonal projector onto k-fold symmetrization of $\mathcal{H}^{(k)}_{\mathrm{sym}} \subset \mathcal{H}^{\otimes k}$. We then have the following inequality*

$$\mathrm{tr}\left(X^{\otimes 4}\,\mathbb{P}^{(4)}_{\mathrm{sym}}\right) \leq C\,\mathrm{tr}\left(X^{\otimes 2}\,\mathbb{P}^{(2)}_{\mathrm{sym}}\right)^2\,,\ \textit{where}\ C = \frac{10.1}{6}\,. \tag{3.7}$$

Finally, the following Lemma 5, proved in Appendix A.1, generalizes Lemma 4 and can be of independent interest. This result will be instrumental in proofs regarding average-case distances between quantum channels.

Lemma 5 (Inequality involving two operators and projections onto 2-fold symmetric subspaces)**.** *Let $X, Y \in \mathrm{Herm}(\mathcal{H})$. Let $\mathbb{P}^{(k)}_{\mathrm{sym}}$ denotes the orthogonal projector onto k-fold symmetrization of $\mathcal{H}^{(k)}_{\mathrm{sym}} \subset \mathcal{H}^{\otimes k}$. We then have the following inequality*

$$\mathrm{tr}\left(X^{\otimes 2}\otimes Y^{\otimes 2}\,\mathbb{P}^{(4)}_{\mathrm{sym}}\right) \leq C\,\mathrm{tr}\left(X^{\otimes 2}\,\mathbb{P}^{(2)}_{\mathrm{sym}}\right)\mathrm{tr}\left(Y^{\otimes 2}\,\mathbb{P}^{(2)}_{\mathrm{sym}}\right)\,,\ \textit{where}\ C = \frac{13}{6}\,. \tag{3.8}$$

Remark 3. *Note that the constant appearing on the right-hand side of* (3.8) *is slightly worse than the one from* (3.7)*.*

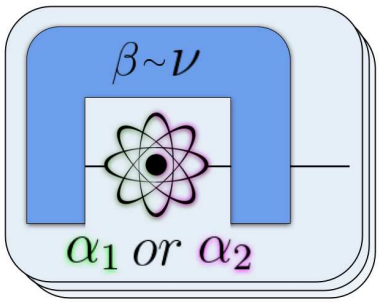


Figure 3.2: *Illustration of the general setup we consider in this work. Two quantum objects α_1, α_2 that can be either quantum states, measurements, or channels, are surrounded by random circuits β drawn from a probability measure ν.*

3.3.3 General methodology of proofs

Consider a general quantum protocol that results in probability distribution $\mathbf{p}^{\alpha,\beta}$ where α denotes a fixed quantum object (state, measurement, or channel), and β is a random variable (usually specifying quantum circuit) distributed according to a probability distribution ν (typically Haar measure, approximate k-design, or random instances of variational circuits). See Fig. 3.2 for illustration. We will be interested in bounding quantities of the type

$$\mathrm{TV}_{\mathrm{av}}(\alpha_1, \alpha_2) \coloneqq \mathop{\mathbb{E}}_{\beta\sim\nu} \mathrm{TV}(\mathbf{p}^{\alpha_1,\beta}, \mathbf{p}^{\alpha_2,\beta}) \,. \tag{3.9}$$

For example, in the case of the distance between quantum states, α would correspond to two fixed quantum states that we want to calculate the distance between, and β would correspond to random quantum measurements (as in Eq. (3.1)).

To estimate $\mathrm{TV}_{\mathrm{av}}(\alpha_1, \alpha_2)$ from above we first expand

$$\mathop{\mathbb{E}}_{\beta\sim\nu} \mathrm{TV}(\mathbf{p}^{\alpha_1,\beta}, \mathbf{p}^{\alpha_2,\beta}) = \frac{1}{2} \sum_{i=1}^{n} \mathop{\mathbb{E}}_{\beta\sim\nu} |p_i^{\alpha_1,\beta} - p_i^{\alpha_2,\beta}|, \tag{3.10}$$

and use Jensen's inequality (see Lemma 1) for the concave function $f(x) = \sqrt{x}$ to upper bound the average of each of the summands

$$\mathrm{TV}_{\mathrm{av}}(\alpha_1, \alpha_2) \leq \frac{1}{2} \sum_{i=1}^{n} \sqrt{\mathop{\mathbb{E}}_{\beta\sim\nu} (p_i^{\alpha_1,\beta} - p_i^{\alpha_2,\beta})^2} \,. \tag{3.11}$$

To establish a lower bound for $\mathrm{TV}_{\mathrm{av}}(\alpha_1, \alpha_2)$ we will apply Berger's inequality (see Lemma 2) to random variables $x_i = p_i^{\alpha_1,\beta} - p_i^{\alpha_2,\beta}$ and insert the obtained result to (3.10). Importantly, in Section 3.4 it will turn out that for probabilities and measures involved in our considerations, we will have

$$\mathop{\mathbb{E}}_{\beta\sim\nu} (p_i^{\alpha_1,\beta} - p_i^{\alpha_2,\beta})^4 \leq C \left[\mathop{\mathbb{E}}_{\beta\sim\nu} (p_i^{\alpha_1,\beta} - p_i^{\alpha_2,\beta})^2 \right]^2 , \tag{3.12}$$

where $C > 0$ is a constant independent of the dimension of the Hilbert space or the number of measurement outcomes. This fact, together with Berger's inequality (Eq. (3.4), yields the bound

$$\frac{1}{C^{1/2}}\frac{1}{2}\sum_{i=1}^{n}\sqrt{\mathop{\mathbb{E}}_{\beta\sim\nu}(p_i^{\alpha_1,\beta}-p_i^{\alpha_2,\beta})^2} \leq \mathrm{TV}_{\mathrm{av}}(\alpha_1,\alpha_2)\ . \tag{3.13}$$

Therefore, we have

$$\frac{1}{C^{1/2}}\frac{1}{2}\sum_{i=1}^{n}\sqrt{\mathop{\mathbb{E}}_{\beta\sim\nu}(p_i^{\alpha_1,\beta}-p_i^{\alpha_2,\beta})^2} \leq \mathrm{TV}_{\mathrm{av}}(\alpha_1,\alpha_2)\ \leq \frac{1}{2}\sum_{i=1}^{n}\sqrt{\mathop{\mathbb{E}}_{\beta\sim\nu}(p_i^{\alpha_1,\beta}-p_i^{\alpha_2,\beta})^2}\,, \tag{3.14}$$

which makes it clear that to calculate both lower and upper bounds for average TV distance we will need to calculate $\mathop{\mathbb{E}}_{\beta\sim\nu}(p_i^{\alpha_1,\beta}-p_i^{\alpha_2,\beta})^2$. Importantly, since both bounds will differ only by a constant (independent of dimension), it will motivate the introduction of quantum average-case distances defined as

$$\mathrm{d}_{\mathrm{av}}(\alpha_1,\alpha_2) \coloneqq \frac{1}{2}\sum_{i=1}^{n}\sqrt{\mathop{\mathbb{E}}_{\beta\sim\nu}(p_i^{\alpha_1,\beta}-p_i^{\alpha_2,\beta})^2}\ . \tag{3.15}$$

Fortunately, as will be shown in Section 3.4, such terms can be expressed via simple, explicit functions of the quantum objects that we want to calculate the distance between, provided that ν forms an approximate 4-design.

Remark 4. *We note that depending on the perspective one adopts, either the upper bound or lower bound on average TV distance might be of particular interest. Namely, if one wishes to compare the ideal implementation of some protocol with its noisy version, then the upper bound might be satisfactory. In such a scenario, the ensemble of random circuits suffices to be approximate 2-design, since only 2nd moments are needed for its calculation. On the other hand, for statistical distinguishability, the lower bound is important (see Remark 1) and thus 4-design property is necessary.*

3.4 Quantum average-case distances

In this section, we present our main technical results. We prove that if random circuits form approximate unitary 4-designs, the average TV distances (see Section 3.3) can be approximated, up to the relative error, by simple functions that can be expressed by degree-2 polynomials in quantum objects in question. We provide explicit expressions for those functions (quantum average-case distances), as well as numerical constants for the relative errors. The proofs given in this section concern *exact* unitary 4-designs, while derivations for approximate designs are relegated to Appendix A.2.

3.4.1 Quantum states

Let $\mathbf{p}^{\rho,U}$ denote the probability distribution obtained when the state ρ (σ) undergoes a unitary transformation according to U and is subsequently measured in the computational basis of $\mathcal{H}_d$. In other words $p_i^{\rho,U} = \mathrm{tr}\left(|i\rangle\langle i| U\rho U^\dagger\right)$, where $\{|i\rangle\}_{i=1}^d$ is a computational basis of $\mathcal{H}$.

Theorem 1 (Average-case distinguishability of quantum sates). *Let $\rho, \sigma \in \mathrm{D}(\mathcal{H}_d)$ be states on $\mathcal{H}_d$ and let U be a random unitary in $\mathcal{H}_d$ drawn from measure ν that forms a δ-approximate 4-design, with $\delta := \frac{\delta'}{2d^4}$, $\delta' \in \left[0, \frac{1}{3}\right]$. We then have the following inequalities*

$$\ell(\delta')\ a\ \mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\rho,\sigma) \leq \mathop{\mathbb{E}}_{U\sim\nu} \mathrm{TV}(\mathbf{p}^{\rho,U}, \mathbf{p}^{\sigma,U}) \leq u(\delta')\ A\ \mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\rho,\sigma)\,, \tag{3.16}$$

where we define the quantum average-case distance between states

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\rho,\sigma) = \frac{1}{2}\sqrt{\mathrm{tr}([\rho-\sigma]^2)} = \frac{1}{2}\|\rho-\sigma\|_{\mathrm{HS}}\,, \tag{3.17}$$

and $a = 0.31$, $A = \sqrt{\frac{d}{d+1}} \leq 1$, $\ell(\delta') = \sqrt{\frac{(1-\frac{\delta'}{d^2})^3}{1+\delta'}}$, $u(\delta') = \left(1+\frac{\delta'}{d^2}\right)^{\frac{1}{2}}$.

Proof. In what follows we prove a version of the theorem for exact 4-designs (i.e., setting $\delta = 0$; the proof for approximate 4-designs can be found in Appendix A.2.1). We start by proving the upper bound in (3.16). To this aim, we utilize the upper bound in (3.11) (from Jensen's inequality) to obtain

$$\mathrm{TV}_{\mathrm{av}}(\rho,\sigma) \leq \frac{1}{2}\sum_{i=1}^{d} \sqrt{\mathop{\mathbb{E}}_{U\sim\nu} \mathrm{tr}(|i\rangle\langle i| U\Delta U^\dagger)^2}\,, \tag{3.18}$$

where we set $\Delta = \rho - \sigma$. Using the assumed 2-design property of ν and the standard techniques of Haar measure integration (cf. Corrolary 1) we get

$$\mathop{\mathbb{E}}_{U\sim\nu} \mathrm{tr}(|i\rangle\langle i| U\Delta U^\dagger)^2 = \frac{1}{d(d+1)}\mathrm{tr}(\Delta^2)\,, \tag{3.19}$$

which follows directly from Eq. (3.6) and the fact that Δ is traceless. Inserting the above into (3.18) we obtain the upper bound from (3.16).

In order to prove the lower bound we use Berger's inequality (cf. Eq. (3.4)) for variable $X = \mathrm{tr}(U|i\rangle\langle i| U^\dagger\Delta)$:

$$\mathop{\mathbb{E}}_{U\sim\nu} |\mathrm{tr}(U|i\rangle\langle i| U^\dagger\Delta)| \geq \frac{\left(\mathop{\mathbb{E}}_{U\sim\nu}[\mathrm{tr}(U|i\rangle\langle i| U^\dagger\Delta)]^2\right)^{3/2}}{\left(\mathop{\mathbb{E}}_{U\sim\nu}[\mathrm{tr}(U|i\rangle\langle i| U^\dagger\Delta)]^4\right)^{1/2}}\,. \tag{3.20}$$

The numerator of the above fraction contains the power of the second moment already calculated in Eq. (3.19), hence we get that it is equal to $K = [\frac{1}{d(d+1)} \operatorname{tr}(\Delta^2)]^{3/2}$. To get the upper bound for the denominator, we first note that from Lemma 3 it follows directly that the denominator is equal to $L = [\binom{d+3}{4}^{-1} \operatorname{tr}(\mathbb{P}^{(4)}_{\text{sym}} \Delta^{\otimes 4})]^{1/2}$, where $\mathbb{P}^{(4)}_{\text{sym}}$ is a projector onto the 4-fold symmetric subspace of $\mathcal{H}_d^{\otimes 4}$. Now we get

$$\operatorname{tr}(\mathbb{P}^{(4)}_{\text{sym}} \Delta^{\otimes 4}) \leq C \left(\operatorname{tr}\left(\mathbb{P}^{(2)}_{\text{sym}} \Delta^{\otimes 2}\right)\right)^2 = \frac{C}{4} \left(\operatorname{tr}(\Delta^2)\right)^2 , \tag{3.21}$$

with $C = \frac{10.1}{6}$. The inequality above is a direct application of Lemma 4, while the equality follows from the fact that Δ is traceless and explicit form of $\mathbb{P}^{(2)}_{\text{sym}}$. Inserting everything into Eq. (3.20) we obtain

$$\mathop{\mathbb{E}}_{U\sim\nu} |\operatorname{tr}(U|i\rangle\langle i| U^\dagger \Delta)| \geq \frac{K}{L} \geq \frac{w}{d} \sqrt{\operatorname{tr} \Delta^2} , \tag{3.22}$$

for $w = \sqrt{\frac{4}{C} \frac{\binom{d+3}{4}}{d^2(d+1)^2}} \geq 0.31 = a$. Finally, summing over $i = 1, \ldots, d$, we obtain lower bound on average TV distance

$$\mathop{\mathbb{E}}_{U\sim\nu} \frac{1}{2} \sum_{i=1}^{d} |\operatorname{tr}(U|i\rangle\langle i| U^\dagger \Delta)| \geq a \, \frac{1}{2} ||\Delta||_{\text{HS}} , \tag{3.23}$$

which concludes the proof. □

Remark 5. *The proof of Theorem 1 is inspired by the proof of Theorem 4 from [49] where Berger inequality was used to prove that two states far apart in Hilbert-Schmidt norm can be information-theoretically distinguished by a POVM constructed from approximate* 4*-design.*

Remark 6. *We note that in existing literature, the trace distance was usually preferred to the Hilbert-Schmidt distance, one of the reasons being the lack of an operational interpretation for the latter. The above considerations provide such an interpretation for H-S distance in terms of average statistical distinguishability between quantum states, thus providing a sound physical motivation for its use.*

Remark 7. *We note that random quantum circuits in the 1D architecture formed from* arbitrary *universal gates that randomly couple neighboring qubits, generate approximate k-designs efficiently with the number of qubits N [51; 52; 53; 54]. Specifically, δ- approximate* 4*-designs are generated by the 1D random brickwork architecture in depth $O\left(\log\left(\frac{N}{\delta}\right)\right)$, with moderate numerical constants (see Appendix B of Ref. [54]). This implies that ensembles appearing in Theorem 1 can be easily realized. Furthermore, it is expected that some of the classes of variational quantum circuits are expected to have, on average, unitary design-like properties [67]. This suggests that quantum average-case distances might be used to quantify the average-case performance of hybrid quantum-classical algorithms. Naturally, the same remarks hold for Theorem 2 for measurements and Theorem 3 for channels.*

3.4.2 Quantum measurements

Let $\mathbf{p}^{\mathbf{M},\psi_V}$ denote the probability distribution of a quantum process in which a fixed pure quantum state ψ_0 on $\mathcal{H}_\mathrm{d}$ is evolved according to unitary V and is subsequently measured via a n-outcome POVM $\mathbf{M} = (M_1, M_2, \ldots, M_n)$. In other words $p_i^{\mathbf{M},\psi_V} = \mathrm{tr}(V\psi_0 V^\dagger M_i)$.

Theorem 2 (quantum average-case distance between quantum measurements). *Let $\mathbf{M}, \mathbf{N}$ be n-outcome POVMs on $\mathcal{H}_d$ and V be a random unitary on $\mathcal{H}_d$ drawn from measure ν that forms a δ-approximate 4-design, with $\delta := \frac{\delta'}{2d^4}$, $\delta' \in \left[0, \frac{1}{3}\right]$. We then have the following inequalities*

$$\ell(\delta')\ a\ \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathbf{M},\mathbf{N}) \leq \mathop{\mathbb{E}}_{V\sim\nu} \mathrm{TV}(\mathbf{p}^{\mathbf{M},\psi_V}, \mathbf{p}^{\mathbf{N},\psi_V}) \leq u(\delta')\ A\ \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathbf{M},\mathbf{N})\,, \tag{3.24}$$

where we define an quantum average-case distance between measurements

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathbf{M},\mathbf{N}) = \frac{1}{2d}\sum_{i=1}^{n}\sqrt{\|M_i - N_i\|_{\mathrm{HS}}^2 + \mathrm{tr}(M_i - N_i)^2}\,. \tag{3.25}$$

and $a = 0.31$, $A = \sqrt{\frac{d}{d+1}} \leq 1$, $\ell(\delta') = \sqrt{\frac{(1-\frac{\delta'}{d^2})^3}{1+\delta'}}$, $u(\delta') = \left(1 + \frac{\delta'}{d^2}\right)^{\frac{1}{2}}$.

Proof. In what follows we prove a version of the theorem for exact 4-designs (the proof for approximate 4-designs can be found in Appendix A.2.1). The proof is in fact almost exactly the same as of Theorem 1. We can define $\Delta_i = M_i - N_i$, now each Δ_i having the role of previous Δ, namely in each summand appearing in the TV distance is of the form $|\mathrm{tr}(V\psi_0 V^\dagger \Delta_i)|$. We now note that arbitrary fixed pure state $\psi_0 = U_0 |0\rangle\langle 0| U_0^\dagger$ is unitarily equivalent to computational basis state via some unitary U_0, and that Haar measure is invariant under transformation $U \longrightarrow UU_0$. From those facts, it follows that we can apply exactly the same steps as for proof of Theorem 1. For the second moment we obtain

$$\mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}} \mathrm{tr}(\psi\Delta_i)^2 = \frac{1}{d(d+1)}\left(\mathrm{tr}(\Delta_i^2) + \mathrm{tr}(\Delta_i)^2\right)\,, \tag{3.26}$$

which differs from Eq. (3.19) by additional summand, because now Δ_i is not necessarily traceless. The rest of the steps is exactly analogous to the proof of Theorem 1.

□

3.4.3 Quantum channels

Let $\mathbf{p}^{\Lambda,\psi_V,U}$ be the probability distribution associated with a quantum process in which a fixed pure state $\psi_0 \in \mathcal{S}(\mathcal{H})$ is transformed by unitary transformation V, channel Λ, unitary U, and is subsequently measured in the computational basis of $\mathcal{H}_\mathrm{d}$. In other words we have $p_i^{\Lambda,\psi_V,U} = \mathrm{tr}(|i\rangle\langle i| U\Lambda(V\psi_0 V^\dagger)U^\dagger)$.

Theorem 3 (Average-case distinguishability of quantum channels)**.** *Let Λ, Γ be quantum channels acting on $\mathcal{H}_d$. let ν be a distribution on on $\mathsf{U}(\mathcal{H}_d)$ forming δ-approximate 4-design for $\delta = \frac{\delta'}{(2d)^8}$. Then we have the following inequalities*

$$\ell^{\mathsf{ch}}(\delta')\, a^{\mathsf{ch}}\, \mathrm{d}^{\mathsf{ch}}_{\mathsf{av}}(\Lambda,\Gamma) \leq \mathop{\mathbb{E}}_{V\sim\nu} \mathop{\mathbb{E}}_{U\sim\nu} \mathsf{TV}(\mathbf{p}^{\Lambda,\psi_V,U}, \mathbf{p}^{\Gamma,\psi_V,U}) \leq u^{\mathsf{ch}}(\delta')\, A^{\mathsf{ch}}\, \mathrm{d}^{\mathsf{ch}}_{\mathsf{av}}(\Lambda,\Gamma)\,, \tag{3.27}$$

where we defined the quantum average-case distance between channels

$$\mathrm{d}^{\mathsf{ch}}_{\mathsf{av}}(\Lambda,\Gamma) \coloneqq \frac{1}{2}\sqrt{\|\mathcal{J}_\Lambda - \mathcal{J}_\Gamma\|^2_{\mathsf{HS}} + \mathrm{tr}\left((\Lambda-\Gamma)[\tau_d]^2\right)}, \tag{3.28}$$

and $a^{\mathsf{ch}} = 0.087$, $A^{\mathsf{ch}} = \frac{d}{d+1} \leq 1$, $\ell^{\mathsf{ch}}(\delta') = \frac{\left(1-\frac{\delta'}{d^2}\right)^3}{1+\delta'}$, $u^{\mathsf{ch}}(\delta') = 1 + \frac{\delta'}{d^2}$, $\tau_d \coloneqq \frac{\mathbb{I}}{d}$.

Proof. In what follows we prove a version of a theorem for exact 4-designs (i.e., setting $\delta = 0$; the proof for approximate 4-designs can be found in Appendix A.2.2). In order to simplify the notation we will use the notation $\Delta \coloneqq \Lambda - \Gamma$ (note that Δ is a superoperator and has a different meaning than Δ used in the proof of Theorem 1). We will make use of the Theorem 1 which implies that for fixed $\psi_V \in \mathcal{S}(\mathcal{H})$ the following inequalities hold

$$\frac{a}{2}\|\Delta[\psi_V]\|_{\mathsf{HS}} \leq \mathop{\mathbb{E}}_{U\sim\nu} \mathsf{TV}(\mathbf{p}^{\Lambda,\psi_V,U}, \mathbf{p}^{\Gamma,\psi_V,U})\,, \tag{3.29}$$

$$\frac{A}{2}\|\Delta[\psi_V]\|_{\mathsf{HS}} \geq \mathop{\mathbb{E}}_{U\sim\nu} \mathsf{TV}(\mathbf{p}^{\Lambda,\psi_V,U}, \mathbf{p}^{\Gamma,\psi_V,U})\,. \tag{3.30}$$

In what follows we prove bounds on $\mathop{\mathbb{E}}_{V\sim\nu} \|\Delta[\psi_V]\|_{\mathsf{HS}} = \mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}} \|\Delta[\psi]\|_{\mathsf{HS}}$. We first establish the upper bound by employing Jensen's inequality

$$\mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}} \|\Delta[\psi]\|_{\mathsf{HS}} \leq \sqrt{\mathop{\mathbb{E}}_{\psi\sim\nu_{\mathcal{S}}} \mathrm{tr}(\Delta[\psi]^2)}\,. \tag{3.31}$$

The average of $\mathrm{tr}(\Delta[\psi]^2)$ can be computed explicitly using the 2-design property and Lemma 3. We first rewrite using the same trick as in the proof of Corollary 1

$$\mathrm{tr}\left(\Delta[\psi]^2\right) = \mathrm{tr}\left(\Delta[\psi]^{\otimes 2}\mathbb{S}\right) \tag{3.32}$$

where $\mathcal{S} = \sum_{i,j=1}^{d} |i\rangle\langle j| \otimes |j\rangle\langle i|$ is the swap operator acting on $\mathcal{H}_d^{\otimes 2}$. Inserting the above into Lemma 3 yields

$$\frac{2}{d(d+1)}\,\mathrm{tr}\left(\mathbb{S}\,\Delta^{\otimes 2}[\mathbb{P}^{(2)}_{\mathsf{sym}}]\right) = \frac{2}{d(d+1)}\frac{1}{2}\left(\mathrm{tr}\left(\mathbb{S}\,\Delta^{\otimes 2}[\mathbb{I}]\right) + \mathrm{tr}\left(\mathbb{S}\,\Delta^{\otimes 2}[\mathbb{S}]\right)\right)\,, \tag{3.33}$$

where we used the identity $\mathbb{P}^{(2)}_{\mathrm{sym}} = \frac{1}{2}(\mathbb{I}\otimes\mathbb{I}+\mathbb{S})$. We now rewrite the first term as $\mathrm{tr}\left(\mathbb{S}\Delta^{\otimes 2}[\mathbb{I}]\right) = \mathrm{tr}\left(\Delta[\mathbb{I}]^2\right)$. Inserting the above into the integral (with multiplication and division by d^2) gives

$$\mathop{\mathbb{E}}_{\psi\sim\nu_S}\mathrm{tr}(\Delta[\psi]^2) = \frac{d^2}{d(d+1)}\left(\mathrm{tr}\left(\Delta\left[\frac{\mathbb{I}}{d}\right]^2\right)+\mathrm{tr}\left(\mathcal{S}\Delta^{\otimes 2}\left[\frac{\mathcal{S}}{d^2}\right]\right)\right) . \tag{3.34}$$

Recall that $\mathcal{J}_\Delta = (\mathbb{I}\otimes\Delta)(\Phi_+)$, where $|\Phi_+\rangle = \frac{1}{\sqrt{d}}\sum_{i=1}^d |i\rangle|i\rangle$ is the maximally entangled state in $\mathcal{H}_\mathrm{d}\otimes\mathcal{H}_\mathrm{d}$. Explicit computation gives $\|\mathcal{J}_\Delta\|^2_{\mathrm{HS}} = \mathrm{tr}\left(\mathcal{S}\Delta^{\otimes 2}\left[\frac{\mathcal{S}}{d^2}\right]\right)$.

$$\mathop{\mathbb{E}}_{\psi\sim\nu_S}\mathrm{tr}(\Delta[\psi]^2) = \frac{d^2}{d(d+1)}\left(\mathrm{tr}\left(\Delta\left[\frac{\mathbb{I}}{d}\right]^2\right)+\|\mathcal{J}_\Delta\|^2_{\mathrm{HS}}\right) . \tag{3.35}$$

Inserting this expression into (3.31) and using (3.30) finally gives the upper bound in Eq. (3.27).

To prove the lower bound integrate both sides of (3.29) and apply Berger's inequality for $X_\psi = \|\Delta[\psi]\|_{\mathrm{HS}} = \sqrt{\mathrm{tr}(\Delta[\psi]^2)}$

$$\mathop{\mathbb{E}}_{\psi\sim\mu_S}\|\Delta[\psi]\|_{\mathrm{HS}} \geq \frac{\left(\mathop{\mathbb{E}}_{\psi\sim\mu_S}\mathrm{tr}(\Delta[\psi]^2)\right)^{3/2}}{\left(\mathop{\mathbb{E}}_{\psi\sim\mu_S}\mathrm{tr}(\Delta[\psi]^2)^2\right)^{1/2}} . \tag{3.36}$$

We proceed by rewriting the integral in the denominator of the above expression

$$\mathop{\mathbb{E}}_{\psi\sim\nu_S}\mathrm{tr}(\Delta[\psi]^2)^2 = 4\mathop{\mathbb{E}}_{\psi\sim\nu_S}\mathrm{tr}(\Delta[\psi]^{\otimes 4}\mathbb{P}^{(2)}_{\mathrm{sym}}\otimes\mathbb{P}^{(2)}_{\mathrm{sym}}) . \tag{3.37}$$

where we used the identity $2\,\mathrm{tr}\,\mathbb{P}^{(2)}_{\mathrm{sym}} X\otimes Y = \mathrm{tr}\,X\,\mathrm{tr}(Y)+\mathrm{tr}(XY)$ and $\mathrm{tr}(\Delta[\psi]) = 0$. By expanding $\Delta[\psi]^{\otimes 4} = \Delta^{\otimes 4}\left[\psi^{\otimes 4}\right]$ and integrating over ψ (cf. Lemma 3) we obtain

$$\mathop{\mathbb{E}}_{\psi\sim\nu_S}\mathrm{tr}(\Delta[\psi]^2)^2 = \frac{4}{\binom{d+3}{4}}\mathrm{tr}\left(\Delta^{\otimes 4}\left[\mathbb{P}^{(4)}_{\mathrm{sym}}\right]\mathbb{P}^{(2)}_{\mathrm{sym}}\otimes\mathbb{P}^{(2)}_{\mathrm{sym}}\right) . \tag{3.38}$$

Substituting $\mathbb{P}^{(2)}_{\mathrm{sym}} = \binom{d+1}{2}\mathop{\mathbb{E}}_{\psi\sim\nu_S}\psi^{\otimes 2}$ we get

$$\mathop{\mathbb{E}}_{\psi\sim\nu_S}\mathrm{tr}(\Delta[\psi]^2)^2 = \frac{4\binom{d+1}{2}^2}{\binom{d+3}{4}}\mathop{\mathbb{E}}_{\psi\sim\nu_S}\mathop{\mathbb{E}}_{\varphi\sim\nu_S}\mathrm{tr}\left(\mathbb{P}^{(4)}_{\mathrm{sym}}\Delta^\dagger[\psi]^{\otimes 2}\otimes\Delta^\dagger[\varphi]^{\otimes 2}\right) . \tag{3.39}$$

We now utilize Lemma 5 to upper bound the function inside the integral

$$\mathrm{tr}\left(\mathbb{P}^{(4)}_{\mathrm{sym}}\Delta^\dagger[\psi]^{\otimes 2}\otimes\Delta^\dagger[\varphi]^{\otimes 2}\right) \leq C\,\mathrm{tr}(\mathbb{P}^{(2)}_{\mathrm{sym}}\Delta^\dagger[\psi]^{\otimes 2})\,\mathrm{tr}(\mathbb{P}^{(2)}_{\mathrm{sym}}\Delta^\dagger[\varphi]^{\otimes 2}) , \tag{3.40}$$

where $C = \frac{13}{6}$. Inserting this into (3.39) and carrying over the integrals over ψ and ϕ (with the help of Corollary 1) we get

$$\mathop{\mathbb{E}}_{\psi\sim\nu_S} \mathrm{tr}(\Delta[\psi]^2)^2 \leq \frac{4C}{\binom{d+3}{4}} \mathrm{tr}\left(\mathbb{P}^{(2)}_{\mathrm{sym}} \Delta^{\otimes 2}\left[\mathbb{P}^{(2)}_{\mathrm{sym}}\right]\right)^2 . \tag{3.41}$$

We now calculate

$$\mathrm{tr}\left(\mathbb{P}^{(2)}_{\mathrm{sym}} \Delta^{\otimes 2}\left[\mathbb{P}^{(2)}_{\mathrm{sym}}\right]\right) = \binom{d+1}{2} \mathop{\mathbb{E}}_{\psi\sim\nu_S} \mathrm{tr}\left(\mathbb{P}^{(2)}_{\mathrm{sym}} \Delta[\psi]^{\otimes 2}\right) = \tag{3.42}$$

$$= \frac{1}{2}\binom{d+1}{2} \mathop{\mathbb{E}}_{\psi\sim\nu_S} \mathrm{tr}\left(\Delta[\psi]^{\otimes 2}\mathbb{S}\right) = \tag{3.43}$$

$$= \frac{1}{2}\binom{d+1}{2} \mathop{\mathbb{E}}_{\psi\sim\nu_S} \mathrm{tr}\left(\Delta[\psi]^2\right) . \tag{3.44}$$

In the first equality, we used the fact that since ν_S forms a 2-design we can substitute the projector onto 2-fold symmetric subspace with corresponding (renormalized) average. In the second equality, we exploited the fact that states of the type $\psi^{\otimes 2}$ are in an invariant subspace of $\mathbb{P}^{(2)}_{\mathrm{sym}}$. Third equality follows directly from Corollary 1 and fact that $\mathrm{tr}(\Delta[\psi]) = 0$. Combining (3.41) and (3.42) we obtain

$$\mathop{\mathbb{E}}_{\psi\sim\nu_S} \mathrm{tr}(\Delta[\psi]^2)^2 \leq \frac{C\binom{d+1}{2}^2}{\binom{d+3}{4}} \left(\mathop{\mathbb{E}}_{\psi\sim\nu_S} \mathrm{tr}\left(\Delta[\psi]^2\right)\right)^2 . \tag{3.45}$$

Inserting the above bound into (3.36) gives

$$\mathop{\mathbb{E}}_{\psi\sim\nu_S} \|\Delta[\psi]\|_{\mathrm{HS}} \geq b_d \sqrt{\mathop{\mathbb{E}}_{\psi\sim\nu_S} \mathrm{tr}(\Delta[\psi]^2)} , \tag{3.46}$$

with $b_d = \frac{1}{\sqrt{13}}\sqrt{\frac{(d+2)(d+3)}{d(d+1)}}$. Integrating both sides of (3.29) over $\psi \sim \nu_S$ and using the the above inequality we finally obtain

$$a^{\mathrm{ch}}\, \mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda, \Gamma) \leq \mathop{\mathbb{E}}_{V\sim\nu} \mathop{\mathbb{E}}_{U\sim\nu} \mathrm{TV}(\mathbf{p}^{\Lambda,\psi_V,U}, \mathbf{p}^{\Gamma,\psi_V,U}) , \tag{3.47}$$

with $a^{\mathrm{ch}} = a \cdot b_{\mathrm{d}} \approx 0.087$. □

Remark 8. *We note that one can view quantum states and measurements as special types of quantum channels. While for state preparation channels the operational procedure of discrimination is equivalent and one gets the same expression (see Example 18), for measurements it is not the case . Moreover, in the case of δ-approximate designs, applying the above Theorem 3 for the average-case distance between state preparation channels gives worse than Theorem 1 constants and functional dependence on δ. This approach thus leads to less tight bounds for states than treating them separately.*

Remark 9. *We note that while the dependence of δ on the dimension of the system d is very high in Theorems 1, 2 and 3, it does not pose a practical problem. Indeed, exponentially accurate δ-approximate unitary designs can be implemented already with logarithmic-depth quantum circuits [54].*

3.4.4 Consequences

The theorems proven in this section suggest defining average-case distances between quantum states, measurements, and channels via formulas $\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}$, $\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}$, $\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}$ appearing in approximations (3.16), (3.24), and (3.27). This approach has several pleasant consequences. First, functions describing these distances can be expressed via simple, degree-two polynomials in underlying objects and can be easily explicitly computed for objects acting on systems of moderate dimension (no optimization is needed as in the case of the diamond norm [47]). Second, all average-case distances utilize in some way the Hilbert-Schmidt norm. This gives this norm an operational interpretation it did not possess before (especially for quantum states for which $\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho,\sigma) = \frac{1}{2}\|\rho-\sigma\|_{\mathrm{HS}}$). Third, it turns out that so-defined distances satisfy plethora of natural properties such as subadditivity: $\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho_1\otimes\rho_2,\sigma_1\otimes\sigma_2) \leq \mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho_1,\sigma_1) + \mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho_2,\sigma_2)$, joint convexity: $\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\sum_\alpha p_\alpha\rho_\alpha,\sum_\alpha p_\alpha\sigma_\alpha) \leq \sum_\alpha p_\alpha \mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho_\alpha,\sigma_\alpha)$, or restricted data-processing inequalities (typically various distances d_{av} are non-increasing under application of unital quantum channels). Fourth, while it may seem that condition of being (approximate) 4-design is quite stringent, from a very recent paper [54] it follows that ensembles of quantum circuits required by Theorems 1-3 can be realized by random circuits in the 1D brickwork architecture in depth $O(\log(N))$ (with moderate prefactors) [54]. Finally, we expect that our average-case distances will more accurately capture the behavior of errors in the performance of quantum objects in generic moderate size quantum algorithms (note that many architectures of variational circuits used in NISQ algorithms are expected to exhibit, on average, design-like behavior [67]). We back up this last claim numerically by testing the usefulness of our distance measures on families of random quantum circuits originating from random instances of variational quantum algorithms on few-qubit systems.

For all the reasons mentioned above, we believe that introduced distances will prove useful in analyzing the practical performance of near-term quantum processors. We expect that they can also be useful in other branches of quantum information requiring the usage of randomized protocols like quantum communication, quantum complexity theory, or quantum machine learning. In the next section we analyze simple examples that illustrate potential usefulness of our results.

3.5 Main applications

3.5.1 Application 1: Noise in quantum advantage experiments.

Here we consider examples which help to understand how noise affects average probability distributions in experiments with random circuits sampling. First, AC distances between noisy and ideal state allow to lower-bound average TVDs between target and noisy distributions. Second, AC distances allow to upper-bound average-case TVD between noisy distribution and trivial (uniform) one. Indeed, to bound average TVD between uniform and noisy distribution, one cal-

culates AC distance to maximally mixed state $\frac{\mathbb{I}}{d}$ (states), trivial POVM $\mathbf{M}^{\mathcal{I}} = (\frac{\mathbb{I}}{d}, \ldots, \frac{\mathbb{I}}{d})$ (measurements), or maximally depolarizing channel Λ_{dep} that acts as $\Lambda_{\text{dep}}(\rho) = \frac{\mathbb{I}}{d}$ for any state ρ (channels). This follows directly from definitions of AC distances – see Lemmas 23, 24, 25 and discussion in 3.7.

In what follows, most of the examples make use of some average noise parameter q^{av} (with different meaning for each example) that describes an average (over qubits) probability of errors of considered type *not* occurring. In most of them, we make an assumption that $q^{av} \leq \sqrt[N]{\frac{1}{2}}$, with N being the number of qubits. This is done solely to achieve a particularly appealing form of lower bounds. One can derive expressions that are more complicated and do not require this assumption (see Section 3.7.3 for details). In general, since $\sqrt[N]{\frac{1}{2}} \xrightarrow{N\to\infty} 1$, the assumption becomes less restrictive for higher-dimensional systems and the presented bounds are intended for use in such cases. In this section, we state the examples and discuss their consequences, while proofs are relegated to Section 3.7.3.

Example 1 (Pauli eigenstates and tensor product Pauli noise). *Consider state $\psi^{pauli} = \otimes_{i=1}^{N} |{\pm}r_i\rangle\langle{\pm}r_i|$, where $r_i \in \{x, y, z\}$, i.e., $|{\pm}r_i\rangle$ is any Pauli eigenstate on qubit i (with eigenvalue $+1$ or -1.). Consider tensor product Pauli channel $\Lambda^{pauli} = \otimes_{i=1}^{N} \Lambda_i^{pauli}$, where single-qubit channel is $\Lambda_i^{pauli}(\rho) = \sum_{j=1} p_j^{(i)} \sigma_j \rho \sigma_j$ with $j \in \{1, x, y, z\}$, $\sigma_1 = \mathbb{I}$, and $p_j^{(i)} \geq 0$, $\sum_j p_j^{(i)} = 1$. Define $q^{(i)} = p_1^{(i)} + p_{r_i}^{(i)}$, i.e., a probability of applying on qubit i a gate that stabilizes the state of that qubit (namely, either identity or Pauli matrix of which $|{\pm}r_i\rangle$ is an eigenstate). Define average properties of noise as $q^{av} = \frac{1}{N}\sum_{i=1}^{N} q^{(i)}$ and $f^{av} = \frac{1}{N}\sum_{i=1}^{N} q^{(i)}(1 - q^{(i)})$. Assume $q^{(i)} \geq \frac{1}{2}$ for each qubit and that $q^{av} \leq \sqrt[N]{\frac{1}{2}}$. Then we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\Lambda^{pauli}(\psi^{pauli}), \frac{\mathbb{I}}{d}) < \frac{1}{2}\exp\left(-2f^{av}\, N\right), \tag{3.48}$$

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\Lambda^{pauli}(\psi^{pauli}), \psi^{pauli}) > \frac{1}{2}\sqrt{1 - 2(q^{av})^N}, \tag{3.49}$$

The above example might be relevant, for example, in QAOA algorithms where input state is often indeed a tensor product Pauli state [57], or can be useful for estimating effects of state-preparation errors for standard setting where input state is $|0\rangle\langle 0|^{\otimes N}$. We see that with growing system size, the average noisy distribution approaches uniform distribution exponentially quickly (while moving away from target distribution).

This demonstrates that even in the absence of noise in random unitaries, the state-preparation errors will quickly aggregate. Exactly the same behaviour is demonstrated for the following simplified measurement noise model.

Example 2 (Symmetric bitflip measurement noise). *Consider a symmetric, uncorrelated stochastic measurement noise map $\mathrm{T}^{sym} = \otimes_{i=1}^{N} \mathrm{T}_i^{sym}$, where for each*

qubit we have $\mathsf{T}_i^{sym} = p^{(i)}\mathbb{I} + (1-p^{(i)})\sigma_x$, *with* $(1-p^{(i)})$ *being a bitflip error probability on* i*th qubit.. It follows that the* k*th effect of a noisy version* $\mathsf{T}^{sym}\mathbf{P}$ *of computational basis measurement* $\mathbf{P}$ *is given by* $(\mathsf{T}^{sym}\mathbf{P})_k = \sum_l \mathsf{T}^{sym}_{kl} |l\rangle\langle l|$. *Define* $f^{av} = \frac{1}{N}\sum_{i=1}^{N} p^{(i)}(1-p^{(i)})$. *Assume* $p^{(i)} \geq \frac{1}{2}$ *for each qubit. Then we have*

$$\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{T}^{sym}\mathbf{P}, \mathbf{M}^{\mathcal{I}}) < \frac{1}{2}\exp\left(-2f^{av}\ N\right)\ , \tag{3.50}$$

The above means that even in the absence of state-preparation and gate errors, for symmetric bitflip noise the resulting average distribution exponentially quickly converges to uniform. We now consider a distance from ideal measurement for more realistic case of generic tensor product measurement noise.

Example 3 (Generic tensor product measurement noise). *Let* $\mathbf{P} = (|\mathbf{x}\rangle\langle\mathbf{x}|)_{\mathbf{x}\in\{0,1\}^N}$ *be a computational basis measurement on* N *qubit system. Let* $\mathbf{M} = (M_{\mathbf{x}})_{\mathbf{x}\in\{0,1\}^N}$ *be a POVM specified by effects* $M_{\mathbf{x}} = \Lambda_1^{\dagger}(|x_1\rangle\langle x_1|) \otimes \ldots \otimes \Lambda_N^{\dagger}(|x_N\rangle\langle x_N|)$, *where* Λ_i *are quantum channels affecting* i*'th qubit, and* $\Lambda_i^{\dagger}$ *is the conjugate of* Λ_i. *Define classical success probability as* $p^{(i)}(\mathbf{x}_i|\mathbf{x}_i) = \mathrm{tr}\left(\Lambda_i^{\dagger}\left(|x_i\rangle\langle x_i|\right)|x_i\rangle\langle x_i|\right)$ *and corresponding average* $q^{(i)}_{av} = \frac{p^i(0|0)+p^{(i)}(1|1)}{2}$. *Let* $q^{av} := \frac{1}{N}\sum_{i=1}^{N} q^{(i)}_{av}$. *Assume that for each qubit* $q^{(i)}_{av} \geq \frac{1}{2}$ *and that* $q^{av} \leq \sqrt[N]{\frac{1}{2}}$. *Then we have*

$$\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathbf{M}, \mathbf{P}) > \frac{1}{2}\sqrt{1-2(q^{av})^N}\ . \tag{3.51}$$

The quantity q^{av} is the survival probability of classical single-qubit state $|x_i\rangle\langle x_i|$ that goes through a channel Λ_i, averaged over all qubits and input states. We note that those quantities are routinely reported in experimental works, which makes the above bound particularly useful. Indeed, data from recent quantum advantage experiments [21; 98] suggests that q^{av} is around 97% (we take average of values reported in both papers). Assume perfect gates, no state preparation errors and $q^{av} = 0.97$. Furthermore, assume that random circuits used in experiments form approximate 4-designs (this assumption is consistent with results of [106]). Then from Theorem 2 it follows that if readout errors remain constant with scaling of the system, for a 54-qubit quantum computer, on average (over realizations of random quantum circuits) output distributions $\mathbf{p}^{\mathbf{M},\psi_V}$ will have a constant ≈ 0.13 TV-distance from the ideal probability distributions $\mathbf{p}^{\mathbf{P},\psi_V}$ solely due to effects of readout noise.

Example 4 (Separable Pauli noise in the middle of the circuit). *Consider tensor product Pauli channel* $\Lambda^{pauli} = \otimes_{i=1}^{N}\Lambda_i^{pauli}$, *where single-qubit channel is* $\Lambda_i^{pauli}(\rho) = \sum_{j=1}\ p_j^{(i)}\sigma_j\rho\sigma_j$ *with* $j \in \{1, x, y, z\}$, $\sigma_1 = \mathbb{I}$, *and* $p_j^{(i)} \geq 0$, $\sum_j p_j^{(i)} = 1$; *the same noise model as in Example 4. For each qubit* i *define* $||\mathbf{p}^{(i)}||_2^2 = \sum_j \left(p_j^{(i)}\right)^2$, *and corresponding average* $p_2^{av} = \frac{1}{N}\sum_{i=1}^{N}||\mathbf{p}^{(i)}||_2^2$, *as well as average*

(over qubits) probability of application of identity channel $p_1^{av} = \frac{1}{N}\sum_{i=1} p_1^{(i)}$ *(note that* $p_1^{(i)}$ *corresponds to probability of no errors on qubit* i*). Assume* $p_1^{av} \leq \sqrt[N]{\frac{1}{2}}$*. Then we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda^{pauli}, \Lambda_{dep}) < \frac{1}{2}\exp\left(-p_2^{av}\ N\right)\ , \tag{3.52}$$

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda^{pauli}, \mathcal{I}) > \frac{1}{\sqrt{2}}\sqrt{1-2\left(p_1^{av}\right)^N}\ . \tag{3.53}$$

Recall that the above scenario corresponds to inserting local Pauli noise "between" two random circuits (two averages in Eq. (3.27)). Similarly to previous cases, whenever there is non-zero noise, we will observe an exponential convergence to the trivial distribution and high separation from ideal distribution corresponding to identity channel $\mathcal{I}$.

Example 5 (Single Pauli error the middle of the circuit). *Consider single-qubit channel* $\Lambda_\sigma^{(i)}$ *that applies some traceless unitary* σ *on qubit* i *(and identity to all other qubits). Then we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_\sigma^{(i)}, \mathcal{I}) = \frac{1}{\sqrt{2}}\ . \tag{3.54}$$

Physically, the above may correspond to a unitary noise applying one of Pauli matrices on qubit i somewhere in the circuit. We then observe a constant separation (value of $\frac{1}{\sqrt{2}}$) between ideal distribution and the noisy distribution. Such significant average distance between noisy and target distribution suggests that local strong coherent errors can dramatically affect the performance of a given device in typical circumstances. This result is in agreement with empirical observations made in Refs. [107; 21] where single-qubit errors were causing "speckle pattern" of output bitstrings probabilities to break, resulting in very low cross-entropy benchmarking fidelity.

3.5.2 Application 2: Sample efficient distinguishability of quantum objects with incoherent access

Example 6. *For any pure state* ψ *and maximally-mixed state* $\tau_d := \frac{\mathbb{I}}{d}$ *on* $\mathcal{H}_d$*, we have* $\mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\psi, \tau_d) = \frac{1}{2}\sqrt{1-\frac{1}{d}}$.

It follows that a single round of a randomized protocol implicit in the definition of $\mathrm{d}_{\mathrm{av}}^{\mathrm{s}}$ (cf. Remark 2), realized via approximate 4-design and computational basis measurements, gives a constant bias in distinguishing *any* pure N-qubit state ψ from the maximally mixed state: $p_{\mathrm{succ}}^{\mathrm{av}} \gtrsim 0.57$. This probability can be made arbitrarily close 1 by repeating the protocol and using the majority-vote strategy. Importantly, this method *does not* utilize the coherent access or a quantum memory (in a sense defined, e. g., in [99; 108]). We note that a related

but distinct scenario is considered in Ref. [99]. There, the authors introduced the task of `PurityTesting` corresponding to discrimination between *unknown* Haar-random pure random state and maximally mixed state. For N qubit systems, Theorem 4 of [99] implies exponential lower bound for the query complexity k (number of usages of unknown quantum state) needed to succeed in this task, given incoherent access to objects in question. In contrast, our randomized measurement protocol gives high statistical distinguishability already for a single query *for all* states ψ. The difference comes from the fact that in the scenario considered in Example 6 the random state is arbitrary but known.

Example 7. *Let Λ_U be a a unitary channel corresponding to a unitary U on $\mathcal{H}_d$ and let Λ_{dep} be a depolarizing channel i.e. $\Lambda_{\text{dep}}(\rho) = \tau_d$ for any ρ. Then we have* $\mathrm{d}_{\text{av}}^{\text{ch}}\left(\Lambda_U, \Lambda_{\text{dep}}\right) = \frac{1}{2}\sqrt{1 - \frac{1}{d^2}}$.

In related task `FixedUnitary` studied in [99], one is asked to distinguish *unknown* Haar-random unitary channel Λ_U from Λ_{dep}. Exponential query complexity lower bound incoherent protocols was shown in [99]. By repeating analogous reasoning as for states, we get that when Λ_U is arbitrary but known, randomized, non-adaptive, and incoherent protocol, utilizing two realizations of approximate 4-designs, gives constant bias in success probability of discrimination of Λ_U from Λ_{dep} using just a single query.

3.5.3 Application 3: Strong complexity of quantum states and unitaries.

The above examples have interesting consequences for the notion of a strong state and unitary complexity investigated in [55]. There, the authors defined complexity C_Δ of N-qubit pure state ψ (resp. unitary circuit Λ_U) as the number of elementary gates needed to construct a circuit necessary to implement a *two-outcome* measurement discriminating between ψ (resp. Λ_U) and the maximally mixed state (resp. maximally depolarizing channel Λ_{dep}) with success probability $p_{\text{succ}} = \frac{1}{2} + \Delta$. Our results imply that if the requirement of two-outcome measurement is relaxed and one accepts the use of *classical* randomness, then measurements realizable with circuit depths $r = O\left(\log(N)\right)$ (i.e., unitary 4-designs from Ref. [54]) can succeed in these discrimination tasks with a constant bias Δ_* *for all* states ψ and unitary channels Λ_U. In particular, for dimension d, the average-case distance between any pure quantum state and the maximally mixed state is equal to $\frac{1}{2}\sqrt{1 - \frac{1}{d}}$ (see Lemma 23), while the AC distance between any unitary channel Λ_U and maximally depolarizing channel is equal to $\frac{1}{2}\sqrt{1 - \frac{1}{d^2}}$ (see Lemma 25). Therefore, in both cases, the bias corresponding to a randomized measurement strategy is equal to $\Delta_* = \frac{1}{4}$ (recall discussion of Total-Variation Distance in Section 2.4). This renders the so-defined notion of complexity trivial - all states and unitaries will have complexity $C_\Delta \leq \log(N)$, unless bias Δ satisfies $\Delta > \frac{1}{4}$ (independent on the system size). Note that the gate complexity of implementing the distinguishing measurement is very small here due to the use

of classical randomness. Indeed, if we didn't allow for classical randomness, but instead implemented the measurement directly via, e.g., Naimark's dilation [101], it could, in general, require higher number of gates (see also Remark 2).

We note that large average-case distance d_{av} implies only information-theoretic distinguishability of quantum objects. The cost of classical post-processing needed to distinguish the probability distributions resulting from randomized protocols can be very large since they operate on exponentially large sample space.

3.5.4 Numerical results.

Here we present results of numerical studies of small-size quantum systems. We compare scaling with the system size for worst-case distance, average-case distance, and a mean TVD taken over ensemble of unitaries. The mean Total-Variation distance is calculated numerically over three types of ensembles of unitaries with a structure of variational circuits. One ensemble has QAOA-like structure, while the other two are standard hardware-efficient VQE ansatze [60]. More specifically, in each case, the p-layer circuit can be written as

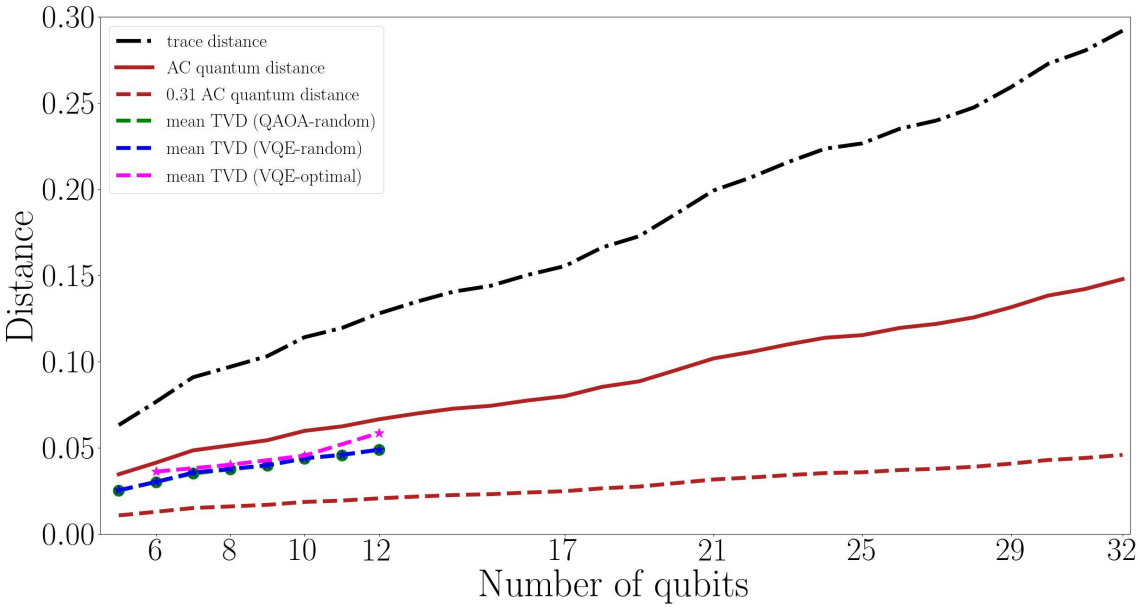
$$U_p = \prod_{j=1}^{p} U_{\mathrm{rot},j}\, U_{\mathrm{ent},j}\ . \tag{3.55}$$

where $U_{\mathrm{rot},j}$ is a "rotation block" and U_{ent} is an "entangling block". Exact form of the evolution, as well as the initial state depends on the ensemble. We consider three such ensembles:

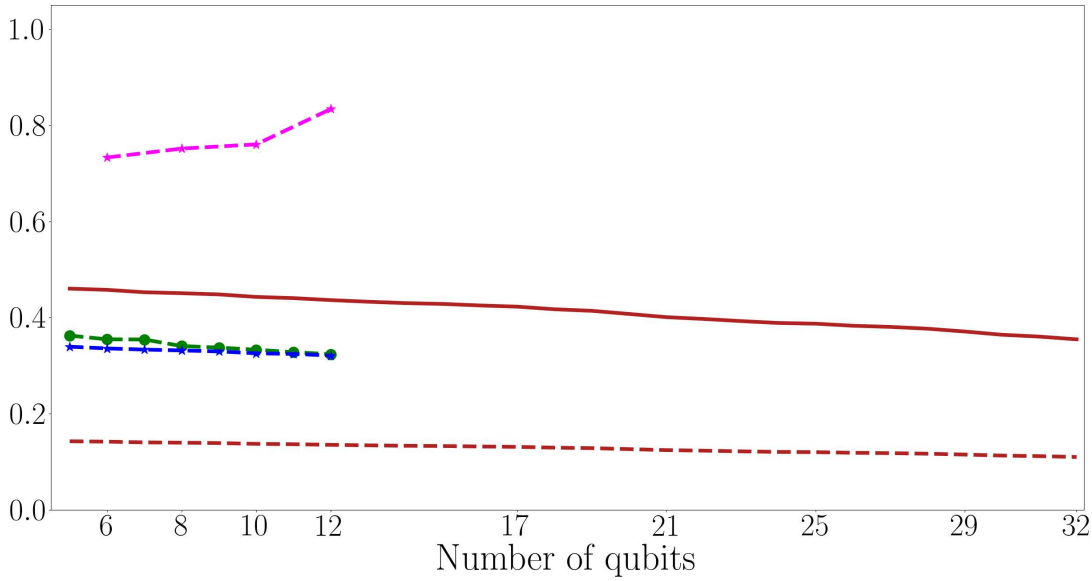
(a) Circuits that originate from QAOA instance for fixed Hamiltonian $\mathsf{H}_{\mathsf{2SAT}}$ encoding fixed (random) instance of random MAX-2-SAT problem [109]. In this case, the initial state is of the form $|+\rangle^{\otimes N}$ with $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, while unitary evolution is given by $U_{\mathrm{rot,j}} := U_{\alpha_j} = \exp\left(-i\alpha_j \sum_{k=1}^{N} \sigma_x^{(k)}\right)$, and $U_{\mathrm{ent},j} := U_{\beta_j} = \exp\left(-i\beta_j \mathsf{H}_{\mathsf{2SAT}}\right)$, with $\sigma_x^{(k)}$ being X gate on kth qubit. For each j, α_j and β_j are N-dimensional vectors of parameters chosen randomly from range $[-\pi, \pi]$.

(b) Circuits of a form of generic Hamiltonian-independent VQE ansatz with initial state being $|0\rangle^{\otimes N}$. We choose the rotation block to be of the form $U_{\mathrm{rot,j}} := U_{\alpha_j} = \bigotimes_{k=1}^{N} \exp\left(-i\alpha_{2j}\, \sigma_Z^{(k)}\right) \circ \exp\left(-i\alpha_{2j+1}\, \sigma_Y^{(k)}\right)$, where σ_Y, σ_Z are Y and Z gates. The entangling block is $U_{\mathrm{ent},j} = U_{\mathrm{ent}} := \prod_{k=1}^{N-1} \mathrm{CX}_{k,k+1}$ with $\mathrm{CX}_{k,l}$ denoting CX gate between qubits k and l. For each j, α_j is a $2N$-dimensional vector of parameters chosen randomly from range $[-\pi, \pi]$.

(c) The third ensemble is similar to the VQE-like ensemble (b), but now rotation block contains only of Y rotations. Furthermore, the angles are *not random*, but they are chosen from a fixed set of parameters that come from solutions of variational optimization. In other words, each used unitary corresponds

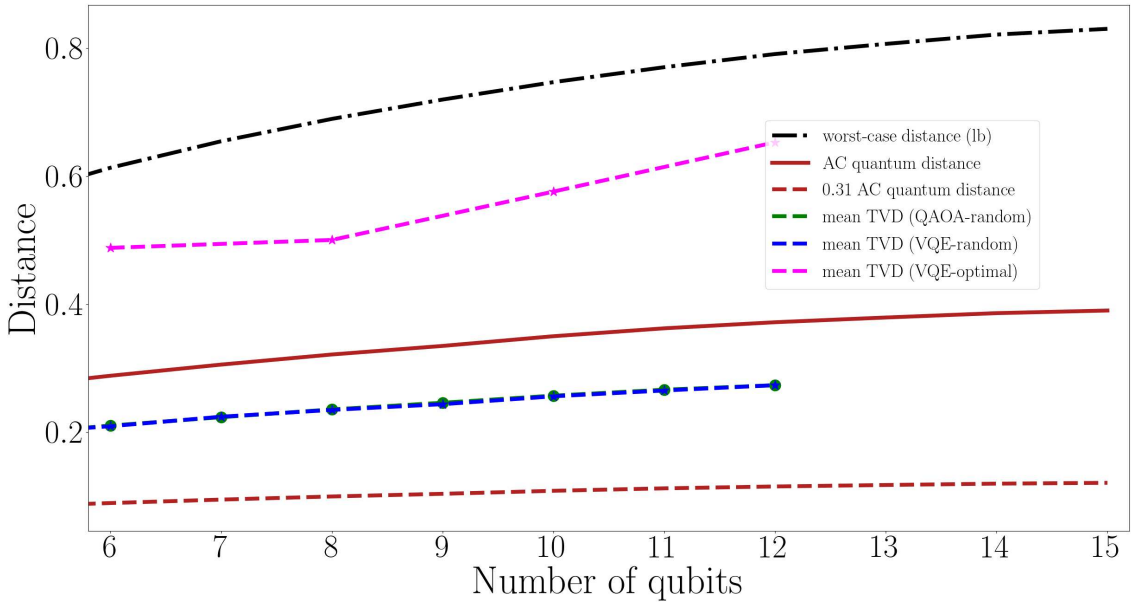


(a) *Quantum states, distance to ideal distribution*

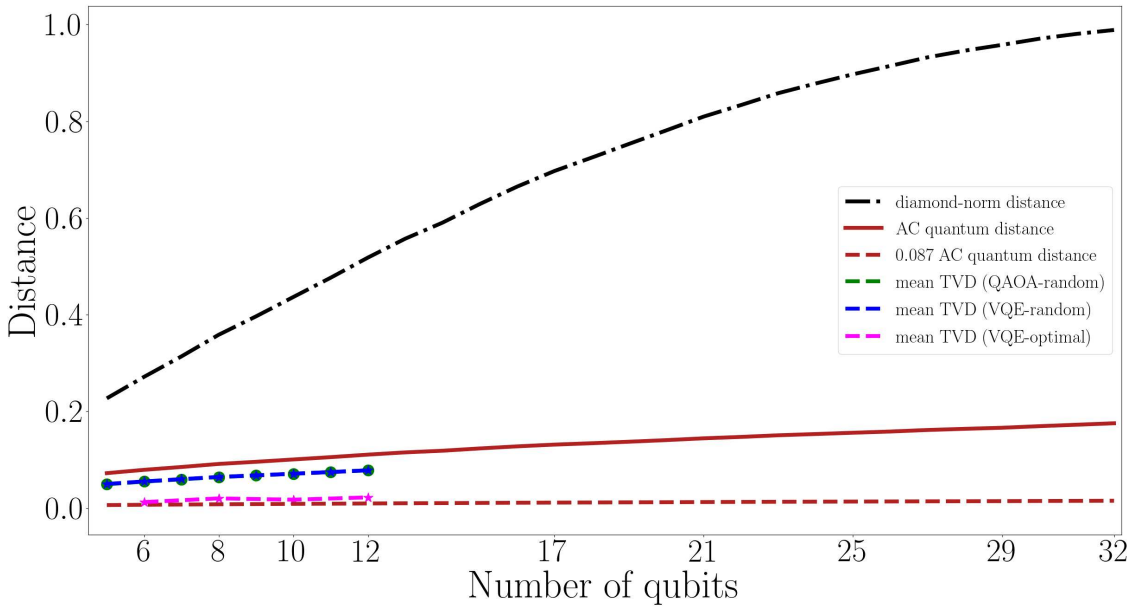


(b) *Quantum states, distance to uniform distribution*

Figure 3.3: *Results of numerical studies for comparison between worst-case distance, quantum average-case distance and numerically calculated mean TVD for quantum states. Figure 3.3a corresponds to distance to ideal (noiseless) distribution, while in Figure 3.3b, we plot distance to uniform (trivial) distribution. For average-case distance, we also plot value corresponding to lower bound on average-case TVD (following from Eqs.* (3.16), (3.24), (3.27)*). In case of worst-case distance, "lb" indicates lower-bound. Quantum average-case distances were calculated explicitly. Mean TVDs were calculated between (exact numerical) probability distributions over* 1000 *random instances of random unitaries.*

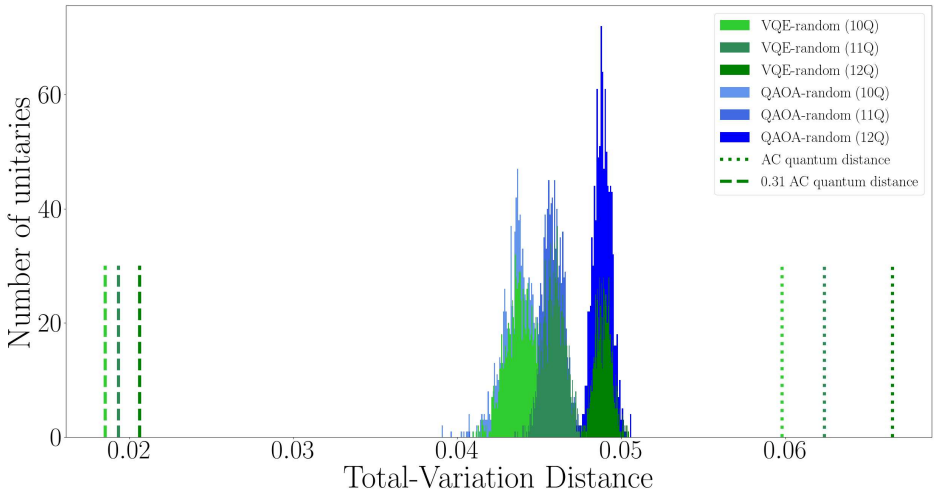


(a) *Quantum measurements, distance to ideal distribution*

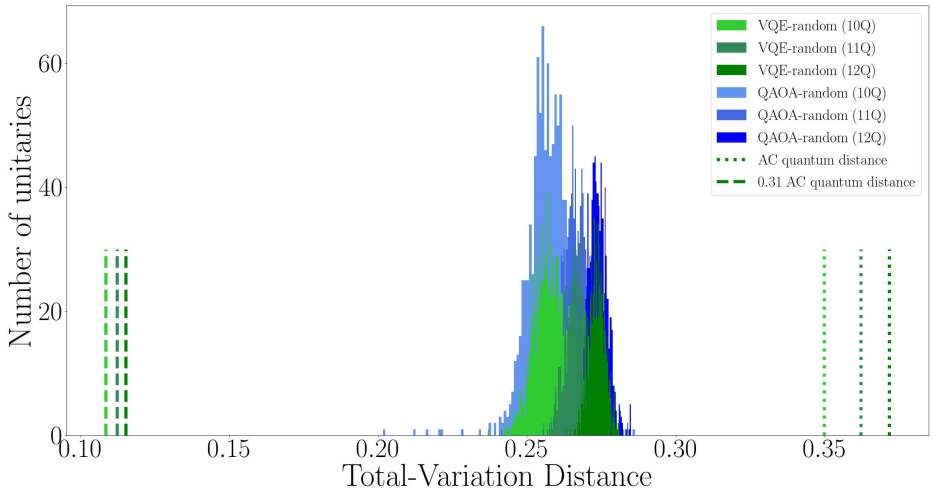


(b) *Quantum channels, distance to ideal distribution*

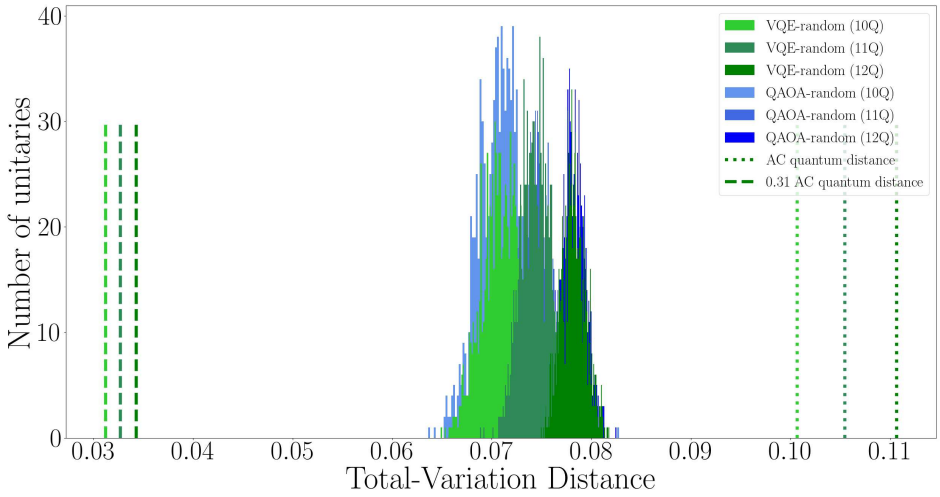
Figure 3.4: *Results of numerical studies for comparison between worst-case distance, quantum average-case distance and numerically calculated mean TVD. For both plots, we show the distance to ideal (noiseless) distribution. The data conventions are exactly the same as in Figure 3.3*



(a) *Quantum states, distance to ideal distribution*



(b) *Quantum measurements, distance to ideal distribution*



(c) *Quantum channels, distance to ideal distribution*

Figure 3.5: *Histograms of TVDs obtained for random ensembles considered in numerical simulations corresponding to Figs. 3.3, 3.4. Different shades of a given color (blue or green) correspond to different system sizes for a given ensemble (QAOA or VQE). Bounds from average-case distances are indicated via dashed lines and for each dimension are the same for both ensembles (they depend only on quantum objects in question, not on the choice of random ensemble).*

to a circuit that was found to be optimal in a VQE optimization (as opposed to uniformly random angles taken for both previous ensembles). We use datasets from Ref. [110] where authors developed an adaptive measurement scheme that improves performance of VQE.

Based on recent results [67], we expect the random ensembles (a) and (b) to form (approximate) unitary 4-designs. Ensemble of type (c), due to limited computational resources, consist of only $7-12$ unitaries (recall that generating each unitary requires performing full VQE optimization). This implies that this ensemble *does not* form even unitary 2-design. It is nevertheless still interesting to investigate its behaviour, since those are circuits of particular practical importance.

To calculate quantum AC distances, we consider the following scenarios:

1. (States) We compare a randomly chosen Pauli eigenstate affected by random local Pauli noise with its ideal version (Fig. 3.3a) and with maximally mixed state $\frac{\mathbb{I}}{d}$ (Fig. 3.3b). This is scenario considered in Example 1. The error probabilities are chosen randomly from range $[0.001, 0.01]$.

2. (Measurements) The noisy measurement is a tensor product POVM constructed from single-qubit measurements obtained via Quantum Detector Tomography [80] of IBM's 15-qubit Melbourne device. We compare it to ideal computational-basis measurement (Fig. 3.4a). Since the measurement noise in superconducting devices is usually highly asymmetric [65], we do not expect it to converge to uniform distribution.

3. (Channels) We compare channel corresponding to random tensor product of 1-qubit rotations around random axis with ideal identity channel $\mathcal{I}$ (Fig 3.4b). Explicitly, the unitary corresponding to the channel has a form $\bigotimes_{k=1}^{N} \exp(-i\gamma_k V^{(k)})$, where $V^{(k)}$ is chosen randomly to be X, Y or Z gate, and $\gamma_k \in [0.025\pi, 0.0313\pi]$. Similarly to POVMs, we do not expect coherent errors to bring noisy distributions close to uniform distribution.

In each case, number of circuit layers is $\lfloor 1.5N \rfloor$ for ensembles a) and b), while for ensemble c) it varied between instances. In Figs. 3.3, 3.4 we collectively present results of all simulations. Recall that the two random ensembles presented in Figs. 3.3, 3.4 consist of circuits that are variational QAOA and VQE circuits with random parameters. From the plots it is clear that in all studied cases for those ensembles, the quantum average-case distance is both significantly closer and more similar in scaling to mean Total Variation distance between distributions in question, as compared to worst-case distance. At the same time, for optimized VQE ensemble, we see that for distances between ideal and noisy distributions, the results are qualitatively similar to random ensembles in case of states and channels, but significantly different for quantum measurements. Recall that POVMs used to generate the plot are results of detector tomography of actual quantum device from IBM. In this case, the noise affects results so much, that empirical TVDs are closer to worst-case than to average-case bounds. In

case of distance between noisy and uniform distribution for states, we also observe that average-case distances do not capture well the behaviour of the distributions for unitaries obtained in VQE optimization. This is not surprising – those circuits are certainly not random.

In Fig 3.5 we present histograms of TVDs over random unitaries. The datapoints correspond to simulations presented in Figs. 3.3, 3.4. The plots show how the TVDs concentrate for small system sizes and demonstrate that all random points lie well within bounds provided by average-case distances.

3.6 Properties of distance measures

While expressions for average-case distances introduced in Section 3.4 might seem abstract, it turns out that they share multiple desired properties with common distances used in quantum information [111; 17]. In particular, our distances indeed fulfill metric axioms, they are subadditive with respect to tensor products, and have a joint convexity property. They are also non-increasing under *unital* quantum channels. Finally, the quantum average-case distance between *unital* channels possesses two additional physically well-motivated properties – stability (it does not change when both channels are extended by identity channel) and chaining (distance between compositions of multiple channels is at most the sum of distances between constituting channels) [111]. In this section, we state and prove those properties for states (Section 3.6.1), measurements (Section 3.6.2), and channels (Section 3.6.3). To make navigation easier, each subsection starts with a table of properties, a comparison with relevant worst-case distance, and the text reference in which the properties are proved – Table 3.1 for states, Table 3.2 for measurements, and Table 3.3 for channels.

3.6.1 Quantum states

The following Table 3.1 summarizes properties of the quantum average-case distance between states and compares it to the worst-case trace distance.

Lemma 6 ($\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}$ fulfills axioms of a metric). *Let $\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}$, denote average distances between states defined in Eq.* (3.17) *as*

$$\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho,\sigma) = \frac{1}{2}\sqrt{\mathrm{tr}([\rho-\sigma]^2)} = \frac{1}{2}\|\rho-\sigma\|_{\mathrm{HS}} \,.$$

Then $\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}$ satisfies axioms of a metric in space of quantum states. Specifically, it satisfies the triangle inequality, symmetry, and identity of indiscernibles:

$$\mathrm{d}^{s}_{\mathrm{av}}(\rho,\sigma) \leq \mathrm{d}^{s}_{\mathrm{av}}(\rho,\tau) + \mathrm{d}^{s}_{\mathrm{av}}(\tau,\sigma) \quad \textit{for all} \;\; \rho,\sigma,\tau \in \mathrm{D}(\mathcal{H}_{dim}) \tag{3.56}$$

$$\mathrm{d}^{s}_{\mathrm{av}}(\rho,\sigma) = \mathrm{d}^{s}_{\mathrm{av}}(\sigma,\rho) \quad \textit{for all} \;\; \rho,\sigma \in \mathrm{D}(\mathcal{H}_{dim}) \tag{3.57}$$

$$\mathrm{d}^{s}_{\mathrm{av}}(\rho,\sigma) = 0 \iff \rho = \sigma \quad \textit{for all} \;\; \rho,\sigma \in \mathrm{D}(\mathcal{H}_{dim}) \,. \tag{3.58}$$

Proof. The result follows directly from the fact, that $\mathrm{d}^{s}_{\mathrm{av}}$ is a Hilbert Schmidt distance. □

Property	Comparison
Function	**Worst-case distance:** $d_{tr}(\rho,\sigma) = \frac{1}{2}\|\rho-\sigma\|_1$
	Average-case distance: $d^s_{av}(\rho,\sigma) = \frac{1}{2}\|\rho-\sigma\|_{HS}$ *(See Theorem 1 and Lemma 6)*
Subadditivity	**Both distances satisfy:** $d(\rho_1\otimes\rho_2,\sigma_1\otimes\sigma_2) \leq d(\rho_1,\sigma_1) + d(\rho_2,\sigma_2)$ *(See Lemma 7)*
Joint Convexity	**Both distances satisfy:** $d(\sum_\alpha p_\alpha\rho_\alpha, \sum_\alpha p_\alpha\sigma_\alpha) \leq \sum_\alpha p_\alpha d(\rho_\alpha,\sigma_\alpha)$ *(See Lemma 8)*
Data Processing Inequality	**Worst-case distance:** $d_{tr}(\Lambda(\rho),\Lambda(\sigma)) \leq d_{tr}(\rho,\sigma)$ for CPTP Λ
	Average-case distance: $d^s_{av}(\Phi(\rho),\Phi(\sigma)) \leq d^s_{av}(\rho,\sigma)$ for unital Φ *(See Lemma 9)*

Table 3.1: *Summary of the mathematical properties of worst-case and average-case distances between quantum states.*

Lemma 7 (d^s_{av} is subadditive). *For arbitrary quantum states $\rho_1,\sigma_1 \in D(\mathcal{H}), \rho_2,\sigma_2 \in D(\mathcal{H})$, we have*

$$d^s_{av}(\rho_1\otimes\rho_2,\sigma_1\otimes\sigma_2) \leq d_{av}(\rho_1,\sigma_1) + d_{av}(\rho_2,\sigma_2)\ . \tag{3.59}$$

Proof. The proof follows from triangle inequality and multiplicativity with respect to the tensor product, i.e.

$$\begin{aligned}\|\rho_1\otimes\rho_2-\sigma_1\otimes\sigma_2\|_{HS} &= \|\rho_1\otimes(\rho_2-\sigma_2)-(\sigma_1-\rho_1)\otimes\sigma_2\|_{HS}\\ &\leq \|\rho_1\|_{HS}\|\rho_2-\sigma_2\|_{HS} + \|\sigma_1\|_{HS}\|\sigma_1-\rho_1\|_{HS}\\ &\leq \|\rho_2-\sigma_2\|_{HS} + \|\sigma_1-\rho_1\|_{HS}.\end{aligned} \tag{3.60}$$

□

Lemma 8 (d^s_{av} has joint-convextiy property). *For arbitrary sets of quantum states $\{\rho_\alpha\}_\alpha$, $\{\sigma_\alpha\}_\alpha$ and probability distributions $\{p_\alpha\}$, we have*

$$d^s_{av}\left(\sum_\alpha p_\alpha\rho_\alpha, \sum_\alpha p_\alpha\sigma\alpha\right) \leq \sum_\alpha p_\alpha d^s_{av}(\rho_\alpha,\sigma_\alpha)\ . \tag{3.61}$$

Proof. The proof follows directly from triangle inequality,

$$\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}\left(\sum_{\alpha} p_\alpha \rho_\alpha, \sum_{\alpha} p_\alpha \sigma\alpha\right) = \left\|\sum_{\alpha} p_\alpha(\rho_\alpha - \sigma_\alpha)\right\|_{\mathrm{HS}} \leq \sum_{\alpha} p_\alpha \left\|\rho_\alpha - \sigma_\alpha\right\|_{\mathrm{HS}} = \\ = \sum_{\alpha} p_\alpha \mathrm{d}^{\mathrm{s}}_{\mathrm{av}}\left(\rho_\alpha, \sigma_\alpha\right) \ . \tag{3.62}$$

□

Lemma 9 (Data-processing inequalities for average-case distance between states)**.** *Average-case distance between states is monotonic with respect to unital maps, i.e., for a unital Φ, we have*

$$\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho, \sigma) \geq \mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\Phi(\rho), \Phi(\sigma)) \ . \tag{3.63}$$

Proof. We begin the proof by reminding celebrated Uhlmann theorem [112], that for unital channel Φ and a Hermitian operator H, we have

$$\Phi(H) \prec H \ , \tag{3.64}$$

where the majorization relation above can be seen as a majorization between real vectors of eigenvalues [17; 113]. We also note, that using the fact, that Hilbert-Schmidt norm is a Schur-convex [113] function of eigenvalues, we get

$$\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho, \sigma) \geq \mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\Phi(\rho), \Phi(\sigma)) \ . \tag{3.65}$$

□

Lemma 10 (Separation between $\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}$ and d_{tr})**.** *Let $\rho, \sigma \in \mathcal{H}_d$ be quantum states. Then from standard inequalities between* 1 *and* 2 *norms, it follows that*

$$\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho, \sigma) \ \leq \ \mathrm{d}_{\mathrm{tr}}(\rho, \sigma) \ \leq \ \sqrt{d}\ \mathrm{d}^{\mathrm{s}}_{\mathrm{av}}(\rho, \sigma) \ . \tag{3.66}$$

We now consider an example that attains the bound in Lemma 10.

Example 8 (Two orthogonal maximally mixed states of rank $\frac{\mathrm{d}}{2}$)**.** *Consider two states $\rho, \sigma \in \mathcal{H}_d$, such that $\rho = \frac{\mathbb{I}_{d'}}{d'}, \sigma = \frac{\mathbb{I}_{d'}}{d'}$ on disjoint halves of the space, i.e., $d' = \frac{d}{2}$ and $\mathrm{tr}(\rho\sigma) = 0$. Direct calculation yields*

$$\mathrm{d}^{\mathrm{s}}_{\mathrm{av}}\left(\rho, \sigma\right) = \frac{1}{\sqrt{d}}, \\ \mathrm{d}_{\mathrm{tr}}\left(\rho, \sigma\right) = 1 \ .$$

Clearly, the above shows that in the asymptotic limit, the average-case distance between states goes to 0. From the perspective of statistical distinguishability, it means that the states can be distinguished perfectly with optimal strategy, while randomized strategy fails dramatically.

Example 9 (Counterexample for data-procesiing inequality for quantum states). *Consider two mixed states $\rho, \sigma \in \mathsf{D}(\mathcal{H})$ from previous Example 8. Now consider a* non-unital *quantum channel Λ s.t. $\Lambda(\rho) = |0\rangle\langle 0|$ and $\Lambda(\sigma) = |1\rangle\langle 1|$. Explicit computation combined with results of the previous example yields*

$$\mathsf{d}^{\mathsf{s}}_{\mathsf{av}}\left(\Lambda(\rho), \Lambda(\sigma)\right) = \frac{1}{\sqrt{2}} \quad > \quad \frac{1}{\sqrt{d}} = \mathsf{d}^{\mathsf{s}}_{\mathsf{av}}\left(\rho, \sigma\right) \ ,$$

for $d > 2$.

3.6.2 Quantum measurements

The following Table 3.2 summarizes properties of the quantum average-case distance between measurements and compares it to the worst-case operational distance. For POVMs $\mathbf{M}$ and $\mathbf{N}$, symbol $\mathbf{M} \otimes \mathbf{N}$ denotes a POVM with effects $\{M_i \otimes N_j\}_{i,j}$. Quantum pre-processing channel Γ acts on the state just before measurement $\mathbf{M}$. This is equivalent to performing new POVM with effects transformed via dual channel $M_i \longrightarrow \Gamma^*(M_i)$ on the original state [114]. The fact that channel Γ is trace-preserving implies that dual channel Γ^* is unital, which ensures that $\{\Gamma^*(M_i)\}_i$ a proper POVM. The post-processing stochastic map described by matrix Λ transforms POVM's effects as $M_i \longrightarrow \sum_j \Lambda_{ij} M_j$ (this can be interpreted as classical post-processing of classical outputs of the measurement).

Lemma 11 ($\mathsf{d}^{\mathsf{m}}_{\mathsf{av}}$ fulfills axioms of a metric). *Let $\mathsf{d}^{\mathsf{m}}_{\mathsf{av}}$, denote average distances between quantum measurements defined in Eq.* (3.25) *as*

$$\mathsf{d}^{\mathsf{m}}_{\mathsf{av}}(\mathbf{M}, \mathbf{N}) = \frac{1}{2d} \sum_{i=1}^{n} \sqrt{\|M_i - N_i\|^2_{\mathsf{HS}} + \mathrm{tr}(M_i - N_i)^2} \ .$$

Then $\mathsf{d}^{\mathsf{m}}_{\mathsf{av}}$ satisfies axioms of a metric in space of POVMs. Specifically, it satisfies the triangle inequality, symmetry, and identity of indiscernibles:

$$\mathsf{d}^{\mathsf{m}}_{\mathsf{av}}(\mathbf{M}, \mathbf{N}) \leq \mathsf{d}^{\mathsf{m}}_{\mathsf{av}}(\mathbf{M}, \mathsf{L}) + \mathsf{d}^{\mathsf{m}}_{\mathsf{av}}(\mathsf{L}, \mathbf{N}) \quad \textit{for all} \ \ \mathbf{M}, \mathbf{N}, \mathsf{L} \in \mathsf{P}(\mathcal{H}) \tag{3.67}$$

$$\mathsf{d}^{\mathsf{m}}_{\mathsf{av}}(\mathbf{M}, \mathbf{N}) = \mathsf{d}^{\mathsf{m}}_{\mathsf{av}}(\mathbf{N}, \mathbf{M}) \quad \textit{for all} \ \ \mathbf{M}, \mathbf{N} \in \mathsf{P}(\mathcal{H}) \tag{3.68}$$

$$\mathsf{d}^{\mathsf{m}}_{\mathsf{av}}(\mathbf{M}, \mathbf{N}) = 0 \Longleftrightarrow \mathbf{M} = \mathbf{N} \quad \textit{for all} \ \ \mathbf{M}, \mathbf{N} \in \mathsf{P}(\mathcal{H}) \ . \tag{3.69}$$

Note, that $\mathsf{d}^{\mathsf{m}}_{\mathsf{av}}(\mathbf{M}, \mathbf{N})$ is absolute homogeneous, i.e. if we extend the definition of $\mathsf{d}^{\mathsf{m}}_{\mathsf{av}}$ to arbitrary collections of operators, we see, that $\mathsf{d}^{\mathsf{m}}_{\mathsf{av}}(s\mathbf{M}, s\mathbf{N}) = |s|\mathsf{d}^{\mathsf{m}}_{\mathsf{av}}(\mathbf{M}, \mathbf{N})$.

Proof. We note first that according to Eq. (3.25), $\mathsf{d}^{\mathsf{m}}_{\mathsf{av}}(\mathbf{M}, \mathbf{N})$ is proportional to the sum of non-negative terms of the form

$$\sqrt{\|M_i - N_i\|^2_{\mathsf{HS}} + \mathrm{tr}(M_i - N_i)^2} \ . \tag{3.70}$$

First, we note, that both terms, treated as a functions $(M, N) \mapsto \|M - N\|_{\mathsf{HS}}$ and $(M, N) \mapsto |\mathrm{tr}(M - N)|$ satisfies triangle inequality, moreover the function

Property	Comparison
Function	**Worst-case distance:** $\mathrm{d}_{\mathrm{op}}(\mathbf{M},\mathbf{N}) = \frac{1}{2}\sup_{\rho\in\mathrm{D}(\mathcal{H})}\sum_{i=1}^{n}\lvert\operatorname{tr}(M_i\rho)-\operatorname{tr}(N_i\rho)\rvert$
	Average-case distance: $\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathbf{M},\mathbf{N}) = \frac{1}{2d}\sum_{i=1}^{n}\sqrt{\lVert M_i-N_i\rVert_{\mathrm{HS}}^2+\operatorname{tr}(M_i-N_i)^2}$ *(See Theorem 2 and Lemma 11)*
Subadditivity	**Both distances satisfy:** $\mathrm{d}(\mathbf{M}_1\otimes\mathbf{M}_2,\mathbf{N}_1\otimes\mathbf{N}_2)\leq\mathrm{d}(\mathbf{M}_1,\mathbf{N}_1)+\mathrm{d}(\mathbf{M}_2,\mathbf{N}_2)$ *(See Lemma 12)*
Joint Convexity	**Both distances satisfy:** $\mathrm{d}(\sum_\alpha p_\alpha\mathbf{M}_\alpha,\sum_\alpha p_\alpha\mathbf{N}_\alpha)\leq\sum_\alpha p_\alpha\mathrm{d}(\mathbf{M}_\alpha,\mathbf{N}_\alpha)$ *(See Lemma 13)*
Data Processing Inequality	**Worst-case distance:** $\mathrm{d}_{\mathrm{op}}(\Lambda\circ\mathbf{M}\circ\Gamma,\Lambda\circ\mathbf{N}\circ\Gamma)\leq\mathrm{d}_{\mathrm{op}}(\mathbf{M},\mathbf{N})$ for CPTP Γ, stochastic Λ
	Average-case distance: $\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\Lambda\circ\mathbf{M}\circ\Phi,\Lambda\circ\mathbf{N}\circ\Phi)\leq\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathbf{M},\mathbf{N})$ for unital Φ, stochastic Λ *(See Lemma 14)*

Table 3.2: *Summary of the mathematical properties of worst-case and average-case distances between quantum measurements.*

$(a,b)\mapsto\sqrt{|a|^2+|b|^2}$ is subadditive and increasing in each argument. Therefore $\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathbf{M},\mathbf{N})$ obeys triangle inequality. Symmetry, absolute homogeneity, and identity of indiscernibles follow from direct inspection. □

Lemma 12 ($\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}$ **is subadditive).** *For arbitrary quantum measurements* $\mathbf{M}_1,\mathbf{N}_1\in\mathrm{P}(\mathcal{H}_d,n),\mathbf{M}_2,\mathbf{N}_2\in\mathrm{P}(\mathcal{H}_{d'},n')$*, we have*

$$\mathrm{d}_{\mathrm{av}}^{m}(\mathbf{M}_1\otimes\mathbf{M}_2,\mathbf{N}_1\otimes\mathbf{N}_2)\leq\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathbf{M}_1,\mathbf{N}_1)+\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathbf{M}_2,\mathbf{N}_2)\ . \tag{3.71}$$

Proof. By triangle inequality we have

$$\mathrm{d}_{\mathrm{av}}^{m}(\mathbf{M}_1\otimes\mathbf{M}_2,\mathbf{N}_1\otimes\mathbf{N}_2)\leq\mathrm{d}_{\mathrm{av}}^{m}(\mathbf{M}_1\otimes\mathbf{M}_2,\mathbf{N}_1\otimes\mathbf{M}_2)+\mathrm{d}_{\mathrm{av}}^{m}(\mathbf{N}_1\otimes\mathbf{N}_2,\mathbf{N}_1\otimes\mathbf{M}_2). \tag{3.72}$$

Now we consider one of the terms from the right-hand side of the inequality above and bound it by

$$\mathrm{d}_{\mathrm{av}}^{m}(\mathbf{M}_1\otimes\mathbf{M}_2,\mathbf{N}_1\otimes\mathbf{M}_2)\leq\mathrm{d}_{\mathrm{av}}^{m}(\mathbf{M}_1,\mathbf{N}_1). \tag{3.73}$$

The above inequality follows from direct calculations, since $\mathrm{d}_{\mathrm{av}}^{m}$ is a sum of square

roots of the formulas for a single effect, for which we can write

$$\begin{aligned}
&\|(M_1)_i \otimes (M_2)_j - (N_1)_i \otimes (M_2)_j\|_{\mathrm{HS}}^2 + \mathrm{tr}((M_1)_i \otimes (M_2)_j - (N_1)_i \otimes (M_2)_j)^2 \\
&= \|(M_1)_i - (N_1)_i\|_{\mathrm{HS}}^2 \|(M_2)_j\|_{\mathrm{HS}}^2 + \mathrm{tr}((M_1)_i - (N_1)_i)^2 \,\mathrm{tr}((M_2)_j)^2 \\
&\leq \mathrm{tr}((M_2)_j)^2 \left(\|(M_1)_i - (N_1)_i\|_{\mathrm{HS}}^2 + \mathrm{tr}((M_1)_i - (N_1)_i)^2 \right) .
\end{aligned} \tag{3.74}$$

Combining the terms above together with the fact, that $\sum_j \mathrm{tr}(M_2)_j = d'$, we obtain Eq. (3.73). Similarly, we can bound $\mathrm{d}_{\mathrm{av}}^m(\mathbf{N}_1 \otimes \mathbf{N}_2, \mathbf{N}_1 \otimes \mathbf{M}_2) \leq \mathrm{d}_{\mathrm{av}}^m(\mathbf{N}_2, \mathbf{M}_2)$ which, together with Eq. (3.72) gives us the result. □

Lemma 13 ($\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}$ **has joint-convexity property).** *For arbitrary sets of quantum measurements* $\{\mathbf{M}_\alpha\}_\alpha$, $\{\mathbf{N}_\alpha\}_\alpha$ *and probability distributions* $\{p_\alpha\}$*, we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}} \left(\sum_\alpha p_\alpha \mathbf{M}_\alpha, \sum_\alpha p_\alpha \mathbf{N}_\alpha \right) \leq \sum_\alpha p_\alpha \mathrm{d}_{\mathrm{av}}^{\mathrm{m}} \left(\mathbf{M}_\alpha, \mathbf{N}_\alpha \right) . \tag{3.75}$$

Proof. The proof is analogous to the one for states and follows from triangle inequality, and absolute homogeneity:

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}} \left(\sum_\alpha p_\alpha \mathbf{M}_\alpha, \sum_\alpha p_\alpha \mathbf{N}_\alpha \right) \leq \sum_\alpha \mathrm{d}_{\mathrm{av}}^{\mathrm{m}} \left(p_\alpha \mathbf{M}_\alpha, p_\alpha \mathbf{N}_\alpha \right) = \sum_\alpha p_\alpha \mathrm{d}_{\mathrm{av}}^{\mathrm{m}} \left(\mathbf{M}_\alpha, \mathbf{N}_\alpha \right) . \tag{3.76}$$

□

Lemma 14 (Data-processing inequalities for average-case distance between measurements). *Average-case distance between quantum measurements is monotonic with respect to a unital pre- and general post-processing, i.e. for a stochastic matrix* Λ *and a general unital CPTP map* Φ*, we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\Lambda \circ \mathbf{M} \circ \Phi, \Lambda \circ \mathbf{N} \circ \Phi) \leq \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathbf{M}, \mathbf{N}) . \tag{3.77}$$

Proof. We will show, that the average-case distance between measurements is monotonic with respect to post-processing. Since the outcome of a measurement is classical, we will consider only classical post-processing, given by a stochastic matrix Λ. We denote by $\Delta_j = M_j - N_j$. We will use the fact, that each term in the sum which defines $\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathbf{M}, \mathbf{N})$, is absolutely homogeneous and obeys

the triangle inequality (see Eq. (3.25) and discussion under Eq. (3.70))

$$\begin{aligned}
&\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\Lambda \circ \mathbf{M}, \Lambda \circ \mathbf{N}) = \\
&= \frac{1}{2d} \sum_i^n \sqrt{|| \sum_j \Lambda_{ij} \Delta_j ||^2_{\mathrm{HS}} + |\operatorname{tr} \sum_j \Lambda_{ij} \Delta_j|^2} \leq \\
&\leq \frac{1}{2d} \sum_i^n \sqrt{\sum_j ||\Lambda_{ij} \Delta_j||^2_{\mathrm{HS}} + \sum_j |\operatorname{tr} \Lambda_{ij} \Delta_j|^2} \\
&= \frac{1}{2d} \sum_i^n \sqrt{\sum_j \Lambda^2_{ij} ||\Delta_j||^2_{\mathrm{HS}} + \sum_j \Lambda^2_{ij} |\operatorname{tr} \Delta_j|^2} \\
&\leq \frac{1}{2d} \sum_{i,j}^n \Lambda_{ij} \sqrt{\operatorname{tr} (\Delta_j)^2 + (\operatorname{tr} \Delta_j)^2} = \\
&= \frac{1}{2d} \sum_j^n \sqrt{\operatorname{tr} (\Delta_j)^2 + (\operatorname{tr} \Delta_j)^2} = \mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathbf{M}, \mathbf{N}) \ .
\end{aligned} \tag{3.78}$$

In the first inequality we used the triangle inequality for both terms inside the square root, the second equality follows from absolute homogeneity of both terms (note that $\Lambda_{i,j}$ are nonegative), while the second inequality is a consequence of subadditivity of the square root function.

In order to show that the average-case distance between quantum measurements is monotonic with respect to unital pre-processing, we consider a general unital CPTP map Φ. Note, that the adjoint map Φ^* is also unital and CPTP. Recall that we can look at the adjoint action of the channel on the effects of $\mathbf{M}$ as

$$\operatorname{tr} M_i \Phi(\rho) = \operatorname{tr} \Phi^*(M_i) \rho \ . \tag{3.79}$$

The fact, that Φ^* is unital assures us that $\mathbf{M'}$ with effects $\{\Phi^*(M_i)\}_i$ forms a POVM . Now we consider the basic terms, which define $\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}$, first we see (again using Uhlmann's theorem [112] and Schur convexity of HS-norm)

$$\|\Phi^*(\Delta_i)\|^2_{\mathrm{HS}} \leq \|\Delta_i\|^2_{\mathrm{HS}} \ . \tag{3.80}$$

Next, since Φ^* is trace-preserving we have

$$\operatorname{tr}(\Phi^*(\Delta_i))^2 = \operatorname{tr}(\Delta_i)^2 \ , \tag{3.81}$$

which finishes the proof of monotonicity with respect to unital pre-processing. □

Lemma 15 (Separation between $\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}$ and d_{op})**.** *For any quantum measurements* $\mathbf{M}, \mathbf{N} \in \mathrm{P}(\mathcal{H}_d)$*, we have*

$$a \ \mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathbf{M}, \mathbf{N}) \ \leq \ \mathrm{d}_{\mathrm{op}}(\mathbf{M}, \mathbf{N}) \leq d \ \mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathbf{M}, \mathbf{N}) \ , \tag{3.82}$$

where $a = 0.31$.

Proof. The lower bound follows from Theorem 2. For the upper bound, we directly calculate

$$\mathrm{d}_{\mathrm{op}}(\mathbf{M},\mathbf{N}) = \frac{1}{2}\max_{\rho}\sum_i |\operatorname{tr}(M_i - N_i)\rho| \leq \frac{1}{2}\sum_i \sqrt{\|(M_i - N_i)\|_{\mathrm{HS}}^2} \leq$$
$$\leq \frac{1}{2}\sum_i \sqrt{\|(M_i - N_i)\|_{\mathrm{HS}}^2 + \operatorname{tr}(M_i - N_i)^2} = d\ \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(M,N)\ .$$

□

The following example attains bound in Lemma 15 up to a constant.

Example 10 (Swapping two outcomes of standard measurement). *Consider computational basis measurement* **P** *in* $\mathcal{H}_d$ *with effects* $P_i = |i\rangle\langle i|$*), and second measurement* **M** *that is obtained from* **P** *by exchanging the first two effects, leaving others intact, i.e.,* $M_1 = |2\rangle\langle 2|$, $M_2 = |1\rangle\langle 1|$, *and* $M_i = |i\rangle\langle i|$ *for* $i = 3,\ldots,d$*. In this scenario, direct calculation yields*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathbf{P},\mathbf{M}) = \sqrt{2}\,\frac{1}{d}\ ,$$
$$\mathrm{d}_{\mathrm{op}}(\mathbf{P},\mathbf{M}) = 1\ .$$

The above implies that in the asymptotic limit, similarly to Example 8 for states, considered measurements can be distinguished perfectly with optimal strategy, while randomized one will not work. On the other hand, if we interpret the second measurement **M** *as a noisy version of target* **P***, then this particular type of noise (that swaps two measurement outcomes) will not highly affect the results of generic experiments.*

We note that the above example, together with asymptotic separation, can be easily generalized to a scenario where the second measurement, instead of swapping only 2 outcomes of **P***, swaps some* constant *number of them.*

Example 11 (Counterexample for data-procesing inequality for quantum measurements). *Consider POVMs* **P** *and* **M** *from previous Example 10. Consider now a* non-unital *channel* Λ *that regardless of the input state prepares a state* $|1\rangle\langle 1|$ *(which is a possible choice for optimal discriminator of POVMs* **M** *and* **P***). Dual action of this channel on POVM's effects is* $\Lambda^\dagger(M_i) = \operatorname{tr}(M_i\,|1\rangle\langle 1|)\,\mathbb{1}$*. The direct calculation, together with results from the previous example, yields*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathbf{P}\circ\Lambda,\mathbf{M}\circ\Lambda) = \sqrt{1+\frac{1}{d}} \quad > \quad \sqrt{2}\,\frac{1}{d} = \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathbf{P},\mathbf{M})\ .$$

3.6.3 Quantum channels

The following Table 3.3 summarizes properties of the quantum average-case distance between channels and compares it to the worst-case diamond-norm distance. Compared to previous Tables, here we also consider two additional

properties relevant for quantum channels, namely stability and chaining [111] which for average-case distance hold for *unital* quantum channels. Stability means that a given distance measure does not change if a channel is extended by an identity channel. In other words, trivial extensions of maps by an ancillary system do not affect their distance measure. Chaining means that distance between multiple compositions of the channel is at most a sum of the distances between constituting channels. If one sequence is a composition of target gates, and the other is their noisy version, this property implies that the total error is at most additive in a given distance measure.

Property	Comparison
Function	**Worst-case distance:** $\mathrm{d}_{\diamond}(\Lambda,\Gamma) = \|\|\Lambda-\Gamma\|\|_{\diamond}$
	Average-case distance: $\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\mathbf{M},\mathbf{N}) = \frac{1}{2}\sqrt{\|\mathcal{J}_{\Lambda}-\mathcal{J}_{\Gamma}\|_{\mathrm{HS}}^2 + \mathrm{tr}\left((\Lambda-\Gamma)[\tau_{\mathrm{d}}]^2\right)}$ *(See Theorem 3 and Lemma 16*
Subadditivity	**Both distances satisfy:** $\mathrm{d}(\Lambda_1\otimes\Lambda_2,\Gamma_1\otimes\Gamma_2) \leq \mathrm{d}(\Lambda_1,\Gamma_1)+\mathrm{d}(\Lambda_2,\Gamma_2)$ *(See Lemma 17)*
Joint Convexity	**Both distances satisfy:** $\mathrm{d}\left(\sum_\alpha p_\alpha\Lambda_\alpha,\sum_\alpha p_\alpha\Gamma_\alpha\right) \leq \sum_\alpha p_\alpha \mathrm{d}\left(\Lambda_\alpha,\Gamma_\alpha\right)$ *(See Lemma 18)*
Data Processing Inequality	**Worst-case distance:** $\mathrm{d}_{\diamond}(\Phi_o\circ\Lambda\circ\Phi_i,\Phi_o\circ\Gamma\circ\Phi_i) \leq \mathrm{d}_{\diamond}(\Lambda,\Gamma)$ for CPTP $\Phi_i,\ \Phi_o$
	Average-case distance: $\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Phi_o\circ\Lambda\circ\Phi_i,\Phi_o\circ\Gamma\circ\Phi_i) \leq \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Gamma)$ for unital $\Phi_i,\ \Phi_o$ *(See Lemma 19)*

Table 3.3: *Summary of the mathematical properties of worst-case and average-case distances between quantum channels.*

Lemma 16 ($\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}$ **fulfills axioms of a metric).** *Let* $\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}$*, denote average distances between channels defined in Eq.* (3.28) *as*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Gamma) \coloneqq \frac{1}{2}\sqrt{\|\mathcal{J}_{\Lambda}-\mathcal{J}_{\Gamma}\|_{\mathrm{HS}}^2 + \mathrm{tr}\left((\Lambda-\Gamma)[\tau_d]^2\right)}.$$

Then $\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}$ *satisfies axioms of a metric in space of quantum channels. Specifically, it satisfies the triangle inequality, symmetry, and identity of indiscernibles:*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Gamma) \leq \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Phi) + \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Phi,\Gamma) \quad \textit{for all} \quad \Lambda,\Gamma,\Phi \in \mathrm{CPTP}(\mathcal{H}_d) \tag{3.83}$$

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Gamma) = \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Gamma,\Lambda) \quad \textit{for all } \Lambda,\Gamma \in \mathrm{CPTP}(\mathcal{H}_d) \tag{3.84}$$

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Gamma) = 0 \iff \Lambda = \Gamma \quad \textit{for all } \Lambda,\Gamma \in \mathrm{CPTP}(\mathcal{H}_d)\ . \tag{3.85}$$

Note, that $\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Gamma)$ *is absolute homogeneous, i.e.* $\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(s\Lambda, s\Gamma) = |s|\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Gamma)$.

Proof. Note that $\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}$ is a function of a distance measure ($\|\mathcal{J}_\Lambda - \mathcal{J}_\Gamma\|_{HS}$) and a term ($\sqrt{\mathrm{tr}\left((\Lambda-\Gamma)[\tau_d]^2\right)}$), which treated as a function, obeys the triangle inequality. Since the function $(a,b) \mapsto \sqrt{|a|^2+|b|^2}$ is subadditive and increasing in each argument, thus $\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}$ obeys triangle inequality. Symmetry and identity of indiscernibles follows from direct inspection. □

Lemma 17 ($\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}$ is subadditive). *For arbitrary quantum channels* $\Lambda_1,\Gamma_1 \in \mathrm{CPTP}(\mathcal{H})$, $\Lambda_2,\Gamma_2 \in \mathrm{CPTP}(\mathbb{C}^{d'})$, *we have*

$$\mathrm{d}_{\mathrm{av}}^{ch}(\Lambda_1\otimes\Lambda_2,\Gamma_1\otimes\Gamma_2) \leq \mathrm{d}_{\mathrm{av}}(\Lambda_1,\Gamma_1) + \mathrm{d}_{\mathrm{av}}(\Lambda_2,\Gamma_2)\ . \tag{3.86}$$

Proof. We begin with triangle inequality

$$\mathrm{d}_{\mathrm{av}}^{ch}(\Lambda_1\otimes\Lambda_2,\Gamma_1\otimes\Gamma_2) \leq \mathrm{d}_{\mathrm{av}}^{ch}(\Lambda_1\otimes\Lambda_2,\Gamma_1\otimes\Lambda_2) + \mathrm{d}_{\mathrm{av}}^{ch}(\Gamma_1\otimes\Gamma_2,\Gamma_1\otimes\Lambda_2)\ . \tag{3.87}$$

Now we consider

$$\mathrm{d}_{\mathrm{av}}^{ch}(\Lambda_1\otimes\Lambda_2,\Gamma_1\otimes\Lambda_2) = \sqrt{\|\mathcal{J}_{\Lambda_1\otimes\Lambda_2} - \mathcal{J}_{\Gamma_1\otimes\Lambda_2}\|_{\mathrm{HS}}^2 + \mathrm{tr}\left(((\Lambda_1-\Gamma_1)\otimes\Lambda_2)\left(\frac{\mathbb{I}_{dd'}}{dd'}\right)^2\right)}\ . \tag{3.88}$$

First we note, that $\mathcal{J}_{\Lambda_1\otimes\Lambda_2}$ is permutationally similar, to $\mathcal{J}_{\Lambda_1}\otimes\mathcal{J}_{\Lambda_2}$. To explain that, let us note that in the definition of $\mathcal{J}_{\Lambda_1}$, the channel $\Lambda_1\otimes\mathcal{I}$ naturally divides the space into subsystems that we can label as 1 and 2 (two halves of the maximally entangled state). In the definition of $\mathcal{J}_{\Lambda_2}$, the analogous division leads to subsystems with labels 3 and 4. Hence, if we take the tensor product of the Choi states, we need to restore the ordering of the subsystems in order to recover the Choi matrix of the tensor product of the channels. This is done by application of the permutation $\mathbb{S}_{23}$ between subsystems 2 and 3, i.e., we have $\mathcal{J}_{\Lambda_1\otimes\Lambda_2} = \mathbb{S}_{23}\left(\mathcal{J}_{\Lambda_1}\otimes\mathcal{J}_{\Lambda_2}\right)\mathbb{S}_{23}$. We can now use this fact, together with the standard properties of the HS norm, to obtain

$$\begin{aligned}\|\mathcal{J}_{\Lambda_1\otimes\Lambda_2} - \mathcal{J}_{\Gamma_1\otimes\Lambda_2}\|_{\mathrm{HS}} &= \|\mathcal{J}_{\Lambda_1}\otimes\mathcal{J}_{\Lambda_2} - \mathcal{J}_{\Gamma_1}\otimes\mathcal{J}_{\Lambda_2}\|_{\mathrm{HS}}\\ &= \|\mathcal{J}_{\Lambda_1} - \mathcal{J}_{\Gamma_1}\|_{\mathrm{HS}}\|\mathcal{J}_{\Lambda_2}\|_{\mathrm{HS}}\\ &\leq \|\mathcal{J}_{\Lambda_1} - \mathcal{J}_{\Gamma_1}\|_{\mathrm{HS}}\ .\end{aligned} \tag{3.89}$$

Next, we note

$$\begin{aligned}\mathrm{tr}\left(((\Lambda_1-\Gamma_1)\otimes\Lambda_2)\left(\frac{\mathbb{I}_{dd'}}{dd'}\right)^2\right) &= \mathrm{tr}\left((\Lambda_1-\Gamma_1)\left(\frac{\mathbb{I}_d}{d}\right)^2\right)\mathrm{tr}\left(\Lambda_2\left(\frac{\mathbb{I}_{d'}}{d'}\right)^2\right)\\ &\leq \mathrm{tr}\left((\Lambda_1-\Gamma_1)\left(\frac{\mathbb{I}_d}{d}\right)^2\right).\end{aligned} \tag{3.90}$$

Combining the above we get

$$\mathrm{d}_{\mathrm{av}}^{ch}(\Lambda_1 \otimes \Lambda_2, \Gamma_1 \otimes \Lambda_2) \leq \sqrt{\|\mathcal{J}_{\Lambda_1} - \mathcal{J}_{\Gamma_1}\|_{\mathrm{HS}}^2 + \mathrm{tr}\left((\Lambda_1 - \Gamma_1)\left(\frac{\mathbb{I}}{d}\right)^2\right)} = \mathrm{d}_{\mathrm{av}}^{ch}(\Lambda_1, \Gamma_1)\ . \tag{3.91}$$

We can analogously bound $\mathrm{d}_{\mathrm{av}}^{ch}(\Gamma_1 \otimes \Gamma_2, \Gamma_1 \otimes \Lambda_2) \leq \mathrm{d}_{\mathrm{av}}^{ch}(\Gamma_2, \Lambda_2)$ and using Eq. (3.87) we obtain the result. □

Lemma 18 ($\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}$ **has joint-convexity property).** *For arbitrary sets of quantum channels* $\{\Lambda_\alpha\}_\alpha$, $\{\Gamma_\alpha\}_\alpha$ *and probability distributions* $\{p_\alpha\}$*, we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}\left(\sum_\alpha p_\alpha \Lambda_\alpha, \sum_\alpha p_\alpha \Gamma_\alpha\right) \leq \sum_\alpha p_\alpha \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}\left(\Lambda_\alpha, \Gamma_\alpha\right)\ . \tag{3.92}$$

Proof. The proof is analogous to the one for states and measurements and follows from triangle inequality, and absolute homogeneity of $\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}$. □

Lemma 19 (Data-processing inequalities for average-case distance between channels). *Average-case distance between quantum channels is monotonic with respect to unital pre- and postprocessing, i.e. for a unital maps* Φ_o, Φ_i*, we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Phi_o \circ \Lambda \circ \Phi_i, \Phi_o \circ \Gamma \circ \Phi_i) \leq \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda, \Gamma). \tag{3.93}$$

Proof. The inequality related to the postprocessing follows directly analogous results for states, in order to show the monotonicity with respect to the preprocessing inequality we write, for unital Φ

$$\begin{aligned} &\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda \circ \Phi, \Gamma \circ \Phi) \\ &= \frac{1}{2}\sqrt{\|\mathcal{J}_{\Lambda\circ\Phi} - \mathcal{J}_{\Gamma\circ\Phi}\|_{\mathrm{HS}}^2 + \mathrm{tr}\left((\Lambda - \Gamma)[\frac{\mathbb{I}}{d}]\right)^2}. \end{aligned} \tag{3.94}$$

We can consider only the term $\|\mathcal{J}_{\Lambda\circ\Phi} - \mathcal{J}_{\Gamma\circ\Phi}\|_{\mathrm{HS}}$, since the second one does not change under preprocessing by a unital map. First, we write a norm in terms of superoperators, i.e.

$$\|\mathcal{J}_{\Lambda\circ\Phi} - \mathcal{J}_{\Gamma\circ\Phi}\|_{\mathrm{HS}} = \left\|(\hat{\Lambda} - \hat{\Gamma})\hat{\Phi}\right\|_{\mathrm{HS}}, \tag{3.95}$$

where $\hat{\Lambda}$ denotes the superoperator matrix ([69]) of channel Λ. Now we use inequality

$$\|AB\|_{\mathrm{HS}} \leq \|A\|_{\mathrm{HS}}\,\|B\|_\infty \tag{3.96}$$

and write

$$\|\mathcal{J}_{\Lambda\circ\Phi} - \mathcal{J}_{\Gamma\circ\Phi}\|_{\mathrm{HS}} \leq \left\|\hat{\Lambda} - \hat{\Gamma}\right\|_{\mathrm{HS}} \left\|\hat{\Phi}\right\|_\infty. \tag{3.97}$$

Now since for any unital map we have $\left\|\hat{\Gamma}\right\|_\infty = 1$ (see [115, Theorem 1]), we obtain the result. □

Lemma 20 (Stability property of average-case distance between unital channels). *Average-case distance between unital quantum channels fulfills stability property [111], i.e. for unital maps Λ, Γ, and identity channel $\mathcal{I}$ acting on arbitrary dimension, we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda \otimes \mathcal{I}, \Gamma \otimes \mathcal{I}) = \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda, \Gamma) \ . \tag{3.98}$$

Proof. For unital channels we have $\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda, \Gamma) = \frac{1}{2}||\mathcal{J}_\Lambda - \mathcal{J}_\Gamma||_{HS}$. We now recall that for any channels Λ, Γ, Choi matrix $\mathcal{J}_{\Lambda\otimes\Gamma}$ is permutationally similar to $\mathcal{J}_\Lambda \otimes \mathcal{J}_\Gamma$. This allows to rewrite the HS norm as

$$||\mathcal{J}_{\Lambda\otimes\mathcal{I}} - \mathcal{J}_{\Gamma\otimes\mathcal{I}}||_{\mathrm{HS}} = ||(\mathcal{J}_\Lambda - \mathcal{J}_\Gamma) \otimes \mathcal{J}_\mathcal{I}||_{\mathrm{HS}} = ||\mathcal{J}_\Lambda - \mathcal{J}_\Gamma||_{\mathrm{HS}} \underbrace{||\mathcal{J}_\mathcal{I}||_{\mathrm{HS}}}_{=1} = ||\mathcal{J}_\Lambda - \mathcal{J}_\Gamma||_{\mathrm{HS}} \ ,$$

which concludes the proof. □

Remark 10. *For generic, non-unital channels, the expression for average-case distance has additional term* $\mathrm{tr}\left((\Gamma - \Lambda)(\frac{\mathbb{I}}{d})\right)^2$. *If we extend our channels by identity $\mathcal{I}_{d'}$ on dimension d', this 'non-unitality' term changes to* $\mathrm{tr}\left(((\Gamma - \Lambda) \otimes \mathcal{I}_{d'})(\frac{\mathbb{I}}{dd'})\right)^2 = \frac{1}{d'}\,\mathrm{tr}\left((\Gamma - \Lambda)(\frac{\mathbb{I}}{d})\right)^2$. *Therefore, the contribution to the average-case distance of the 'non-unitality' decreases as d' increases. Note that this scenario corresponds to channel discrimination (via random circuits) with the use of an ancillary system.*

Lemma 21 (Chaining property of average-case distance between unital channels). *Average-case distance between unital quantum channels fulfills chaining property [111], i.e. for unital maps $\Lambda_1, \Lambda_2, \Gamma_1, \Gamma_2$, we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_1 \circ \Lambda_2, \Gamma_1 \circ \Gamma_2) \leq \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_1, \Gamma_1) + \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_2, \Gamma_2). \tag{3.99}$$

Proof. To prove the theorem, we apply triangle inequality followed by the data-processing inequality for unital channels (Lemma 19)

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_1 \circ \Lambda_2, \Gamma_1 \circ \Gamma_2) \leq \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_1 \circ \Gamma_2, \Gamma_1 \circ \Gamma_2) + \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_1 \circ \Gamma_2, \Lambda_1 \circ \Lambda_2) \leq \tag{3.100}$$

$$\leq \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_1, \Gamma_1) + \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_2, \Gamma_2) \ . \tag{3.101}$$

□

Remark 11. *We note that for generic, non-unital channels, the chaining property of average-case distance does not hold. To see that, we note that if we choose channels $\Lambda_1 = \Gamma_1$ to be the same, the chaining property effectively reduces to data-processing inequality, which we know does not hold for generic channels (see below for a counterexample).*

Lemma 22 (Separation between $\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}$ and $\mathrm{d}_\diamond$). *For any quantum channels $\Lambda, \Gamma \in \mathrm{CPTP}(\mathcal{H}_d)$, we have*

$$a^{\mathrm{ch}}\ \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda, \Gamma) \ \leq \ \mathrm{d}_\diamond(\Lambda, \Gamma) \leq d^{\frac{3}{2}}\ \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda, \Gamma) \ , \tag{3.102}$$

where $a^{\mathrm{ch}} = 0.087$.

Proof. The lower bound is a consequence of Theorem 3. To show the other inequality we begin with the upper bound for diamond norm (see [116, Thm. 2] and [117, Prop. 1]), which for Hermiticity preserving operation can be written in our notation as

$$\mathrm{d}_\diamond(\Lambda,\Gamma) \leq \frac{d}{2}\,\|\mathrm{tr}_2(|\mathcal{J}_\Lambda - \mathcal{J}_\Gamma|)\|_\infty\ , \tag{3.103}$$

Next express the operator norm via maximization over pure states on the first subsystem

$$\|\mathrm{tr}_2(|\mathcal{J}_\Lambda - \mathcal{J}_\Gamma|)\|_\infty = \max_{\psi\in\mathcal{S}(\mathcal{H}_\mathrm{d})} |\,\mathrm{tr}\,(\psi\otimes\mathbb{I}_\mathrm{d}|\mathcal{J}_\Lambda - \mathcal{J}_\Gamma|)\,|\ . \tag{3.104}$$

Applying to the above Cauchy - Schwarz inequality we obtain

$$\|\mathrm{tr}_2(|\mathcal{J}_\Lambda - \mathcal{J}_\Gamma|)\|_\infty \leq \max_{\psi\in\mathcal{S}(\mathcal{H}_\mathrm{d})} \|\psi\otimes\mathbb{I}_\mathrm{d}\|_\mathrm{HS}\,\|\mathcal{J}_\Lambda - \mathcal{J}_\Gamma\|_\mathrm{HS} = \sqrt{d}\,\|\mathcal{J}_\Lambda - \mathcal{J}_\Gamma\|_\mathrm{HS}\ . \tag{3.105}$$

Combining the above we obtain the desired result

$$\mathrm{d}_\diamond(\Lambda,\Gamma) \leq \frac{d}{2}\sqrt{d}\,\|\mathcal{J}_\Lambda - \mathcal{J}_\Gamma\|_\mathrm{HS} \leq d^{\frac{3}{2}}\,\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma)\ . \tag{3.106}$$

□

Example 12 (Separation example). *Let us consider even dimensional Hilbert space $\mathcal{H}_d$ and a Hermitian matrix A, such that* $\mathrm{tr}\,A = 0$ *and* $A^2 = \mathbb{I}_d$*. Next, we define a pair of channels Λ and Γ by their Jamiołkowski states as*

$$\begin{aligned}\mathcal{J}_\Lambda &= \frac{1}{d^2}\,\mathbb{I}_{d^2}\,,\\ \mathcal{J}_\Gamma &= \frac{1}{d^2}\,\mathbb{I}_{d^2} - \frac{1}{d^2}\,|\psi\rangle\langle\psi|\otimes A\,,\end{aligned} \tag{3.107}$$

where $\psi\in\mathcal{S}(\mathcal{H}_d)$ is an arbitrary pure state. The diamond norm between Λ and Γ can be calculated easily, using an alternative formula for the diamond norm for Hermiticity preserving operations (see e.g. [118, Eqn. (11)]), i.e.,

$$\begin{aligned}\|\Lambda-\Gamma\|_\diamond &= d\max_{\rho\in\mathrm{D}(\mathcal{H}_d)}\|(\sqrt{\rho}\otimes\mathbb{I})\mathcal{J}_{\Lambda-\Gamma}(\sqrt{\rho}\otimes\mathbb{I})\|_1\\ &= d\max_{\rho\in\mathrm{D}(\mathcal{H}_d)}\|(\sqrt{\rho}\otimes\mathbb{I})(\frac{1}{d^2}\,|\psi\rangle\langle\psi|\otimes A)(\sqrt{\rho}\otimes\mathbb{I})\|_1 = \frac{1}{d}\|A\|_1 = 1.\end{aligned} \tag{3.108}$$

The average distance can be evaluated as

$$\begin{aligned}\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda,\Gamma) &= \frac{1}{2}\sqrt{\|\mathcal{J}_{\Lambda-\Gamma}\|^2_\mathrm{HS} + \|\Lambda(\mathbb{I}/d)-\Gamma(\mathbb{I}/d)\|^2_\mathrm{HS}}\\ &= \frac{1}{2}\sqrt{\|\frac{1}{d^2}\,|\psi\rangle\langle\psi|\otimes A\|^2_\mathrm{HS} + \|\frac{1}{d^2}\,\mathrm{tr}_1(|\psi\rangle\langle\psi|\otimes A)\|^2_\mathrm{HS}}\\ &= \frac{1}{2}\sqrt{\frac{1}{d^4}\|A\|^2_\mathrm{HS} + \frac{1}{d^4}\|A\|^2_\mathrm{HS}} = \frac{\sqrt{2}}{2d^{\frac{3}{2}}}.\end{aligned} \tag{3.109}$$

Which gives us finally the separation of order $d^{\frac{3}{2}}$,

$$\frac{1}{2} = \mathrm{d}_\diamond(\Lambda, \Gamma) = d^{\frac{3}{2}} \frac{1}{\sqrt{2}} \mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda, \Gamma). \tag{3.110}$$

Example 13 (Counterexample for general post-processing monotonicity for quantum channels). *Consider two state-preparation channels acting on N-qubit system as $\Lambda(\rho) = \mathrm{tr}(\rho)\,|0\rangle\langle 0| \otimes \frac{\mathbb{I}}{2^{N-1}}$ and $\Gamma(\rho) = \mathrm{tr}(\rho)\,|1\rangle\langle 1| \otimes \frac{\mathbb{I}}{2^{N-1}}$, for any input state $\rho \in \mathrm{D}(\mathcal{H})$. Then we have*

$$\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda, \Gamma) = \frac{1}{2}\sqrt{\frac{1}{d}(1 + \frac{1}{d})} \tag{3.111}$$

$$\mathrm{d}_\diamond(\Lambda, \Gamma) = 1\,, \tag{3.112}$$

where expression for average-case distance follows from direct calculation, and the value of diamond norm follows from the fact that channels always prepare states that are orthogonal on first qubits, and thus can be perfectly distinguished.

Now consider additional non-unital *conditional state-preparation channel $\tilde{\Lambda}$ that acts as $\tilde{\Lambda}(|0\rangle\langle 0| \otimes \sigma) = \psi$ and $\tilde{\Lambda}(|1\rangle\langle 1| \otimes \sigma) = \psi^\perp$ for any σ, where $\psi, \psi^\perp$ are two orthogonal pure states. Note that the composed action of the channels reduces to state-preparation channels $\tilde{\Lambda} \circ \Lambda(\rho) = \psi$ and $\tilde{\Lambda} \circ \Gamma(\rho) = \psi^\perp$ for any ρ. Direct computation together with Eq.* (3.111) *yields*

$$\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\tilde{\Lambda} \circ \Lambda, \tilde{\Lambda} \circ \Gamma) = \sqrt{\frac{1}{2}(1 + \frac{1}{d})} \;>\; \frac{1}{2}\sqrt{\frac{1}{d}(1 + \frac{1}{d})} = \mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda, \Gamma)\,.$$

Example 14 (Counterexample for general pre-processing monotonicity for quantum channels). *Consider two perfectly distinguishable unitary channels of size $d > 2$, $\Lambda_U : \rho \mapsto U\rho U^\dagger$ and $\Lambda_V : \rho \mapsto V\rho V^\dagger$.*

The average distance can be calculated directly and is equal to (see also Example 19)

$$\mathrm{d}^{\mathrm{ch}}_{\mathrm{av}}(\Lambda_U, \Lambda_V) = \frac{1}{2}\sqrt{2 - \frac{2}{d^2}|\,\mathrm{tr}\, U^\dagger V|^2}\,. \tag{3.113}$$

Since channels Λ_U and Λ_V are perfectly distinguishable, let $|\psi\rangle$ be the optimal discriminator, i.e. the state for which $\langle\psi|\, U^\dagger V\, |\psi\rangle = 0$. Note, that in the case of unitary channels, one does not need to attach an additional system in order to perform optimal discrimination. Now we consider a channel $\Gamma : \rho \mapsto \mathrm{tr}(\rho)\,|\psi\rangle\langle\psi|$, which prepares the optimal discriminator. We then have

$$\begin{aligned} \mathcal{J}_{\Lambda_U \circ \Gamma} &= \mathbb{I}/d \otimes U\,|\psi\rangle\langle\psi|\,U^\dagger\,, \\ \mathcal{J}_{\Lambda_V \circ \Gamma} &= \mathbb{I}/d \otimes V\,|\psi\rangle\langle\psi|\,V^\dagger\,. \end{aligned} \tag{3.114}$$

Direct computations yield the following result

$$\begin{aligned}&\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_U \circ \Gamma, \Lambda_V \circ \Gamma) = \\ &= \frac{1}{2}\sqrt{\|\mathbb{I}/d \otimes (U|\psi\rangle\langle\psi|U^\dagger - V|\psi\rangle\langle\psi|V^\dagger)\|_{\mathrm{HS}}^2 + \mathrm{tr}(U|\psi\rangle\langle\psi|U^\dagger - V|\psi\rangle\langle\psi|V^\dagger)^2} = \\ &= \frac{1}{2}\sqrt{\frac{2}{d}+2}.\end{aligned} \tag{3.115}$$

Finally, we obtain that the data processing inequality for general pre-processing does not hold.

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_U \circ \Gamma, \Lambda_V \circ \Gamma) \;>\; \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_U, \Lambda_V). \tag{3.116}$$

If we choose $U = \mathbb{I}, V = diag(1, -1, 1, \ldots, 1)$ *we get* $\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_U, \Lambda_V) = \frac{1}{2}\sqrt{2 - \frac{2}{d^2}(d-2)^2} = \frac{1}{d}\sqrt{2(d-1)}$.

In fact, similar calculations can be performed on any distinguishable channels, with the pre-processing channel chosen to be the preparation of the optimal discriminator.

3.7 More use cases

In previous parts of the Chapter, we discussed some specific scenarios in which scaling of quantum average-case distances with system size provided some insight into various areas of quantum information. In this part, we investigate some further exemplary scenarios, and we provide a discussion of the consequences of our findings.

3.7.1 Convergence to uniform distribution

One particularly interesting consequence of our main theorems is that average-case distances allow us to easily study a convergence of the average Total-Variation distance between the noisy distribution (generated by a non-ideal quantum device) and the uniform distribution (the fact which was used multiple times in Section 3.5). To this aim, one needs to calculate an average-case distance between a noisy state, measurement, or channel, and the maximally mixed state, trivial POVM, or maximally depolarizing channel, respectively. We summarize those observations in the following Lemmas 23, 24, 25 – the proofs follow directly from Theorems 1, 2, and 3, respectively. In what follows we denote uniform distribution as $\mathbf{p}^{\mathrm{uniform}}$, meaning $p_i^{\mathrm{uniform}} = \frac{1}{d}$ for all $i = 1, \ldots, d$.

Lemma 23. *[Noisy states – convergence to uniform distribution] Let* ψ *be a pure state and* Λ *a quantum channel. Then we have*

$$\mathop{\mathbb{E}}_{U\sim\nu} \mathrm{TV}(\mathbf{p}^{\Lambda(\psi),U}, \mathbf{p}^{\mathrm{uniform}}) \approx \mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\Lambda(\psi), \frac{\mathbb{I}}{d}) = \frac{1}{2}\sqrt{\mathrm{tr}\left((\Lambda(\psi))^2\right) - \frac{1}{d}}\,. \tag{3.117}$$

In the above, the ≈ sign means the approximation in a sense of Eq. (3.16)*, i.e., that we have*

$$\ell(\delta')\ a\ \mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\rho,\sigma) \leq \mathop{\mathbb{E}}_{U\sim\nu} \mathrm{TV}(\mathbf{p}^{\rho,U},\mathbf{p}^{\sigma,U}) \leq u(\delta')\ A\ \mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\rho,\sigma)\ ,$$

for constants and functions specified in Theorem 1, with $\rho = \Lambda(\psi)$ *and* $\sigma = \frac{\mathbb{I}}{d}$*. The notation* $\mathbf{p}^{\Lambda(\psi),U}$ *is the same as for Theorem 1.*

From the above, it follows that the convergence of noisy distribution to the uniform in random circuits setting is controlled by the purity of the output state. For quantum measurements and channels, we have similar expressions.

Lemma 24. *[Noisy measurements – convergence to uniform distribution] Let* $\mathbf{M}$ *be a generic* d*-outcome quantum measurement on* d*-dimensional space, and* $\mathbf{M}^{\mathcal{I}}$ *a trivial POVM s.t.* $M_i^{\mathcal{I}} = \frac{\mathbb{I}}{d}$ *for each* $i = 1,\ldots,d$*. Then we have*

$$\mathop{\mathbb{E}}_{V\sim\nu} \mathrm{TV}(\mathbf{p}^{\mathbf{M},\psi_V},\mathbf{p}^{\mathrm{uniform}}) \approx \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathbf{M},\mathbf{M}^{\mathcal{I}}) = \frac{1}{2d}\sum_{i=1}^{d} \sqrt{\mathrm{tr}\left(M_i^2\right) + (\mathrm{tr}\, M_i - 1)^2 - \frac{1}{d}}\ . \tag{3.118}$$

In the above, the ≈ sign means the approximation in a sense of Eq. (3.24)*, i.e., that we have*

$$\ell(\delta')\ a\ \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathbf{M},\mathbf{N}) \leq \mathop{\mathbb{E}}_{V\sim\nu} \mathrm{TV}(\mathbf{p}^{\mathbf{M},\psi_V},\mathbf{p}^{\mathbf{N},\psi_V}) \leq u(\delta')\ A\ \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathbf{M},\mathbf{N})\ ,$$

for constants and functions specified in Theorem 2, with $\mathbf{N} = \mathbf{M}^{\mathcal{I}}$*. The notation* $\mathbf{p}^{\mathbf{M},\psi_V}$ *is the same as for Theorem 2.*

Lemma 25. *[Noisy channels – convergence to uniform distribution] Let* Λ *be a generic quantum channel and* Λ_{dep} *be a maximally depolarizing channel, i.e.,* $\Lambda_{\mathrm{dep}}(\rho) = \frac{\mathbb{I}}{d}$ *for any state* ρ*. Then we have*

$$\mathop{\mathbb{E}}_{V\sim\nu}\mathop{\mathbb{E}}_{U\sim\nu} \mathrm{TV}(\mathbf{p}^{\Lambda,\psi_V,U},\mathbf{p}^{uniform}) \approx \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Lambda_{\mathrm{dep}}) = \frac{1}{2}\sqrt{\mathrm{tr}\left(\mathcal{J}_\Lambda^2\right) + \mathrm{tr}\left(\left(\Lambda(\frac{\mathbb{I}}{d})\right)^2\right) - \frac{1}{d}\left(1+\frac{1}{d}\right)} \tag{3.119}$$

In the above, the ≈ sign means the approximation in a sense of Eq. (3.27)*, i.e., that we have*

$$\ell^{\mathrm{ch}}(\delta')\ a^{\mathrm{ch}}\ \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Gamma) \leq \mathop{\mathbb{E}}_{V\sim\nu}\mathop{\mathbb{E}}_{U\sim\nu} \mathrm{TV}(\mathbf{p}^{\Lambda,\psi_V,U},\mathbf{p}^{\Gamma,\psi_V,U}) \leq u^{\mathrm{ch}}(\delta')\ A^{\mathrm{ch}}\ \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Gamma)\ ,$$

for constants and functions specified in Theorem 3, with $\Gamma = \Lambda_{\mathrm{dep}}$*. The notation* $\mathbf{p}^{\Lambda,\psi_V,U}$ *is the same as for Theorem 3.*

3.7.2 Further examples

Example 15 (Two pure states). *For two pure states ψ and ϕ we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\psi,\phi) = \frac{1}{\sqrt{2}}\sqrt{1-\mathrm{tr}(\psi\phi)}\,,$$
$$\mathrm{d}_{\mathrm{tr}}(\psi,\phi) = \sqrt{1-\mathrm{tr}(\psi\phi)}\,.$$

Therefore, in this case, we see that $\mathrm{d}_{\mathrm{av}}(\psi,\phi) = \frac{1}{\sqrt{2}}\mathrm{d}_{\mathrm{tr}}(\psi,\phi)$, which gives only constant separation between average-case and worst-case scenarios.

The consequences of the above example are twofold, depending on the perspective we adopt. First, if we wish to perform a task of state discrimination between two pure states, then the above identity implies that there exists a strategy that uses random quantum circuits that is worse than the optimal strategy only by a constant. Second, if we treat ψ as our target state and ϕ as its noisy version affected by unwanted unitary rotation, then this type of noise will highly affect the quality of our results. Specifically, for generic quantum states, it will behave similarly to the worst-case scenario.

Example 16 (Pauli eigenstates and tensor product Pauli noise – general case). *Consider state $\psi^{pauli} = \otimes_{i=1}^{N} |{\pm}r_i\rangle\langle{\pm}r_i|$, where $r_i \in \{x,y,z\}$, i.e., $|{\pm}r_i\rangle$ is any Pauli eigenstate on qubit i (with eigenvalue $+1$ or -1.). Consider tensor product Pauli channel $\Lambda^{pauli} = \otimes_{i=1}^{N}\Lambda_i^{pauli}$, where single-qubit channel is $\Lambda_i^{pauli}(\rho) = \sum_{j=1} p_j^{(i)}\sigma_j\rho\sigma_j$ with $j \in \{1,x,y,z\}$, $\sigma_1 = \mathbb{I}$, and $p_j^{(i)} \geq 0$, $\sum_j p_j^{(i)} = 1$. Define $q^{(i)} = p_1^{(i)} + p_{r_i}^{(i)}$, i.e., a probability of applying on qubit i a gate that stabilizes the state of that qubit (namely, either identity or Pauli matrix of which $|{\pm}r_i\rangle$ is an eigenstate). Furthermore, assume that for each qubit i we have $q^{(i)} \geq \frac{1}{2}$. Then we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\Lambda^{pauli}(\psi^{pauli}),\frac{\mathbb{I}}{d}) = \frac{1}{2}\sqrt{\Pi_{i=1}^{N}\left(1-2q^{(i)}(1-q^{(i)})\right)-\frac{1}{d}}\,, \tag{3.120}$$

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\Lambda^{pauli}(\psi^{pauli}),\psi^{pauli}) = \frac{1}{2}\sqrt{1-2\Pi_{i=1}^{N}q^{(i)}+\Pi_{i=1}^{N}(1-2q^{(i)}(1-q^{(i)}))}\,, \tag{3.121}$$

Proof. We start by analyzing the effects of Pauli noise on single-qubit Pauli eigenstate. We first write $|{\pm}r_i\rangle\langle{\pm}r_i| = \frac{1}{2}\left(\mathbb{I} \pm \sigma_{r_i}\right)$ and evaluate

$$\Lambda_i^{\mathrm{pauli}}(|{\pm}r_i\rangle\langle{\pm}r_i|) = \frac{1}{2}\left(\mathbb{I} \pm \left((p_1^{(i)} + p_r^{(i)} - p_{k\neq r_i}^{(i)} - p_{l\neq r_i}^{(i)}\right)\right) = \tag{3.122}$$
$$= \frac{1}{2}\left(\mathbb{I} \pm \left((2(p_1^{(i)} + p_r^{(i)}) - 1\right)\right) = \tag{3.123}$$
$$= \frac{1}{2}\left(\mathbb{I} \pm \left(2q^{(i)} - 1\right)\right)\,, \tag{3.124}$$

where $p^{(i)}_{k\neq r_i}$ and $p^{(i)}_{l\neq r_i}$ are error probabilities corresponding to two Pauli matrices that are not σ_{r_i}. We now notice that the above state has two eigenvalues which are $\frac{1}{2}(1 \pm |2q^{(i)} - 1|)$, which for assumed regime $q^{(i)} \geq \frac{1}{2}$ gives eigenvalues $q^{(i)}$ and $(1 - q^{(i)})$.

To get Eq. (3.120) we refer to Lemma 23 and use the fact that the purity of tensor product states is a product of purities. For a single qubit, the purity of a noisy state is $\mathrm{tr}\left(\Lambda_i^{\mathrm{pauli}}\left(|\pm r_i\rangle\langle\pm r_i|\right)^2\right) = (q^{(i)})^2 + (1-q^{(i)})^2 = 1 - 2q^{(i)}(1-q^{(i)})$, which for multiple qubits yields Eq (3.120).

To get Eq. (3.121), we first diagonalize all noisy Pauli states, getting global state represented as $\bigotimes_{i=1}^{N}\left(q^{(i)}\,|0\rangle\langle 0| + (1-q^{(i)})\,|1\rangle\langle 1|\right)$. In this basis, the noiseless Pauli eigenstate is simply $|0\rangle\langle 0|^{\otimes N}$ (note that both states are simultaneously diagonalizable). Having this in mind, we want to decompose the distance between states $||\Lambda^{\mathrm{pauli}}(\psi^{\mathrm{pauli}}) - \psi^{\mathrm{pauli}})||^2_{HS}$ into parts that are easy to handle. To this aim, we use the fact that for any states ρ and $\tilde{\rho}$, the HS distance can be written as $||\rho - \tilde{\rho}||^2_{HS} = \mathrm{tr}\,\rho^2 + \mathrm{tr}\,\tilde{\rho}^2 - 2\,\mathrm{tr}\,(\rho\,\tilde{\rho})$. In our case $\rho = \Lambda^{\mathrm{pauli}}(\psi^{\mathrm{pauli}})$ and $\tilde{\rho} = \psi^{\mathrm{pauli}}$. Since the Pauli state is pure we get $\mathrm{tr}\,\tilde{\rho}^2 = 1$, while the purity of ρ was already calculated above. The cross-term can be evaluated by recalling that in basis we consider Pauli eigenstate is simply $|0\rangle\langle 0|^{\otimes N}$, we thus need to simply take the value of the first matrix element of ρ, obtaining $\mathrm{tr}\,(\rho\,\tilde{\rho}) = \Pi_i q^{(i)}$. Summing up and inserting into the definition of average-case distance yields Eq. (3.121). □

We note that the above derivation is a more general form of Example 5, in which we provide a discussion on its consequences for simplified figures of merit (averages of noise parameters over qubits).

We now consider a scenario where our target POVM is computational-basis measurement **P**, and we wish to calculate its distance from some other POVM **M**. This choice is motivated by the fact that in quantum computing the computational-basis measurement is often a model for ideal detector [17] , and **M** can be thought of as its noisy implementation. In particular, we considered a situation in which $\mathbf{M} = \mathsf{T}\,\mathbf{P}$, where T is a left-stochastic map, i.e., its columns' are probability distributions. Such noise is equivalent to classical post-processing of ideal statistics (i.e., probabilities one would have obtained on **P**), hence we call it classical noise. This is a practically relevant scenario, as it has been experimentally observed that classical noise is a dominant type of readout noise in contemporary quantum devices based on superconducting qubits [42].

We now define, in analogy to quantum average-case distance, the *average-case classical distance* between POVMs **M** and **N**

$$\mathrm{d}^{\mathrm{classical}}_{\mathrm{av}}(\mathbf{M},\mathbf{N}) := \mathop{\mathbb{E}}_{|k\rangle\langle k|} \mathrm{TV}\left(\mathbf{p}\left(|k\rangle\langle k|,\mathbf{M}\right),\mathbf{p}\left(|k\rangle\langle k|,\mathbf{N}\right)\right)\,, \tag{3.125}$$

where by $\mathop{\mathbb{E}}_{|k\rangle\langle k|}$ we denote average over all *classical* deterministic states $|k\rangle\langle k|$. The above distance turns out to be a helpful tool in investigating some of the

properties of quantum average-case distances for quantum measurements. In our considerations about the distances between measurements, the following Lemma 26 and Lemma 27 proved useful.

Lemma 26 (Average-case quantum vs classical distance). *Let* $\mathbf{P}$ *be measurement in computational basis, and* T *arbitrary stochastic map, i.e.,* $\sum_i T_{ij} = 1$. *Define POVM* $\mathsf{T}\mathbf{P}$ *via* $(\mathsf{T}\,\mathbf{P})_i = \sum_j \mathsf{T}_{ij} P_j$. *Then we have*

$$\frac{1}{2}\,\mathrm{d}_{\mathrm{av}}^{\mathrm{classical}}(\mathsf{T}\mathbf{P}, \mathbf{P}) \leq \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathsf{T}\,\mathbf{P}, \mathbf{P})\ . \tag{3.126}$$

Proof. We start by directly computing classical distance from Eq. (3.125)

$$\begin{aligned}\mathrm{d}_{\mathrm{av}}^{\mathrm{classical}}(\mathsf{T}\mathbf{P}, \mathbf{P}) &= \frac{1}{2d}\sum_{k=1}^{d}\sum_{i=1}^{d} |\operatorname{tr}(|k\rangle\langle k|\,(\sum_j T_{ij}\,|j\rangle\langle j| - |i\rangle\langle i|))| = \frac{1}{2d}\sum_{k=1}^{d}\sum_{i=1}^{d} |T_{ik} - \delta_{k,i}| \\ &= \frac{1}{2d}\sum_{k=1}^{d}(1 - T_{kk} + \sum_{i\neq k} T_{ik}) = \frac{1}{2d}\sum_{k=1}^{d} 2(1 - T_{kk}) = 1 - \frac{\operatorname{tr}(T)}{d}\ ,\end{aligned}$$

where we used the fact that T is left-stochastic, hence $\sum_{i\neq k} T_{ik} = 1 - T_{kk}$. Now we notice that

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathsf{T}\mathbf{P}, \mathbf{P}) = \frac{1}{2d}\sum_{i=1}^{d}\sqrt{(1 - \mathsf{T}_{ii})^2 + (1 - \sum_j T_{ij})^2 + \sum_{k\neq i} T_{ik}^2}\ \geq \tag{3.127}$$

$$\geq \frac{1}{2}\,\frac{1}{d}\sum_{i=1}^{d}\sqrt{(1 - T_{ii})^2} = \frac{1}{2}\,(1 - \frac{\operatorname{tr}(T)}{d})\ , \tag{3.128}$$

thus the expression on the RHS of Eq. (3.127) is exactly equal to $\frac{1}{2}\mathrm{d}_{\mathrm{av}}^{\mathrm{classical}}(\mathsf{T}\mathbf{P}, \mathbf{P})$, which concludes the proof. □

Lemma 27 (Distance of classical part of the measurement noise). *Let* $\mathbf{M}$ *be an arbitrary* d*-outcome POVM, and* $\mathbf{P}$ *measurement in the computational basis. Decompose* $\mathbf{M}$ *as* $\mathbf{M} = \mathsf{T}\,\mathbf{P} + \Delta$, *where* $\mathsf{T}\,\mathbf{P}$ *is a POVM obtained by taking only diagonal elements of operators* $\mathbf{M}$, *i.e.,* $(\mathsf{T}\,\mathbf{P})_i := \operatorname{diag}(M_i)$. *Then we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathsf{T}\,\mathbf{P}, \mathbf{P}) \leq \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathbf{M}, \mathbf{P})\ . \tag{3.129}$$

Proof. Consider an action of (the dual of) completely dephasing noise $\Lambda_{\mathrm{deph}}^{\dagger}$ on POVMs' effects, namely $\Lambda_{\mathrm{deph}}^{\dagger}(M_i) = (\mathsf{T}\mathbf{P})_i$ (this is because, to begin with, we defined $\mathsf{T}\,\mathbf{P}$ as diagonal part of POVM $\mathbf{M}$). Since dephasing noise is unital and it preserves computational-basis measurement $\mathbf{P}$, the above property follows directly from data-processing inequality for unital pre-processing of quantum measurements proved in Lemma 14. □

Remark 12. *We note that while the decomposition of POVM* $\mathbf{M} = \mathsf{T}\,\mathbf{P} + \Delta$ *into diagonal and off-diagonal parts may seem arbitrary, it has been in fact previously used in the context of measurement error mitigation. In particular, the* $\mathsf{T}\mathbf{P}$ *can be interpreted as a "classical part" of the noise and if we are able to reconstruct* T *(for example, using Diagonal Detector Tomography), we can use it to reduce the noise via classical post-processing of the statistics estimated on faulty detector* $\mathbf{M}$ *[65; 119; 120].*

Corollary 2. *By combining Lemma 26 with Lemma 27, we immediately get that for any POVM decomposed into the diagonal and off-diagonal part as* $\mathbf{M} = \mathsf{T}\,\mathbf{P} + \Delta$*, its distance from standard measurement can be bounded from below via*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathbf{M}, \mathbf{P}) \geq \frac{1}{2}\,\mathrm{d}_{\mathrm{av}}^{classical}(\mathsf{T}\mathbf{P}, \mathbf{P})\ . \tag{3.130}$$

Let us now consider a simplified scenario where the target POVM is the computational basis measurement, and its noisy version corresponds to local, symmetric classical noise.

Example 17 (Computational basis and local symmetric bitflip). *Let* $\mathbf{P}$ *denote measurement in computational basis and its noisy version* $\mathsf{T}^{sym}\mathbf{P}$ *affected by noise* $\mathsf{T} = \otimes_{i=1}^{N} \Lambda_i^{(\mathrm{sym})}$*, where* $\Lambda_i^{(\mathrm{sym})} = p^{(i)}\mathbb{I} + (1-p^{(i)})\sigma_x$ *denotes local stochastic noise describing symmetric bitflip specified by parameter* p_i *(bitflip error probability). In this case, we have*

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}\left(\mathsf{T}^{sym}\mathbf{P}, \mathbf{P}\right) = \frac{1}{2}\sqrt{1 - 2\Pi_{i=1}^{N}(1-p_i) + \Pi_{i=1}^{N}(1-2p_i(1-p_i))}, \tag{3.131}$$

$$\mathrm{d}_{\mathrm{op}}\left(\mathsf{T}^{sym}\mathbf{P}, \mathbf{P}\right) = 1 - \Pi_{i=1}^{N}(1-p_i)\ , \tag{3.132}$$

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathsf{T}^{sym}\mathbf{P}, \mathbf{M}^{\mathcal{I}}) = \frac{1}{2}\sqrt{\Pi_{i=1}^{N}(1-2p_i(1-p_i)) - \frac{1}{d}}\ , \tag{3.133}$$

where N *is the number of qubits.*

Proof. To obtain (3.131) we calculate explicitly

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathbf{P}, \mathsf{T}\mathbf{P}) = \frac{1}{2d}\sum_{i=1}^{d}\sqrt{(1-\mathsf{T}_{ii})^2 + (1-\sum_{j} T_{ij})^2 + \sum_{k\neq i} T_{ik}^2} = \tag{3.134}$$

$$= \frac{1}{2d}\sum_{i=1}^{d}\sqrt{(1-\mathsf{T}_{ii})^2 + \sum_{k\neq i} T_{ik}^2}\ , \tag{3.135}$$

where the first equality follows from the fact that T is bistochastic. Then we notice that for identical symmetric bitflip, each term on RHS is the same, namely for each i we have

$$(1-\mathsf{T}_{ii}^{\mathrm{sym}})^2 + \sum_{k\neq i}(\mathsf{T}_{ik}^{\mathrm{sym}})^2 = 1 - 2\mathsf{T}_{ii}^{\mathrm{sym}} + \sum_{k}(\mathsf{T}_{ik}^{\mathrm{sym}})^2\ . \tag{3.136}$$

Furthermore, from the product structure of T it follows that

$$\mathsf{T}^{\mathrm{sym}}_{kk} = \Pi_{i=1}^{N}(1-p_i) \tag{3.137}$$

$$,\sum_k (\mathsf{T}^{\mathrm{sym}}_{lk})^2 = \Pi_{i=1}^{N}((1-p_i)^2+p_i^2)\ . \tag{3.138}$$

Summing over $i = 1, \ldots, d$ yields Eq. (3.131).

To compute (3.132) we notice that in case of (any) stochastic noise T affecting standard measurement we have

$$\mathrm{d}_{\mathrm{op}}(\mathsf{T}\mathbf{P},\mathbf{P}) = \max_j(1-T_{jj})\ , \tag{3.139}$$

which after substituting T_{jj} from Eq. (3.137) yields Eq. (3.132).

To obtain Eq. (3.133), we first notice that multiqubit symmetric bitflip is represented by a bistochastic map that does not change trace – thus second term in Eq. (3.118) vanishes. Then we calculate explicitly the purity of ith effect as

$$\mathrm{tr}\left(((\mathsf{T}^{\mathrm{sym}}\mathbf{P})_i)^2\right) = \sum_k (\mathsf{T}^{\mathrm{sym}}_{ik})^2\ . \tag{3.140}$$

Combining the above observations with Eq. (3.137) yields Eq. (3.133). □

Lemma 28 (Computational basis and local asymmetric bitflip). *Consider a noisy version* $\mathsf{T}^{asym}\mathbf{P}$ *of computational basis measurement* $\mathbf{P}$*, where* $\mathsf{T}^{asym} = \otimes_{i=1}^{N}\mathsf{T}_i^{asym}$ *is a tensor product,* asymmetric *stochastic map. For each qubit* i*, such map is characterized by two parameters,* $p_i(1|0)$ *and* $p_i(0|1)$*, specifying the probability of erroneously measuring 1 (0) if the input state was* $|0\rangle$ *(*$|1\rangle$*). Define average error probability*

$$q_i^{av} = \frac{p_i(1|0)+p_i(0|1)}{2}\ , \tag{3.141}$$

and corresponding symmetric bitflip map $\mathsf{T}_i^{av} = (1-q_i^{av})\mathbb{I} + q_i^{av}\sigma_x$*, together with global map* $\mathsf{T}^{av} = \bigotimes_{i=1}^{N}\mathsf{T}_i^{av}$*. Then we have*

$$\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{T}^{av}\mathbf{P},\mathbf{P}) \leq \mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{T}^{asym}\mathbf{P},\mathbf{P})\ , \tag{3.142}$$

$$\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{T}^{av}\mathbf{P},\mathbf{M}^{\mathcal{I}}) \leq \mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathsf{T}^{asym}\mathbf{P},\mathbf{M}^{\mathcal{I}})\ . \tag{3.143}$$

Proof. The proof uses data-processing inequality for unital channels and stochastic post-processing proved in Lemma 14, as well as joint convexity property from Lemma 13. The idea is to present a strategy that "symmetrizes" stochastic noise on each qubit via randomized measurements and post-processing (while not changing computational basis measurement $\mathbf{P}$ or trivial POVM $\mathbf{M}^{\mathcal{I}}$). Consider a strategy that applies combinations of X and $\mathbb{I}$ gates uniformly at random just before measurement, and then applies a post-processing strategy that combines the outcomes of measurements to "undo" the effects of the unital channel.

Namely, for each qubit, if the applied gate was $\mathbb{I}$ do nothing, and if it was X then flip the outcome. Such a strategy does not affect the computational basis measurement nor a trivial POVM $\mathbf{M}^{\mathcal{I}}$, so we need to work out the expressions for twirled noisy measurement. The whole procedure corresponds to application of a so-called superchannel (i.e., a generalization of a quantum channel to operations that transform quantum channels themselves) [40], but we can derive the bounds for each "sub-channel" and then apply the joint-convexity property to derive the desired lower bound. More explicitly, for a given combination of X and $\mathbb{I}$ gates, let us label the corresponding unital pre-processing as Φ_α (this is application of given combination of gates) and corresponding post-processing as Λ_α (this is relabeling of measurement outcomes). We can write the resulting twirled noisy measurement as

$$\sum_\alpha p_\alpha \Lambda_\alpha \circ \mathrm{T}^{\mathrm{asym}}\mathbf{P} \circ \Phi_\alpha \ . \tag{3.144}$$

Note that in our case, $p_\alpha = \frac{1}{d}$. From Lemma 14, we get that for each element α, we have

$$\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathrm{T}^{\mathrm{asym}}\mathbf{P}, \mathbf{P}) \geq \mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\Lambda_\alpha \circ \mathrm{T}^{\mathrm{asym}}\mathbf{P} \circ \Phi_\alpha, \Lambda_\alpha \circ \mathbf{P} \circ \Phi_\alpha) \ . \tag{3.145}$$

We now apply the above inequality for each element of a trivial convex combination and make use of the joint-convexity property of the AC distance

$$\begin{aligned}
\mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathrm{T}^{\mathrm{asym}}\mathbf{P}, \mathbf{P}) &= \sum_\alpha p_\alpha \mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathrm{T}^{\mathrm{asym}}\mathbf{P}, \mathbf{P}) \geq \\
&\geq \sum_\alpha p_\alpha \mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\Lambda_\alpha \circ \mathrm{T}^{\mathrm{asym}}\mathbf{P} \circ \Phi_\alpha, \Lambda_\alpha \circ \mathbf{P} \circ \Phi_\alpha) \ \geq \\
&\geq \mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\sum_\alpha p_\alpha \Lambda_\alpha \circ \mathrm{T}^{\mathrm{asym}}\mathbf{P} \circ \Phi_\alpha, \sum_\alpha p_\alpha \Lambda_\alpha \circ \mathbf{P} \circ \Phi_\alpha) = \\
&= \mathrm{d}^{\mathrm{m}}_{\mathrm{av}}(\mathrm{T}^{\mathrm{av}}\mathbf{P}, \mathbf{P}) \ .
\end{aligned} \tag{3.146}$$

In the above, the final equality follows from direct calculation. Since the trivial POVM is also invariant under this procedure, the same argument applies and we conclude the proof.

We note that the above strategy was used for single-qubit error mitigation in Ref. [121], and more general multi-qubit versions were considered in the context of noise characterization and mitigation in Refs. [122; 123; 124]. We briefly discuss it further in Section 4.2.3 in the next Chapter. □

From the Lemma 28 it follows that when studying the separation between asymmetric stochastic noise and ideal measurement in the computational basis, one can instead study symmetric noise with "average" error probability (Eq. (3.141)), which is easier to handle computationally. The same holds for studying separation from a uniform distribution. The usefulness of this comes from the fact that asymmetric bitflip is a more realistic model of measurement noise than symmetric bitflip, (see, e.g., [42; 119]).

We now consider a few interesting scenarios for distances between channels.

Example 18 (Two arbitrary state preparation channels). *Denote by Λ_ρ and Λ_σ the state preparation channels that regardless of the input state always prepare state $\rho \in \mathsf{D}(\mathcal{H}_d)$ or $\sigma \in \mathsf{D}(\mathcal{H}_d)$, respectively. Then we have*

$$\mathsf{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_\rho, \Lambda_\sigma) = \sqrt{1 + \frac{1}{d}}\ \frac{1}{2}||\rho - \sigma||_{\mathrm{HS}} \,. \tag{3.147}$$

Example 19 (Two arbitrary unitary channels). *Denote by Λ_U and Λ_V the unitary channels associated with unitaries U and V, i.e., $\Lambda_U(\rho) = U\rho U^\dagger$ for any state $\rho \in \mathsf{D}(\mathcal{H}_d)$. Then we have*

$$\mathsf{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda_U, \Lambda_V) = \sqrt{\frac{1}{2}\left(1 - \frac{|\operatorname{tr}\left(U^\dagger V\right)|^2}{d^2}\right)} \,. \tag{3.148}$$

Example 20 (Identity channel and tensor product of unitary rotations). *Let $\mathcal{I}$ denote identity channel, and Λ_V be unitary channel corresponding to tensor-product rotation $V = \bigotimes_{j=1}^N \exp(i\ \mathbf{n}_j \cdot \boldsymbol{\sigma}\ \frac{\phi_j}{2})$, where $|\mathbf{n}_j| = 1$ and $\phi_j > 0$. Assume that $\sum_{j=1}^N \phi_j \leq \frac{\pi}{2}$. Define $\phi_{\max} = \max_j \phi_j$ and $\phi_{\min} = \min_j \phi_j$. Then we have*

$$\mathsf{d}_{\mathrm{av}}^{\mathrm{ch}}(\mathcal{I}, \Lambda_V) \leq \sqrt{N}\ \frac{\phi_{\max}}{\sqrt{8}} \,, \tag{3.149}$$

$$\mathsf{d}_{\diamond}(\mathcal{I}, \Lambda_V) \geq \frac{1}{\sqrt{2}} N \phi_{\min} \,. \tag{3.150}$$

To obtain the above, we first note that since the distances are unitarily invariant, we can rotate each unitary so it is a phase shift gate with an angle ϕ_j. To get the first inequality, we calculate explicitly (see Example 19) $\mathsf{d}_{\mathrm{av}}^{\mathrm{ch}}(\mathcal{I}, \Lambda_V) = \sqrt{\frac{1}{2}(1 - \prod_{j=1}^N \cos^2(\frac{\phi_j}{2}))}$. Then we use inequality $\cos^2(\frac{\phi_j}{2}) \leq \cos^2(\frac{\phi_{\max}}{2})$ for $\phi_j \in [0, \pi]$, and employ inequalities $\cos(x)^2 \geq 1 - x^2$ and $(1 - x)^N \geq 1 - Nx$. To get the second inequality we calculate diamond norm explicitly $\mathsf{d}_{\diamond}(\mathcal{I}, \Lambda_V) = 2|\sin(\sum_{j=1}^N \frac{\phi_j}{2})|$, and employ inequality $|\sin(x)| \geq \frac{x}{2\sqrt{2}}$ for $x \in \left[0, \frac{\pi}{2}\right]$.

From derivations in the above example it follows that if we adopt the perspective of average-case statistical distinguishability, any local coherent noise (when the target operation is identity) can be viewed simply as a phase shift error. Furthermore, for angles such that $\frac{\phi_{\max}}{\phi_{\min}} = O(1)$, we see that worst-case distance grows quadratically faster than average-case.

Example 21. *[Separable Pauli noise in the middle of the circuit – general case] Consider tensor product Pauli channel Λ^{pauli} defined in Example 16. Then we have*

$$\mathsf{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda^{pauli}, \Lambda_{dep}) = \frac{1}{2}\sqrt{\Pi_{i=1}^N ||\mathbf{p}^i||_2^2 - \frac{1}{d^2}} \,, \tag{3.151}$$

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda^{pauli}, \mathcal{I}) = \frac{1}{2}\sqrt{1 + \Pi_{i=1}^{N}||\mathbf{p}^i||_2^2 - 2\Pi_{i=1}^{N} p_1^i}\,, \tag{3.152}$$

where $||\mathbf{p}^i||_2^2 = \sum_j (p_j^i)^2$ *is a Euclidean norm of the vector of noise coefficients on* i*th qubit.*

Proof. To begin the proof, we notice that the Pauli noise is a mixed unitary channel and is thus unital. Since both completely depolarizing and identity channels are unital as well, in both average-case distances the terms that include the difference of the images of the maximally-mixed state under action of said channels equal 0 . We are therefore left with the task of calculating Hilbert-Schmidt norms of relevant Choi matrices.

To show that Eq. (3.151) holds, we note that the purity of a tensor product of Choi states is a product of purities – this follows from the fact that any Choi matrix of product channel is permutationally similar to a tensor product of Choi matrices of those channels. We thus need to consider only single-qubit purity (note that this is analogous to proof for states in Example 1). Denote by $\mathcal{J}_{\mathrm{pauli}}^{(i)}$ a Choi matrix of Pauli channel on qubit i. By directly evaluating the action of that channel on operators of the form $|k\rangle\langle l|$ (recall the definition of Choi matrix) we explicitly write down matrix representation of $\mathcal{J}_{\mathrm{pauli}}^{(i)}$ and calculate

$$4\,\mathrm{tr}\left(\mathcal{J}_{\mathrm{pauli}}^{(i)}\right)^2 = \mathrm{tr}\left(\Lambda_{\mathrm{pauli}}^{(i)}(|0\rangle\langle 0|)\right)^2 + \mathrm{tr}\left(\Lambda_{\mathrm{pauli}}^{(i)}(|1\rangle\langle 1|)\right)^2 + \tag{3.153}$$

$$+ 2\,\mathrm{tr}\left(\Lambda_{\mathrm{pauli}}^{(i)}(|0\rangle\langle 1|)\left(\Lambda_{\mathrm{pauli}}^{(i)}(|0\rangle\langle 1|)\right)^{\dagger}\right) \tag{3.154}$$

From direct evaluation, we get that

$$\mathrm{tr}\left(\Lambda_{\mathrm{pauli}}^{(i)}(|k\rangle\langle k|)\right)^2 = (p_1 + p_{z_i})^2 + (p_{x_i} + p_{y_i})^2 \tag{3.155}$$

and

$$\mathrm{tr}\left(\Lambda_{\mathrm{pauli}}^{(i)}(|k\rangle\langle l|)\left(\Lambda_{\mathrm{pauli}}^{(i)}(|k\rangle\langle l|)\right)^{\dagger}\right) = (p_1 - p_{z_i})^2 + (p_{x_i} - p_{y_i})^2 \tag{3.156}$$

for $k \neq l$. Summing up everything we get that cross-terms cancel and $||\mathcal{J}_{\mathrm{pauli}}^{(i)}||_{HS}^2 = \sum_j p_j^{(i)} = ||\mathbf{p^{(i)}}||_2^2$ which combined with Lemma 25 yields Eq. (3.151).

To get Eq. (3.152) we follow an identical strategy as for Example 16. Namely, we recall the fact that for any states ρ and $\tilde{\rho}$, the HS distance can be written as $||\rho - \tilde{\rho}||_{HS}^2 = \mathrm{tr}\,\rho^2 + \mathrm{tr}\,\tilde{\rho}^2 - 2\,\mathrm{tr}\,(\rho\,\tilde{\rho})$. Now in our case $\rho = \mathcal{J}_{\Lambda^{\mathrm{pauli}}}$ and $\tilde{\rho} = \mathcal{J}_{\mathcal{I}}$. The Choi of the identity channel is a maximally-entangled state, its purity is thus equal to 1, while the purity of the Choi of the noisy channel was already calculated above. To evaluate cross-term, we note that it factorizes into a product of single-qubit terms (as for purity, it follows from the permutational equivalence

between the Choi matrix of product channel and tensor product of Choi matrices), each of them being equal to

$$\mathrm{tr}\left(\mathcal{J}^{(i)}_{\mathrm{pauli}}\mathcal{J}^{(i)}_{\mathcal{I}}\right) = \frac{1}{4}\sum_{k,l\in\{0,1\}} \mathrm{tr}\left(\Lambda\left(|k\rangle\langle l|\right)|l\rangle\langle k|\right) . \tag{3.157}$$

This evaluates to

$$\mathrm{tr}\left(\Lambda^{(i)}_{\mathrm{pauli}}\left(|k\rangle\langle k|\right)|k\rangle\langle k|\right) = p^{(i)}_1 + p^{(i)}_{z_i} , \tag{3.158}$$

and

$$\mathrm{tr}\left(\Lambda^{(i)}_{\mathrm{pauli}}\left(|k\rangle\langle l|\right)|k\rangle\langle l|\right) = p^{(i)}_1 - p^{(i)}_{z_i} , \tag{3.159}$$

for $k \neq l$. Summing up we obtain $\mathrm{tr}\left(\mathcal{J}^{(i)}_{\mathrm{pauli}}\mathcal{J}_{\mathcal{I}}\right) = p^{(i)}_1$. Combining all of the above with the definition of average-case distance yields Eq. (3.152). □

3.7.3 Proof of claims in Examples 1, 2, 3, and 4

Now we are ready to prove Examples 1, 2, 3, and 4 from Section 3.5.

Uncorrelated Pauli noise and symmetric measurement noise

Recall that in Examples 1 and 4 we consider product Pauli channels. Specifically, we have a Pauli channel $\Lambda^{\mathrm{pauli}} = \otimes_{i=1}^{N}\Lambda_i^{\mathrm{pauli}}$, where single-qubit channel is $\Lambda_i^{\mathrm{pauli}}(\rho) = \sum_{j=1}\ p_j^{(i)}\sigma_j\rho\sigma_j$ with $j \in \{1, x, y, z\}$, $\sigma_1 = \mathbb{I}$, and $p_j^{(i)} \geq 0$, $\sum_j p_j^{(i)} = 1$. It is useful to define $q^{(i)} = p^{(i)}_1 + p^{(i)}_{r_i}$, i.e., a probability of applying on qubit i a gate that stabilizes the state of that qubit (namely, either identity or Pauli matrix of which $|\pm r_i\rangle$ is an eigenstate). Moreover, we will consider average properties of noise as $q^{av} = \frac{1}{N}\sum_{i=1}^{N} q^{(i)}$ and $f^{av} = \frac{1}{N}\sum_{i=1}^{N} q^{(i)}(1-q^{(i)})$. In Example 2, we consider a stochastic measurement noise that is uncorrelated, and symmetric. Due to that last property, it can be modeled as a probabilistic application of Pauli X gate, i.e., a special type of Pauli channel. Therefore, Examples 1, 2, and 4 all consider the same type of noise, which we analyze now.

In what follows, we will derive an upper bound for Eq. (3.120) and a lower bound for Eq. (3.121), which correspond to inequalities Eq. (3.48) and Eq. (3.49) from Example 1, respectively.

To begin, we consider a function $f^{(i)} = q^{(i)}(1-q^{(i)})$, as well as average noise properties $q^{av} = \frac{1}{N}\sum_{i=1}^{N} q^{(i)}$ and $f^{av} = \frac{1}{N}\sum_{i=1}^{N} f^{(i)}$. We then bound Eq. (3.120) from above as

$$\sqrt{\Pi_{i=1}^{N}\left(1-2f^{(i)}\right) - \frac{1}{d}} \leq \sqrt{\Pi_{i=1}^{N}\left(1-2f^{(i)}\right)} , \tag{3.160}$$

and continue with bounding the (positive) expression inside square root as

$$\Pi_{i=1}^{N}\left(1-2f^{(i)}\right)=\left(\sqrt[N]{\Pi_{i=1}^{N}\left(1-2f^{(i)}\right)}\right)^{N}\leq\left(\frac{\sum_{i=1}^{N}(1-2f^{(i)})}{N}\right)^{N}= \tag{3.161}$$

$$=(1-2f^{av})^{N}\leq\exp(-2f^{av}N)\,, \tag{3.162}$$

where in first inequality we used inequality between geometric and arithmetic means together with a fact that $x^N \geq y^N$ for $x > y > 0$. In second inequality we used that for $0 \leq x \leq 1$ and $N \geq 1$, we have $(1-x)^N \leq \exp(-xN)$. Note that each term $2f^{(i)}$ lies in interval $2f^{(i)} \in \left[0, \frac{1}{2}\right]$. Combining everything we obtain

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\Lambda^{\mathrm{pauli}}(\psi^{\mathrm{pauli}}),\frac{\mathbb{I}}{d})\leq\frac{1}{2}\exp(-f^{av}N)\,, \tag{3.163}$$

which concludes the proof of first bound.

To bound Eq. (3.121) from below, we start by again employing inequality between geometric and arithmetic mean, namely

$$1-2\Pi_{i=1}^{N}q^{(i)}=1-2\left(\sqrt[N]{\Pi_{i=1}^{N}q^{(i)}}\right)^{N}\geq 1-2\left(\frac{\sum_{i=1}q^{(i)}}{N}\right)^{N}=1-2\,(q^{av})^{N}\,, \tag{3.164}$$

which after combining with Eq. (3.121) yields

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{s}}(\Lambda^{\mathrm{pauli}}(\psi^{\mathrm{pauli}}),\psi^{\mathrm{pauli}})\geq\frac{1}{2}\sqrt{1-2\,(q^{av})^{N}+\Pi_{i=1}^{N}(1-2q^{(i)}(1-q^{(i)}))}\;\geq \tag{3.165}$$

$$\geq\frac{1}{2}\sqrt{1-2\,(q^{av})^{N}}\,. \tag{3.166}$$

The above bound is valid provided that argument is still contained in the domain of square root, i.e., we need to impose

$$1-2\,(q^{av})^{N}\geq 0\implies q^{av}\leq\sqrt[N]{\frac{1}{2}}\,. \tag{3.167}$$

Note that $\sqrt[N]{\frac{1}{2}} \xrightarrow{N\to\infty} 1$, and since q^{av} is by definition lower than 1, the bound becomes less restrictive for higher system sizes. For small systems it is valid only for high noise (small q^{av}), but in such cases one can simply use the exact expressions from Eqs. (3.120) and (3.121).

The exactly same reasoning is applied for Examples 2 and 4, for which all expressions have almost the same functional forms (see Examples 17 and 21).

Generic uncorrelated measurement noise

Finally, recall that in Example 3, we consider a more general measurement noise applied to computational basis measurement $\mathbf{P} = (|\mathbf{x}\rangle\langle\mathbf{x}|)_{\mathbf{x}\in\{0,1\}^N}$ on N qubit system. We denote $\mathbf{M} = (M_{\mathbf{x}})_{\mathbf{x}\in\{0,1\}^N}$ to be a POVM specified by effects $M_{\mathbf{x}} = \Lambda_1^\dagger(|x_1\rangle\langle x_1|) \otimes \ldots \otimes \Lambda_N^\dagger(|x_N\rangle\langle x_N|)$, where Λ_i are quantum channels affecting i'th qubit, and $\Lambda_i^\dagger$ is the conjugate of Λ_i. We define classical success probability as $p^{(i)}(\mathbf{x}_i|\mathbf{x}_i) = \mathrm{tr}\left(\Lambda_i^\dagger\left(|x_i\rangle\langle x_i|\right)|x_i\rangle\langle x_i|\right)$, the corresponding average as $q_{av}^{(i)} = \frac{p^i(0|0)+p^{(i)}(1|1)}{2}$, and average over qubits $q^{av} := \frac{1}{N}\sum_{i=1}^N q_{av}^{(i)}$. Since this model is more general than Pauli noise, it requires separate treatment.

We will now derive the lower bound on the distance between the above-described noisy measurement and computational basis measurement, that is the inequality statement in Eq. (3.51) from Example 3. The first part of the proof is now slightly more involved due to more general noise model considered, and its goal is to reduce the noise form to a stochastic, symmetric measurement noise model that we considered above. To prove the claim, first one applies maximally-dephasing channel to both measurements and uses data-processing inequality for average-case distance to bound the distance from below by the diagonal part of the POVM $\mathbf{M}$. Specifically, define dephased POVM $\Phi_{\mathrm{dep}}(\mathbf{M})$ via its effects $\Phi_{\mathrm{dep}}(\mathbf{M})_i = \Phi_{\mathrm{dep}}(M_i)$, where maximally dephasing channel acts on any operator A as $\Phi_{\mathrm{dep}}(A) = \mathsf{diag}(A)$, with $\mathsf{diag}(A)$ denoting diagonal part of A. Note that for compuational basis measurement $\mathbf{P}$ we have $\Phi_{\mathrm{dep}}(\mathbf{P}) = \mathbf{P}$. Thus we have

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\Phi_{\mathrm{dep}}(\mathbf{M}), \Phi_{\mathrm{dep}}(\mathbf{P})) \geq \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\Phi_{\mathrm{dep}}(\mathbf{M}), \mathbf{P}) \,. \tag{3.168}$$

The above allows to treat noise as classical and look only on assignment infidelities for classical states (i.e., error probabilites when measured states are computational-basis states). Note that, importantly, maximally dephasing channel does not change the product structure of $\mathbf{M}$. Thus we can treat this dephased POVM $\Phi_{\mathrm{dep}}(\mathbf{M})$ as related to computational basis measurement via some tensor product of stochastic maps $\mathsf{T} = \bigotimes_{i=1}^N \mathsf{T}^{(i)}$, where $\mathsf{T}^{(i)}$ acts on ith qubit and is specified by two success probabilities $p^{(i)}(0|0)$ and $p^{(i)}(1|1)$ (see, for example, Ref. [42] and Chapter 4 for more details on stochastic readout noise). Thus we have

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathbf{M}, \mathbf{P}) \geq \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathsf{T}\mathbf{P}, \mathbf{P}) \,, \tag{3.169}$$

where $\mathsf{T}\mathbf{P}$ is a POVM with ith effect given by $(\mathsf{T}\mathbf{P})_i = \sum_i \mathsf{T}_{ij}\,|j\rangle\langle j|$ and stochastic map T is defined via diagonal elements of original POVM $\mathbf{M}$ (as in discussion above).

Now one applies Lemma 28 that lower bounds the distance via symmetrized version of T, where now both error probabilities are the same and equal to $q_{av}^{(i)} = \frac{p^{(i)}(0|0)+p^{(i)}(1|1)}{2}$ (note that this is equivalent to Pauli bitflip channel applied with probability $q_{av}^{(i)}$). Denote such symmetrized version of T as $\mathsf{T}^{\mathrm{sym}}$. This gives

$$\mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathsf{T}\mathbf{P}, \mathbf{P}) \geq \mathrm{d}_{\mathrm{av}}^{\mathrm{m}}(\mathsf{T}^{\mathrm{sym}}\mathbf{P}, \mathbf{P}). \tag{3.170}$$

Therefore we reduced the lower bound to scenario considered in Example 2, for which the bound was proved above in the previous subsection.

3.8 Summary

In this Chapter, we introduced and analyzed a novel approach to measuring similarity between quantum objects (states, measurements, and channels) based on their average-case statistical distinguishability through random quantum circuits. Unlike conventional distance measures such as trace distance or diamond norm that focus on optimal distinguishability protocols, our average-case quantum distances capture the typical behavior of quantum objects in experiments involving low-depth quantum circuits – approximate unitary 4-design that can be implemented in depth $\log(N)$ (N being number of qubits).

The core theoretical contribution is proving that for circuits forming approximate 4–designs, the average Total Variation Distance between output statistics can be approximated by simple explicit functions expressible as degree-2 polynomials in the underlying quantum objects. These functions, which we denoted as average-case (AC) distances, possess several desirable properties including subadditivity with respect to tensor products, joint convexity, and restricted data-processing inequalities. Notably, all distances include terms utilizing the Hilbert-Schmidt norm, providing this norm with a new operational interpretation. We also established bounds on the relationship between worst-case and average-case distances, showing that their ratio is at most $d^{1/2}$, d, and $d^{3/2}$ for quantum states, measurements, and channels, respectively. Through explicit examples, we demonstrated that these bounds are tight.

Our results have potentially significant practical implications for analysis of quantum computing protocols. The average-case distances provide a more realistic assessment of device performance than worst-case measures, particularly for near-term applications where circuits will typically be of moderate depth. In particular, we have demonstrated that our tools can be used to study average-case convergence (resp. divergence) of noisy distributions to uniform, useless distribution (resp. from ideal, noiseless distribution). This provides a new way to study effects of noise on quantum computing experiments.

Importantly, AC distances demonstrate that, in principle, an efficient discrimination of high-dimensional quantum objects is possible with low-depth quantum circuits if one allows for classical randomness (however, this is, in general, at the cost of exponential classical processing requirements). Moreover, we have discussed consequences of our findings to certain complexity measures of pure quantum states and unitary channels.

We supplemented our findings with numerical simulations that further validated that AC distances might be better suited for quantifying the quality of noisy quantum computing protocols (such as variational optimization), as compared to standard worst-case measures.

4. Modeling and mitigation of correlated readout noise

4.1 Chapter overview

4.1.1 Summary in the context of the thesis

One of the main impediments in the execution of measurements on quantum computing apparatus is the pervasive influence of noise. In this chapter, we introduce a scalable method of characterizing correlated measurement noise based on so-called quantum overlapping tomography – we call the method Diagonal Detector Overlapping Tomography (DDOT). In conjunction with DDOT, we present classical post-processing methods designed to reduce the effects of noise in the task of, among others, estimating local quantum observables. We test our methods in experiments on superconducting quantum devices developed by IBM and Rigetti.

The chapter is based on work [3] which the author of this Thesis co-authored. It contains large excerpts from that publication, with edits necessary to adjust them for the format of a Ph.D. thesis.

4.1.2 Technical abstract

Measurement noise is one of the main sources of errors in currently available quantum devices based on superconducting qubits. At the same time, the complexity of its characterization and mitigation often exhibits exponential scaling with the system size. In this work, we introduce a correlated measurement noise model that can be efficiently described and characterized, and which admits effective noise-mitigation on the level of marginal probability distributions. Noise mitigation can be performed up to some error for which we derive upper bounds. Characterization of the model is done efficiently using Diagonal Detector Overlapping Tomography – a generalization of the recently introduced Quantum Overlapping Tomography to the problem of reconstruction of readout noise with restricted locality. The procedure allows to characterize k-local measurement cross-talk on N-qubit device using $O\left(k2^k \log\left(N\right)\right)$ circuits containing random combinations of X and identity gates. We perform experiments on 15 (23) qubits using IBM's (Rigetti's) devices to test both the noise model and the error-mitigation scheme, and obtain an average reduction of errors by a factor > 22 (> 5.5) compared to no mitigation. Interestingly, we find that correlations in the measurement

noise do not correspond to the physical layout of the device. Furthermore, we study numerically the effects of readout noise on the performance of the Quantum Approximate Optimization Algorithm (QAOA). We observe in simulations that for numerous objective Hamiltonians, including random MAX-2-SAT instances and the Sherrington-Kirkpatrick model, the noise-mitigation improves the quality of the optimization. Finally, we provide arguments why in the course of QAOA optimization the estimates of the local energy (or cost) terms often behave like uncorrelated variables, which greatly reduces sampling complexity of the energy estimation compared to the pessimistic error analysis. We also show that similar effects are expected for Haar-random quantum states and states generated by shallow-depth random circuits.

4.2 Introduction

4.2.1 Motivation

Outstanding progress has been made in the last years on the path to development of scalable and fully functional quantum devices. With state of the art quantum processors reaching a scale of 50-100 qubits [125], the scientific community is approaching a regime in which quantum systems cannot be modeled on modern supercomputers using known methods [126]. This situation is unarguably exciting since it opens possibilities for the first demonstrations of quantum advantage and, potentially, useful applications [127; 128; 129]. At the same time, however, it entails a plethora of non-trivial problems related to fighting effects of experimental imperfections on the performance of quantum algorithms. As quantum devices in the near-term will be unable to implement proper error correction [129], various methods of *noise mitigation* have been recently developed [130; 131; 132; 133; 134; 135; 136; 42; 41; 120]. Those methods aim at reducing the effects of errors present in quantum gates and/or in quantum measurements. In this work, we focus on the latter, i.e., noise affecting quantum detectors.

Indeed, it has been found in currently available quantum devices that the noise affecting measurements is quite significant. Specifically, errors of the order of a few percent in a single qubit measurement and non-negligible effects of cross-talk were reported [125; 42; 41; 120; 119]. Motivated by this, a number of methods to characterize and/or reduce measurement errors were proposed [42; 41; 119; 120; 137; 138; 139; 140; 141; 142; 143]. Readout noise mitigation usually relies on the classical post-processing of experimental statistics, preceded by some procedure of noise characterization. Existing techniques typically suffer from the curse of dimensionality due to characterization cost, sampling complexity, and the complexity of post-processing – which scale exponentially with the number of qubits N. Fortunately, some interesting problems in quantum computing do not require measurements to be performed across the whole system. An important class of algorithms that have this feature are the Quantum Approximate Optimization Algorithms (QAOA) [144], which only require the estimation of a number of few-particle *marginals*. QAOA

is a heuristic, hybrid quantum-classical optimization technique [144; 145], that was later generalized as a standalone ansatz [146] shown to be computationally universal [147; 148]. In its original form, QAOA aims at finding approximate solutions for hard combinatorial problems, notable examples of which are maximum satisfiability (SAT) [109], maximum cut (MAXCUT) [149], or the Sherrington-Kirkpatrick (SK) spin-glass model [150; 58]. Regardless of the underlying problem, the main QAOA subroutine is the estimation of the energy of the local classical Hamiltonians on a quantum state generated by the device. Those Hamiltonians are composed of a number of a few-body commuting operators, and hence estimation of energy can be done via estimation of the local terms (which can be performed simultaneously). Since the estimation of marginal distributions is the task of low sampling and post-processing complexity, this suggests that error mitigation techniques can be efficiently applied in QAOA. In this work, we present a number of contributions justifying the usage of measurement error-mitigation in QAOA, even in the presence of significant cross-talk effects.

4.2.2 Summary of results

Our first contribution is to provide an efficiently describable measurement noise model that incorporates asymmetric errors and cross-talk effects. Importantly, our noise model admits efficient error-mitigation on the marginal probability distributions, which can be used, e.g., for improvement of the performance of variational quantum algorithms. We show how to efficiently characterize the noise using a number of circuits scaling logarithmically with the number of qubits. To this aim, we generalize the techniques of the recently introduced Quantum Overlapping Tomography (QOT) [151] to the problem of readout noise reconstruction. Specifically, we introduce notion of Diagonal Detector Overlapping Tomography (DDOT) which allows to reconstruct noise description with k-local cross-talk on N-qubit device using $O\left(k2^k \log\left(N\right)\right)$ quantum circuits consisting of single layer of X and identity gates. Furthermore, we explain how to use that characterization to mitigate noise on the marginal distributions and provide a bound for the accuracy of the mitigation. Importantly, assuming that cross-talk in readout noise is of bounded locality, the sampling complexity of error-mitigation is not significantly higher than that of the starting problem of marginals estimation.

We test our error-mitigation method in experiments on 15 qubits using IBM's device and on 23 qubits using Rigetti's device, both architectures based on superconducting transmon qubits [152]. We obtain a significant advantage by using a correlated error model for error mitigation in the task of ground state energy estimation. Interestingly, the locality structure of the reconstructed errors in these devices does not match the spatial locality of qubits in these systems.

We also study statistical errors that appear in the simultaneous estimation of multiple local Hamiltonian terms that appear frequently in QAOAs. In particular, we provide arguments why one can expect that the estimated energies of local terms behave effectively behave as uncorrelated variables, for the quantum states appearing at the beginning and near the end of the QAOA algorithm. This allows to prove significant reductions in sampling complexity of the total energy

estimation (compared to the worst-case upper bounds).

Finally, we present a numerical study and detailed discussion of the possible effects that measurement noise can have on the performance of the Quantum Approximate Optimization Algorithm. This includes the study of how noise distorts the quantum-classical optimization in QAOAs. We simulate the QAOA protocol on an 8-qubit system affected by correlated measurement noise inspired by the results of IBM's device characterization. For a number of random Hamiltonians, we conclude that the error-mitigation highly improves the accuracy of energy estimates, and can help the optimizer to converge faster than when no error-mitigation is used.

4.2.3 Related works

The effects of simple, uncorrelated noisy quantum channels on QAOAs were analyzed in Refs. [153; 154; 155]. In particular, the symmetric bitflip noise analyzed in [153] can be used also to model symmetric, uncorrelated measurement noise - a type of readout noise which has been demonstrated to be not very accurate in currently available devices based on transmon qubits [42; 119]. A simple readout noise-mitigation technique on the level of marginals for QAOAs was recently experimentally implemented on up to 23 qubits in a work by Google AI Quantum team and collaborators [21]. Importantly, the authors assumed uncorrelated noise on each qubit – we believe that our approach which accounts for correlations could prove beneficial in those kinds of experiments. In a similar context, readout noise-mitigation using classical post-processing on global probability distributions was implemented in QAOA and Variational Quantum Eigensolver (VQE) experiments on up to 6 qubits [156; 157]. While for such small system sizes it is possible to efficiently perform global error mitigation, we emphasize that our approach based on the noise-mitigation on the marginals could be performed also in larger experiments of this type.

Alternative methods of characterization of correlated measurement noise were recently proposed in Refs. [119; 140; 120]. Out of the above-mentioned works, perhaps the most related to ours in terms of studied problems is Ref. [120], therefore we will now comment on it thoroughly. The authors introduced a correlated readout noise model based on Continuous Time Markov Processes (CTMP). They provide both error-characterization and error-mitigation methods. The CTMP noise model assumes that the noisy stochastic process in a measurement device can described by a set of two-qubit generators with corresponding error rate parameters. The total number of parameters required to describe a generic form of such noise for N-qubit device is $2N^2$. The authors propose a method of noise characterization that requires preparation of a set of suitably chosen classical states and performing a post-selection on the noise-free outcomes (i.e., correct outcomes given known input classical state) on the subset of $(N-2)$ qubits. Since probability of noise-free outcomes can be exponentially small even for the uncorrelated readout noise, this method can prove relatively costly in practice. We believe that our DDOT characterization technique could prove useful in reducing the number of parameters needed to describe CTMP model (by show-

ing which pairs of qubits are correlated, and for which the cross-talk can be neglected), hence allowing for much more efficient version of noise reconstruction presented in Ref. [120]. The authors also present a novel noise-mitigation method that allows to estimate the noise-reduced expected values of observables. The method is based on decomposition of inverse noise matrix into the linear combination of stochastic matrices, and constructing a random variable (with the aid of classical randomness and post-processing) that agrees in the expected value of noise-free observable. In general, both sampling complexity and classical post-processing cost scale exponentially with the number of qubits. Since the authors aim to correct observables with arbitrary locality, the problem they consider is different to our approach that aims to correct only *local* observables (and therefore does not exhibit exponential scalings). It is an interesting problem to see whether the CTMP noise model can also be interpreted on the level of marginal probability distributions in a way that allows for mitigation analogous to ours in terms of complexity.

Another related method was developed in Ref. [122] (see also [123; 124]), where the authors propose to use random implementation of X gates, together with a simple classical post-processing, to symmetrize (or "twirl") a generic stochastic measurement noise (we note that we used similar twirling trick to prove Lemma 28 in Chapter 3). Intuitively, the twirling procedure makes the error probabilities independent on the corresponding ideal measurement outcome (i.e., all columns of the noise matrix T become identical up to permutation of their elements). For example, in the case of single-qubit stochastic map with error probabilities $p(0|1)$ and $p(1|0)$, the twirling procedure transforms the effective noise model to symmetric stochastic map described by a single parameter $\frac{p(0|1)+p(1|0)}{2}$ (recall Lemma 28 and Example 17). When estimating expected values of Pauli Z operators, such a symmetric measurement noise results in rescaling of the expected values. Ref. [122] then proposes to estimate the rescaling factors by implementation of random circuits, and using them as a noise mitigation technique for expected values (by simply dividing the estimators by corresponding factors). This simplification of the noise model comes at a cost of randomized implementation of the twirling gates for every experiment (not only for calibration) and an increase in sample complexity (similarly to most quantum error mitigation techniques [34]). We note that the DDOT characterization technique that we introduce in Section 4.4 could be used to estimate the above-mentioned noise rescaling factors. Moreover, the measurement noise models we consider in this Chapter (see Section 4.3) could, in principle, benefit from such twirling (in a sense of reducing the number of parameters needed to describe a given model). Making connections between our techniques and similar randomized readout error-mitigation/characterization methods [122; 123; 124] remains an interesting future research direction.

4.2.4 Structure of the chapter

In Section 4.3 we discuss correlations in classical readout noise and introduce a "clusters and neighbors" model that we propose to effectively model them.

In Section 4.4, we introduce the Diagonal Detector Overlapping Tomography (DDOT), a method to efficiently characterize local readout noise using certain collections of very simple quantum circuits. The section includes experimental results of noise characterization performed on IBM's and Rigetti's quantum devices. In Section 4.5, we describe how the results of DDOT can be used for efficient (approximate) noise mitigation on the level of marginal probability distributions. This allows to reduce effects of readout noise, for example, on the protocols that involve the estimation of energy of local Hamiltonians. In Section 4.6, we experimentally benchmark the proposed noise mitigation strategy on various classes of such Hamiltonians in IBM's and Rigetti's quantum devices. Then, in Section 4.7, we focus our attention on Quantum Approximation Optimization Algorithm (QAOA) protocol, and investigate how different approximations necessary for our noise-mitigation technique contribute to the errors in the final energy estimates. The section is followed by Section 4.8, where we present extensive numerical studies on how the correlated measurement noise affects QAOA optimization. Finally, in Section 4.9, we summarize our findings.

4.3 Correlated readout noise model

4.3.1 Preliminaries

Recall from Section 2.3.2 that, to a good approximation, measurement noise can be modeled by a stochastic map T. Due to the linearity of Born's rule, it follows that probabilities $\mathbf{p}^{noisy}$ from which noisy detector samples are related to the noiseless probabilities $\mathbf{p}^{ideal}$ via the same stochastic map (recall Equation (2.3))

$$\mathbf{p}^{noisy} = \mathsf{T}\mathbf{p}^{ideal} \,. \tag{4.1}$$

Then the noise can by reversed by first estimating LHS, and then multiplying it by T^{-1}. Note that while this method is perhaps the most natural (and simple) method to reduce the noise and has been shown to be useful in practical situations [41; 42], there exist more sophisticated techniques of noise-mitigation that do not exhibit this problem. For example, Iterative Bayesian Unfolding [138] always returns physical probability vectors.

4.3.2 Correlations in readout noise

The size of the matrix T scales exponentially with the number of qubits. Thus, if one wants to estimate such a generic T using standard methods, both the number of circuits and sampling complexity scale exponentially. Indeed, the standard method of reconstructing T is to create all the 2^N computational basis and estimate the resulting probability distributions (which constitute the columns of T). We refer to such characterization as Diagonal Detector Tomography (DDT), since it probes the diagonal elements of the measurement operators describing the detector. This is restricted version of more general Quantum Detector Tomography (QDT) [80; 158; 159; 160].

These complexity issues can be circumvented if one assumes some *locality* structure in the measurement errors. For example, in the simplest model with completely uncorrelated readout noise, the T matrix is a simple tensor product of single-qubit error matrices T_{Q_i}

$$\mathsf{T} = \bigotimes_i \mathsf{T}_{Q_i} \qquad \text{(uncorrelated noise)}. \tag{4.2}$$

In this model, we need to estimate only single-qubit matrices, which renders the complexity of the problem to be linear in the number of qubits. However, for contemporary quantum devices based on superconducting qubits, it was demonstrated that such a noise model is not very accurate due to the cross-talk effects [125; 42; 41; 119]. At the same time, the completely correlated noise is not realistic as well, which motivates the search for a model that can account for correlations in readout errors while still giving an efficient description of T.

In this work, we propose such a model and give a method to characterize it. Let us lay out the basic concepts of our model. Consider the correlated errors between some group of qubits $\mathcal{C}_\chi$. The most general way of describing those errors is to treat the qubits in $\mathcal{C}_\chi$ as a single object, i.e., to always consider their measurement outcomes together. In terms of the noise matrix description, this means that the noise matrix on $\mathcal{C}_\chi$ is some generic $\mathsf{T}_{\mathcal{C}_\chi}$ acting on $\mathcal{C}_\chi$. This gives rise to the first basic object in our model – the *clusters* of qubits. The cluster $\mathcal{C}_\chi$ is a group of qubits with correlations between them so strong, that one can not consider outcomes of their measurements separately. At the same time, it is unlikely that in actual devices the correlations between all the qubits will be so strong that one should assign them all to a single cluster. This motivates the introduction of another, milder possibility. Consider a measurement performed on qubits in cluster $\mathcal{C}_\chi$ and some other qubits $\mathbf{N}_\chi$ (not being in that cluster). It is conceivable to imagine some complicated physical process, which results in the situation in which the noise matrix $\mathsf{T}_{\mathcal{C}_\chi}$ on cluster $\mathcal{C}_\chi$ slightly depends on the state of the qubits in $\mathbf{N}_\chi$. To account for that, we introduce the second basic object of our model – the *neighborhood* of the cluster. The neighborhood $\mathbf{N}_\chi$ of a cluster $\mathcal{C}_\chi$ is a group of qubits the state of which *just before the measurement* affects slightly the noise matrix acting on the cluster $\mathcal{C}_\chi$.

For example, if $\mathcal{C}_\chi$ contains only a single qubit, say Q_0, it is possible that due to some effective ferromagnetic-type interaction, the probability of erroneously detecting state "0" of Q_0 as "1" rises when the neighboring qubit Q_1 is in state "1" (compared to when it is in state '0').

A notion related to our "neighborhood" has appeared in recent literature. Specifically in the context of measurement error characterization, 'spectator qubits' are the qubits that affect measurement noise on other qubits [119; 140]. However, so far this effect was treated rather as an undesired complication, while here it is an inherent element of the proposed noise model.

Now we are ready to provide an efficient noise model. We construct a global

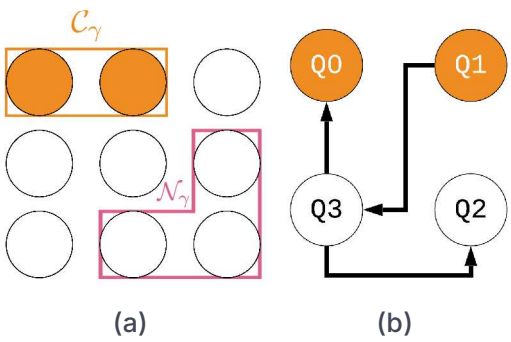


Figure 4.1: *Exemplary correlations in the measurement noise that can be captured by our model. Each circle represents a qubit. Subfigure a) represents a 9-qubit device with some cluster C_χ (orange envelope) consisting of two qubits. The noise on the qubits from that cluster is dependent on the state of qubits from its neighborhood $\mathbf{N}_\chi$ (magenta envelope). Note that the "neighborhood" does not have to correspond to spatial arrangement of qubits in the device. Subfigure b) shows a more detailed example of a four-qubit device. Qubits $Q0$ and $Q1$ are in one cluster, which is indicated by coloring them with the same color. Qubits $Q2$ and $Q3$ are white, meaning they do not belong to any cluster. Qubits at the beginning of the arrow are the neighbors of the qubits at the end of the arrow. Explicitly, the clusters in the example are $C_1 = \{0, 1\}$, $C_2 = \{2\}$, and $C_3 = \{3\}$, while their neighborhoods are $\mathbf{N}_1 = \{3\}$, $\mathbf{N}_2 = \{3\}$, and $\mathbf{N}_3 = \{1\}$. Correlations in the readout errors for qubits Q_0 and Q_1 can be arbitrary, while for the rest of the qubits the dependencies are restricted by the structure of clusters and the neighbors. In particular, the noise matrix on Q_3 depends on the state of Q_1, while the state of Q_3 affects the noise on qubits Q_0 and Q_2. At the same time, qubit 2 does not affect the noise matrix on any other qubit. See the description in the main text.*

noise matrix T with matrix elements of the following form

$$\mathsf{T}_{X_1 \ldots X_N | Y_1 \ldots Y_N} = \prod_\chi \mathsf{T}^{\mathbf{Y}_{\mathbf{N}_\chi}}_{\mathbf{X}_{C_\chi} | \mathbf{Y}_{C_\chi}} \,. \tag{4.3}$$

In the next subsection we give some illustrative examples, but first let us thoroughly describe the notation used in the above equation. A collection $\{C_\chi\}_\chi$ gives us the partitioning of the set of all qubits. Explicitly, $C_\chi \cap C_{\chi'} = \varnothing$ if $\chi \neq \chi'$ and $\cup_\chi C_\chi = [N]$, where N is the number of qubits. To each cluster C_χ the model associate its *neighborhood* $\mathbf{N}_\chi$. Equation (4.3) can be now understood in the following way. The noise matrix $\mathsf{T}^{\mathbf{Y}_{\mathbf{N}_\chi}}$ describing the measurement noise occurring in cluster C_χ depends on the state *just before measurement* of the qubits in the neighborhood $\mathbf{N}_\chi$ of that cluster (hence the superscript $\mathbf{Y}_{\mathbf{N}_\chi}$ denoting that state). Importantly, each $\mathsf{T}^{\mathbf{Y}_{\mathbf{N}_\chi}}$ is left-stochastic for any state of the neighbors. By $\mathbf{X}_{C_\chi}$ (or $\mathbf{Y}_{C_\chi}$) we denote bit-strings of qubits belonging to cluster C_χ which were measured (or put inside the device just before measurement). Finally, $\mathbf{Y}_{\mathbf{N}_\chi}$ indicates the bit-string denoting the state just before the measurement of the qubits from the neighborhood $\mathbf{N}_\chi$ of the cluster C_χ (see Fig. 4.1a for illustration). Note that

in general the correlations in measurement errors (expressed by the structure of the clusters and neighborhoods) do not need to be directly correlated with the physical layout of the device. In general T in Eq.(4.3) is specified by only $\approx \sum_\chi (4^{|C_\chi|})2^{|\mathbf{N}_\chi|}$ parameters, where $|C_\chi|$ and $|\mathbf{N}_\chi|$ are sizes of χ'th cluster and its neighbourhood respectively. Therefore this description is *efficient* provided sizes of the clusters and their neighborhoods are bounded by a constant.

4.3.3 Illustrative examples

In what follows present examples of readout correlation structures that can be described with our model. It is instructive to start with a simple example of a hypothetical four-qubit device depicted in Fig. 4.1b. Note that in this example we have only one non-trivial (i.e., with size ≥ 2) cluster. Let us write explicitly the matrix elements of the global noise matrix acting on that exemplary 4-qubit device

$$\mathsf{T}_{X_0X_1X_2X_3|Y_0Y_1Y_2Y_3} = \mathsf{T}^{Y_3}_{X_0X_1|Y_0Y_1}\mathsf{T}^{Y_3}_{X_2|Y_2}\mathsf{T}^{Y_1}_{X_3|Y_3}\,. \tag{4.4}$$

Note that on the RHS of Eq. (4.4), the superscript Y_3 appears two times, indicating that noise matrices on the cluster C_1 and on the cluster C_2 both depend on the state just before measurement of the qubit 3. At the same time, there is no superscript Y_2, which follows from the fact that qubit 2 does not affect the noise on any other qubits. Note that while a generic noise matrix on 4 qubits would require reconstruction of 16×16 matrix, here we need a number of smaller dimensional matrices to fully describe the noise.

We now move to a more general readout error model, which is particularly inspired by current superconducting qubits implementations of quantum computing devices. Consider a collection of qubits arranged on a device with a limited connectivity. This can be schematically represented as a graph $\mathcal{G}(V,E)$, where each vertex in V represents a qubit and each edge in E connects two qubits that can interact in the device. In such a scenario, a natural first step beyond an uncorrelated readout noise model can be a nearest-neighbour correlated model, where readout errors on each qubit are assumed to be influenced at most by the state of the neighbouring ones. By using the notation introduced in the previous section, we can represent such a model by associating a single-qubit cluster to each vertex, $C_i = \{i\}$, for $i \in V$, and defining the neighbourhoods according to the graph structure, namely $\mathcal{N}_i = \{j|(i,j) \in E\}$. The global noise matrix then reads

$$\mathsf{T}_{X_1\ldots X_{|V|}|Y_1\ldots Y_{|V|}} = \prod_{i\in V}\mathsf{T}^{\mathbf{Y}_{\mathbf{N}_i}}_{\mathbf{X}_i|\mathbf{Y}_i}\,. \tag{4.5}$$

If we specialise this to the case of a $2D$ rectangular lattice of size L, the neighbourhood of generic (i.e. not belonging to the boundary) vertex becomes $\mathcal{N}_i = \{i+1, i-1, i+L, i-L\}$. It follows that each $\mathsf{T}^{\mathbf{Y}_{\mathbf{N}_i}}_{\mathbf{X}_i|\mathbf{Y}_i}$ can be represented by a collection of $2^4 = 16$ matrices of size 2×2, which is an exponential improvement with respect to a general $2^{L^2}\times 2^{L^2}$.

Although the above correlated noise model seems a very natural one, we will see in the following Section that it does not encompass all the correlated readout errors in current superconducting devices, for which it will be more convenient to resort to models (4.3) with more general cluster and neighborhood structures that do not necessarily correspond to the physical layout of the devices.

4.4 Efficient characterization of readout noise

Here we outline a strategy to determine a noise matrix in the form (4.3) which closely represents the readout noise of a given device. We proceed in two steps: at first we infer the structure of clusters ($\{C_\chi\}$ and neighbourhoods ($\{\mathcal{N}_\chi\}$ by making use of Diagonal Detector Tomography (DDT); then we proceed to experimentally determine noise matrices $\{\mathrm{T}^{\mathbf{Y}_{\mathbf{N}_\chi}}_{\mathbf{X}_{C_\chi}|\mathbf{Y}_{C_\chi}}\}$

For the first step, we propose to reconstruct all two-qubit noise matrices (averaged over all other qubits) by means of DDT and calculate the following quantities

$$c_{j\to i} = \frac{1}{2}||\mathrm{T}^{Y_j='0'}_{Q_i} - \mathrm{T}^{Y_j='1'}_{Q_i}||_{1\to 1} \tag{4.6}$$

where $||A||_{1\to 1} := \sup_{||v||_1=1} ||Av||_1 = \max_j \sum_i |A_{ij}|$. The above quantity has an operational interpretation in terms of Total-Variation Distance (TVD). Recall from Section 2.4 that this distance quantifies statistical distinguishability of probability distributions $\mathbf{p}$ and $\mathbf{q}$ and can be defined by (Eq. (2.7))

$$\mathrm{TVD}\,(\mathbf{p},\mathbf{q}) = \frac{1}{2}||\mathbf{p}-\mathbf{q}||_1 = \frac{1}{2}\sum_i |p_i - q_i|\ . \tag{4.7}$$

We can give the following, intuitive interpretation of the quantity from Eq. (4.6): $c_{j\to i}$ represents the maximal TVD for which the output probability distributions on qubit Q_i differs due to the impact of the state of the qubit Q_j on the readout noise on Q_i. Note that in general $c_{j\to i} \neq c_{i\to j}$, which encapsulates the asymmetry in the correlations which is built into our noise model.

We propose to use the values of $c_{j\to i}$ to decide whether the qubits should belong to the same clusters, to the neighborhoods, or should be considered uncorrelated (a simple, intuitive algorithm for this procedure is presented in Appendix B.2.2, Algorithm 3 – in the future, we intend to extend those methods). After doing so, the noise matrices $\{\mathrm{T}^{\mathbf{Y}_{\mathbf{N}_\chi}}_{\mathbf{X}_{C_\chi}|\mathbf{Y}_{C_\chi}}\}$ can be reconstructed by means of joint DDT over the sets of qubits $\{C_\chi \cup \mathcal{N}_\chi\}_\chi$.

In the above construction we assumed that the joint size of a cluster and its neighborhood is at most k. This makes it so that one has to gather DDT data on subsystems of fixed size, implying a number of different circuits that scales at most as $O(N^k)$. However, for any characterization procedure, it is expedient to utilize as few resources as possible. In order to reduce the number of circuits even further, in the next Section we generalize the recently introduced Quantum

Overlapping Tomography (QOT) [151] (see also recent followups [161; 162]) to the context of Diagonal Detector Tomography. We will refer to our method as Diagonal Detector Overlapping Tomography (DDOT).

4.4.1 Diagonal Detector Overlapping Tomography

Quantum Overlapping Tomography is a technique that was introduced for a problem of efficient estimation of all k-particle marginal quantum states. The main result of Ref. [151] was to use the concept of hashing functions [163; 164; 165; 166] to reduce the number of circuits needed to reconstruct all k-qubit marginal states. Specifically, it was shown there that $O\left(\log\left(N\right) k e^k\right)$ circuits suffice for this purpose. Here we propose to use an analogous technique to estimate all noise matrices corresponding to k-particle subsets of qubits. Specifically, we propose to construct a collection of circuits consisting of certain combinations of $\mathbb{I}$ and X gates in order to initialize qubits in states $|0\rangle$ or $|1\rangle$. With fixed k, the collection of quantum circuits for DDOT must have the following property – for each subset of qubits of size k, all of the computational-basis states on that subset must appear at least once in the whole collection of circuits. Intuitively, if a collection has this property, then the implementation of all circuits in the collection allows us to perform tomographic reconstruction (via standard DDT) of noise matrices on all k-qubit subsets. One can think about DDOT as a method of parallelizing multiple local DDTs in order to minimize number of circuits needed to obtain description of all local k-qubit noise processes. In Appendix B.2.4 we show that it suffices to implement $O\left(k2^k \log\left(N\right)\right)$ quantum circuits consisting of random combinations of X and identity gates in order to construct a DDOT circuits collection that allows to capture all k-qubit correlations in readout errors (see Algorithm 1 and Algorithm 2). It is an exponential improvement over standard technique of performing local DDTs separately (which, as mentioned above, requires $O\left(N^k\right)$ circuits). For example, if one chooses $k = 5$ for $N = 15$-qubit device, the naive estimation of all 5-qubit marginals would require the implementation of $2^5\binom{15}{5} \approx 10^5$ quantum circuits, while DDOT allows doing so using ≈ 350 circuits. We note that this efficiency, however, comes with a price. Namely, since different circuits are sampled with different frequencies, some false-positive correlations might appear. This may cause some correlations in the reconstructed noise model to be overestimated (see Appendix B.2.6 for a detailed explanation of this effect). This effect can be mitigated either by certain post-processing of experimental results (see Appendix B.2.6), or by constructing DDOT collections that sample each term the same number of times. Using probabilistic arguments in Appendix B.2.3 we show that still the number of circuits exponential in k and logarithmic in N suffices if we want to have all k particle subsets sampled with *approximately* equal frequency.

4.4.2 Experimental noise reconstruction

We implemented the procedure described in previous subsections with $k = 5$ for IBM's 15q *Melbourne* device and 23-qubit subset of Rigetti's *Aspen-8* device.

The obtained correlation models are depicted in Fig. 4.2. In the case of Rigetti's device, our procedure reports a very complicated structure of multiple correlations in readout noise, while in the case of IBM's device the correlations are fairly simple. We discuss this issue in detail in further sections while presenting results of noise-mitigation benchmarks. Here we conclude by making an observation that, despite common intuition, the structure of the correlations in the readout noise can not be directly inferred from the physical layout of the device.

4.5 Noise mitigation on marginals

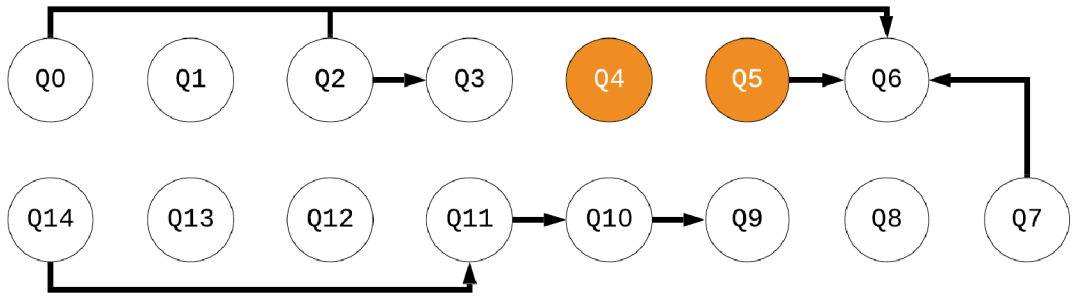
So far, we have described how to effectively characterize readout noise in a quantum device. However, to make this knowledge useful in practice, we need to first understand what are the effects of that noise on quantities we wish to measure in experiments. To this aim, in this Section, we will study how the correlated readout noise affects marginal probability distributions. This has applications, for example, in estimation of expected values of local Hamiltonians. Then we will explain how the reconstructed noise model can be used to reduce the errors on the estimators of marginal probability distributions, and comment on sample complexity of such error-mitigation strategy. The more technical aspects of the discussion can be found in Appendix B.1.

4.5.1 Noise on marginals – overview

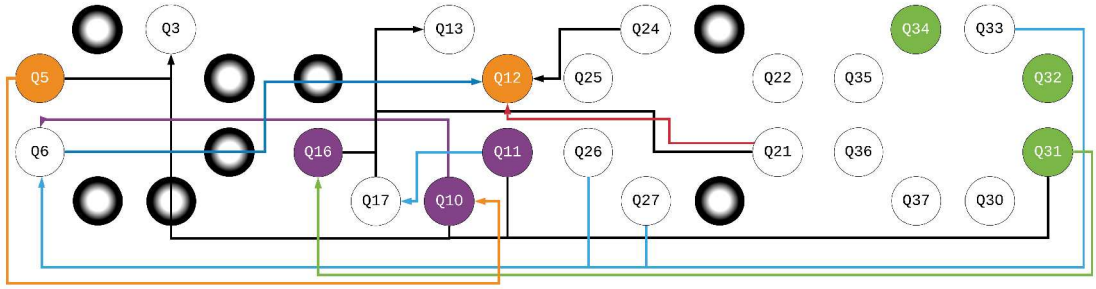
Let us denote by $\mathbf{p}^{noisy}$ a global probability distribution generated by measurement of arbitrary quantum state on the noisy detector for which Eq. (4.3) holds. As mentioned previously, for many interesting problems, such as QAOA or VQE algorithms, one is interested not in the estimation of $\mathbf{p}$ itself (which is an exponentially hard task), but instead in the estimation of multiple marginal probability distributions obtained from $\mathbf{p}$. Let us say that we are interested in the marginal on a subset $\mathcal{S}$ formed by clusters $\mathcal{C}_\chi$ indexed by a set $\mathcal{A}$, $\mathcal{S} := \cup_{\chi\in\mathcal{A}}\mathcal{C}_\chi$, where each $\mathcal{C}_\chi$ is some cluster of qubits (see green envelope on Fig. 4.3 for illustration). Our goal is to perform error mitigation on $\mathcal{S}$. To achieve this, we need to understand how our model of noise affects marginal distribution on $\mathcal{S}$.

From the definition of the noise model in Eq. (4.3) we get that the marginal probability distribution $\mathbf{p}_{\mathcal{S}}^{noisy}$ on $\mathcal{S}$, is a function of the local noise matrices acting on the qubits from $\mathcal{S}$ and the "joint neighborhood" of $\mathcal{S}$, $\mathcal{N}(\mathcal{S}) := \cup_{\chi\in\mathcal{A}}\mathbf{N}_\chi \setminus \mathcal{S}$ (the set $\mathcal{N}(\mathcal{S})$ consists of qubits which are neighbors of points from $\mathcal{S}$ but are not in $\mathcal{S}$) – we derive the exact form of that noise in Appendix B.1.1. Because of this, one can not simply use the standard mitigation strategy: i.e., estimate $\mathbf{p}_{\mathcal{S}}^{noisy}$ and reconstruct probability distribution $\mathbf{p}_{\mathcal{S}}^{ideal}$ by applying the inverse of T.

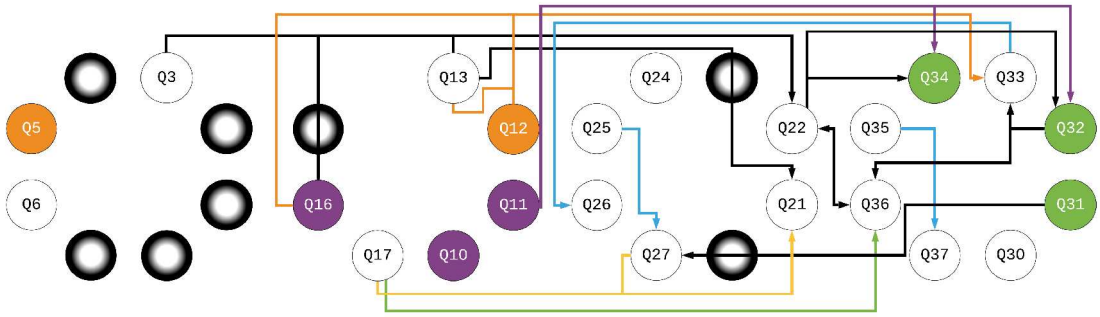
To circumvent the above problem we propose to use the following natural ansatz for the construction of an *approximate* effective noise model on the marginal



(a) *IBM's Melbourne device, 15 qubits.*



(b) *Rigetti's Aspen-8 device. Arrows indicate qubits which affect the measurement noise on the left half of the device.*



(c) *Rigetti's Aspen-8 device. Arrows indicate qubits which affect the measurement noise on the right half of the device.*

Figure 4.2: *Depiction of the correlation model obtained with Diagonal Detector Overlapping Tomography on a) IBM's 15-qubit Melbourne device and b,c) a 23-qubit subset of Rigetti's Aspen-8 device. Due to the complicated structure of Rigetti's correlations, for clarity we divided the plots into two parts which show correlations on the left and right halves of the device (the merged plot can be found in the Appendix B.5.2). The 8 black-and-white circles without label represent qubits which were not included in the characterization due to poor fidelity of the single-qubit gates (below 98%). The meaning of the rest of the symbols is described in the caption of Fig. 4.1b. The colors of the lines connecting the neighbors on b,c) are provided such that the crossings of the lines are unambiguous (and have no other meaning otherwise). For the layout of the graphs, we used the qubits actual connectivity in the devices (i.e., it is possible to physically implement two-qubit entangling gates on all nearest-neighbors in the graph). For IBM's device, we included the qubits in the cluster if the correlations given by Eq. (4.6) were higher than 4% in any direction, while we marked qubits as neighbors if the correlations were higher than 1%. In the case of Rigetti, the respective thresholds were chosen to be 6% and 2%. Moreover, for Rigetti we imposed locality constraints by forcing the joint size of the cluster and the neighborhood to be at most 5 by disregarding the smallest correlations. In practice the correlations within clusters were significantly higher than the chosen thresholds – heatmaps of all correlations can be found in Appendix B.5.2.*

$\mathcal{S}$

$$\mathsf{T}^{\mathcal{S}}_{av} := \frac{1}{2^{|\mathcal{N}(\mathcal{S})|}} \sum_{\mathbf{Y}_{\mathcal{N}(\mathcal{S})}} \mathsf{T}^{\mathbf{Y}_{\mathcal{N}(\mathcal{S})}} , \tag{4.8}$$

where summation is over states of qubits in the joint neighborhood $\mathcal{N}(\mathcal{S})$ defined above. In other words, it is a noise matrix averaged over all states of the neighbors of the clusters in $\mathcal{S}$, *excluding* potential neighbors which themselves belong to the clusters in $\mathcal{S}$. Indeed, note that it might happen that a qubit from one cluster is a neighbor of a qubit from another cluster – in that case, one does not average over it but includes it in a noise model. Importantly, the average matrices $\mathsf{T}^{\mathcal{S}}_{av}$ can be calculated explicitly using data obtained in the characterization of the readout noise.

4.5.2 Approximate noise mitigation

The noise matrix $\mathsf{T}^{\mathcal{S}}_{av}$ (Eq. (4.8)) can be used to construct the corresponding *effective correction matrix*

$$\mathsf{A}^{\mathcal{S}}_{av} := \left(\mathsf{T}^{\mathcal{S}}_{av}\right)^{-1} . \tag{4.9}$$

Correcting the marginal distribution via left-multiplication by the above ansatz matrix is not perfect and can introduce error in the mitigation. In the following Proposition 1 we provide an upper bound on that error measured in TV distance.

Proposition 1. *Let $\mathbf{p}^{noisy}$ be a probability distribution on N qubits obtained from the N qubit probability distribution $\mathbf{p}^{ideal}$ via stochastic transformation T of the form given in Eq. (4.3). Consider the subset of qubits $\mathcal{S} = \cup_{\chi\in\mathcal{A}} C_\chi$. Let $\mathbf{p}^{corr}_{\mathcal{S}} = \mathsf{A}^{\mathcal{S}}_{av}\mathbf{p}^{noisy}_{\mathcal{S}}$ be the result of the application to the marginal distribution $\mathbf{p}^{noisy}_{\mathcal{S}}$ of the correction procedure using the effective correction matrix $\mathsf{A}^{\mathcal{S}}_{av}$ from Eq. (4.9). We then have the following inequality*

$$TVD\left(\mathbf{p}^{corr}_{\mathcal{S}}, \mathbf{p}^{ideal}_{\mathcal{S}}\right) \leq \tag{4.10}$$

$$\leq \frac{1}{2}||\mathsf{A}^{\mathcal{S}}_{av}||_{1\to 1} \max_{\mathbf{Y}_{\mathcal{N}(\mathcal{S})}} ||\mathsf{T}^{\mathcal{S}}_{av} - \mathsf{T}^{\mathbf{Y}_{\mathcal{N}(\mathcal{S})}}||_{1\to 1} , \tag{4.11}$$

where the maximization goes over all possible states of the neighbors of $\mathcal{S}$.

The proof of the above Proposition is given in Appendix B.1.2 – it uses the convexity of the set of stochastic matrices, together with standard properties of matrix norms and with a triangle inequality (similar methods were used for providing error bounds on mitigated statistics in Ref. [42]). Note that the quantity on RHS of Eq. (4.10) shows resemblance to $c_{i\to j}$ in Eq. (4.6) (which we used to quantify correlations). Hence $\frac{1}{2}\max_{\mathbf{Y}_{\mathbf{N}(\mathcal{S})}} ||\Lambda^{\mathcal{S}}_{av} - \Lambda^{\mathbf{Y}_{\mathbf{N}(\mathcal{S})}}||_{1\to 1}$ can be interpreted as the maximal TVD between states on $\mathcal{S}$ generated by $\Lambda^{\mathcal{S}}_{av}$ and states generated

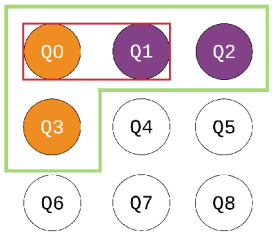


Figure 4.3: *Illustration of the cluster structure of an exemplary 9-qubit device. There are two non-trivial clusters present ($\{0,3\}$ and $\{1,2\}$). For clarity, no neighborhood dependencies are shown, though in reality noise on the clusters can be dependent on the qubits outside the clusters. When one measures all the qubits but the goal is to estimate four-qubit marginal distribution on $\mathcal{S} = \{0,3\} \cup \{1,2\} = \{0,1,2,3\}$ (green envelope), noise-mitigation should be performed based on the noise model for the whole set of qubits $\mathcal{S}$. On the other hand, when the goal is to estimate the two-qubit marginal on qubits $\mathcal{S}_k = \{0,1\}$ (red envelope), it is still preferable to first perform error-mitigation on the four-qubit marginal on qubits $\mathcal{S} = \{0,1,2,3\}$, and then take marginals over qubits 2 and 3 to obtain corrected marginal on $\mathcal{S}_k$. See the description in the text.*

by $\Lambda^{\mathbf{Y}_{\mathbf{N}_x}}$ (which appear in the description of the noise model). This can be also interpreted as a measure of dependence of noise between qubits in $\mathcal{S}$ and the state of their neighbors just before measurement. Indeed, if the true noise does not depend on the state of the neighbors, the RHS of inequality Eq. (4.10) yields 0, and it grows when the noise matrices $\left\{\Lambda^{\mathbf{Y}_{\mathbf{N}(x)}}\right\}$ increasingly differ.

In practice, it might happen that one is interested in the marginal distribution on the qubits from some subset $\mathcal{S}_k \subset \mathcal{S}$ (red envelope in Fig. 4.3). In principle, one could then consider a coarse-graining of noise-model *within* $\mathcal{S}$ (i.e., construction of noise model averaged over qubits from $\mathcal{S}$ that do not belong to $\mathcal{S}_k$, treating those qubits like neighbors) and perform error-mitigation on the coarse-grained subset $\mathcal{S}_k$. However, due to the high level of correlations within clusters, we expect such a strategy to work worse than performing error-mitigation on $\mathcal{S}$, and then taking marginal to $\mathcal{S}_k$. Indeed, we observed numerous times that the latter strategy works better in practice. Yet, it is also more costly (since, by definition, $\mathcal{S}$ is bigger than $\mathcal{S}_k$), hence in actual implementations with restricted resources (e.g., a limited number of available circuit executions) one may also consider implementing a coarse-grained strategy. In the following sections, we will focus on error-mitigation on the set $\mathcal{S}$. All those considerations can be easily generalized to the case of $\mathcal{S}_k \subset \mathcal{S}$.

4.5.3 Sample complexity of error-mitigation

Let us now briefly comment on the sampling complexity of this error-mitigation scheme (the detailed discussion is postponed to Section 4.7). If one is interested in estimating an expected value of local Hamiltonian, a standard strategy is to estimate the local marginals and calculate the expected values of local Hamiltonian terms on those marginals. Without any error-mitigation, this has

exponential sampling complexity in locality of marginals (which for local Hamiltonians is small), and logarithmic complexity in the number of local terms (hence, for typically considered Hamiltonians, also logarithmic in the number of qubits) – see Eq. (4.17) and its derivation in Appendix B.1.3. Now, if one adds to this picture error-mitigation *on marginals*, this, under reasonable assumptions, does not significantly change the scaling of the sampling complexity. We identify here two sources of sampling complexity increase (as compared to the non-mitigated marginal estimation). First, the noise mitigation via inverse of noise matrix does propagate statistical deviations – the bound on this quantitatively depends on the norm of the correction matrix (see Ref. [42] and detailed discussion around Eq. (4.17) in Section 4.7). Assuming that the local noise matrices are not singular (which is anyway required for error-mitigation to work), this increases sampling complexity by a constant factor (in particular, for a given marginal, it is proportional to the norm of the correction matrix acting on that marginal, see Appendix B.1.2 for details). Second, the additional errors can come from the fact that, as described above, sometimes it is desirable to perform noise-mitigation on higher-dimensional marginals (if some qubits are highly correlated). However, assuming that readout noise has bounded locality, this can increase a sampling complexity only by a constant factor (this factor is proportional to the increase of the marginal size as compared to estimation without error-mitigation). In both cases, for a fixed size of marginals (as is the case for local Hamiltonians), it does not change the *scaling* of the sampling complexity with the number of qubits, which remains logarithmic.

4.6 Experimental benchmark of the noise model

4.6.1 Energy estimation of local Hamiltonians

After having characterized the noise model, how to assess whether it is accurate? To answer this question we propose the following, application-driven heuristic benchmark. The main idea is to test whether the error-mitigation of local marginals based on the adopted noise model is accurate. To check this we propose to to consider the problem of estimation of the expectation value $\langle \mathcal{H} \rangle$ of a local classical Hamiltonian

$$\mathcal{H} = \sum_{\alpha} H_{\alpha} \tag{4.12}$$

measured on its ground state $|\psi_0(\mathcal{H})\rangle$. Here by "local" we mean that the maximal number of qubits on which each H_α acts non-trivially does not scale with the system size. Classicality of the Hamiltonian means that every $\mathcal{H}_\alpha$ is a linear combination of products of σ_z Pauli matrices. In turn the ground state $|\psi_0(\mathcal{H})\rangle$ can be chosen as classical i.e.,

$$|\psi_0(\mathcal{H})\rangle = |\mathbf{X}(\mathcal{H})\rangle \ , \tag{4.13}$$

for some bit-string $\mathbf{X}(\mathcal{H})$ representing one of the states from the computational basis. This problem is a natural candidate for error-mitigation benchmark due

to at least three reasons. First, a variety of interesting optimization problems can be mapped to Ising-type Hamiltonians from Eq. (4.12). Indeed, this is the type of Hamiltonians appearing in the Quantum Approximate Optimization Algorithm. The goal of the QAOA is to get as close as possible to the ground state $|\psi_0(\mathcal{H})\rangle$. Second, the estimation of $\langle\mathcal{H}\rangle$ can be solved by the estimation of energy of local terms $\langle H_\alpha\rangle$ and therefore error-mitigation on marginals can be efficiently applied. Finally, the preparation of the classical ground state $|\psi_0\rangle$, once it is known, is very easy and requires only the application of local σ_x (NOT) gates. This works in our favor because we want to extract the effects of the readout noise, and single-qubit gates are usually of high quality in existing devices.

To perform the benchmark we propose to implement quantum circuits preparing ground states of many different local classical Hamiltonians, measure them on the noisy device, and perform two estimations of the energy - first from the raw data, and second with error-mitigation based on our characterization. Naturally, it is also desirable to compare both with the error-mitigation based on a completely uncorrelated noise model (cf. Eq. (4.2)). We propose that if the mitigation based on a particular noise model works well on average (over some number of Hamiltonians), one can infer that the model is more accurate as well.

4.6.2 Experimental readout noise mitigation

We applied the benchmark strategy described above on the 15-qubit IBM's *Melbourne* device and 23-qubit subset of Rigetti's *Aspen-8* device. We implemented two classes of random Hamiltonians – MAX-2-SAT instances (600 on IBM and 399 on Rigetti) and fully-connected graphs with random interactions and local fields (600 on IBM).

The MAX-2-SAT problem is the simplest variant of the maximal satisfability problem (MAX-k-SAT) [167]. In those problems, we are given some Boolean formula, typically represented in the standard conjuctive normal form (CNF), that is a set of logical AND statements between multiple clauses. Here, a clause means an OR statement between k binary variables – in the case of MAX-2-SAT, each clause is an OR statement between 2 variables. The task is to find the maximum number of clauses that can be assigned the "true" value when the assignment of binary values is made to the variables. The hardness of the problem is controlled by the clause density parameter, i.e., the (average) number of clauses per variable.

In our experiments, for MAX-2SAT we chose the clause density 4, and the corresponding instances were generated by randomly choosing $4 * N$ clauses with N variables (N being the number of qubits), and mapping the resulting CNF to an Ising Hamiltonian via the standard mapping (see, e.g., Refs. [168; 169]). For fully connected graphs, the random interactions and local fields were chosen uniformly at random from range $[-1, 1]$. Figure 4.4 presents the results of our experiments, together with a comparison with the uncorrelated noise model.

Let us first analyze the results of experiments performed on a 15-qubit IBM's device. Here it is clear that the error-mitigation based on our noise model performs well, often reducing errors in estimation by as much as one order of mag-

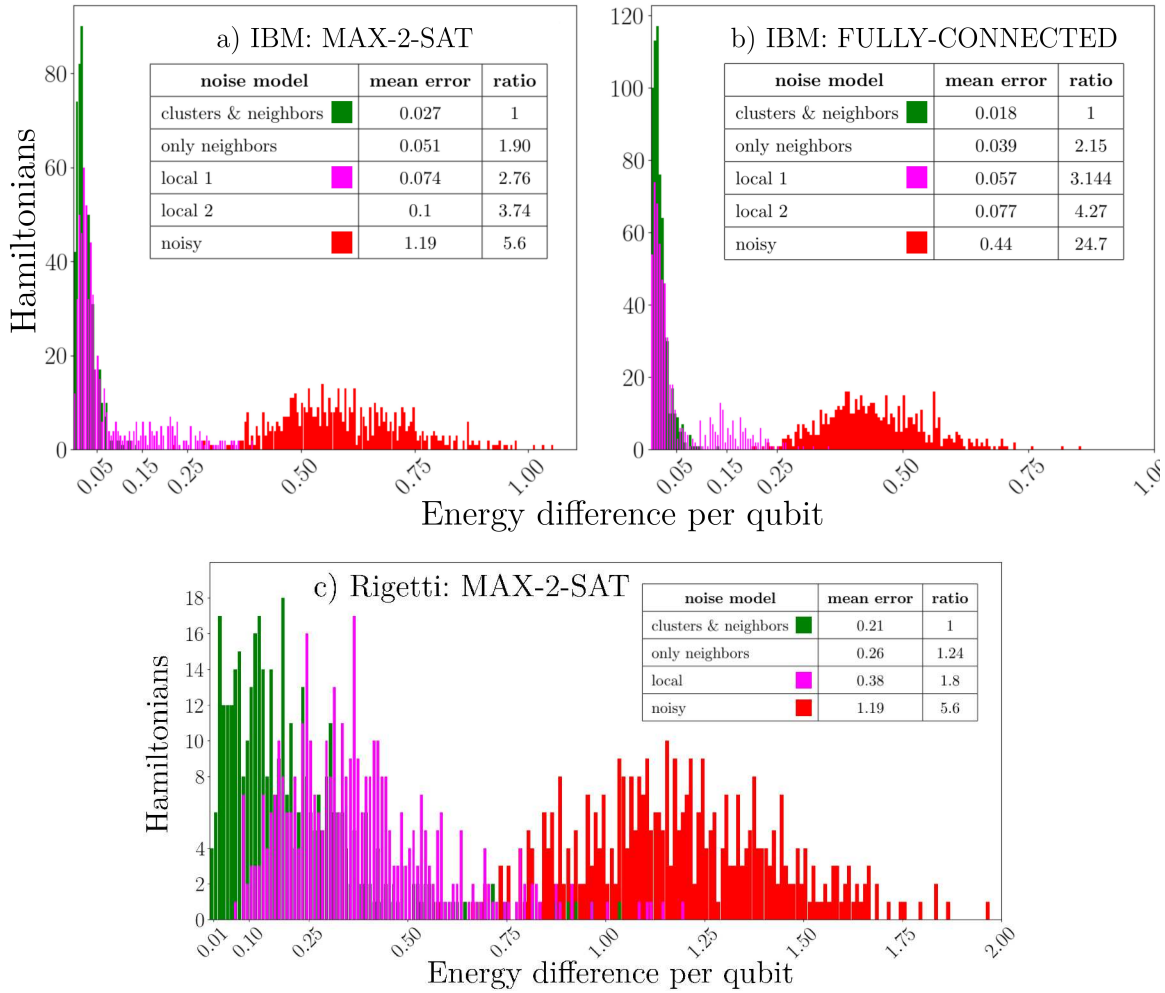


Figure 4.4: *Results of an experimental benchmark of the readout noise mitigation on (a-b) IBM's 15-qubit Melbourne device, and (c) a subset of 23 qubits on Rigetti's Aspen-8 device. Each histogram shows data for 600 (IBM's) or 399 (Rigetti) different random Hamiltonians – (a,c) random MAX-2-SAT and (b) fully-connected graph with random interactions and local fields. The horizontal axis shows the absolute energy difference (between estimated and theoretical) divided by the number of qubits. The histogram comparison is done with no mitigation and with uncorrelated noise model characterization. The embedded tables show average errors depending on the adopted noise model. Here "ratio" refers to ratio of means. Additional second row in each figure shows data for noise model with only trivial (single-qubit) clusters and their neighborhoods. In case of IBM data, the additional fifth row illustrates memory effects. See description in the main text.*

nitude. We further note that the uncorrelated noise model performs quite well (yet being visibly worse than ours). To compare the accuracy by using a single number (as opposed to looking at the whole histogram), we take the ratio of the mean deviations from the ideal energy for the error-mitigated data based on two models. The results are presented in tables embedded into Fig. 4.4. In those tables, we provide also additional experimental data. Namely, for each figure there the second row shows data for noise model labeled as "only neighbors". This corresponds to noise model in which each qubit is a trivial, single-qubit cluster, and correlations are included only via neighborhoods. The worse results of error-mitigation for such model as compared to full clusters-neighborhoods model motivates the introduction of non-trivial clusters. Furthermore, we found experimentally that the characterization of the uncorrelated noise model exhibits significant memory effects (see, for example, Ref. [170]). Particularly, if one performs uncorrelated noise characterization in a standard way, i.e., by performing characterization in a separate job request to a provider, without any other preceding experiments ("local 2" in tables), the accuracy (measured by the error in energy after mitigation based on a given noise model) is much lower than for the characterization with some other experiments performed *prior* to the characterization of the uncorrelated model ("local 1" in tables). Indeed, the difference in mean accuracy can be as big as $\approx 26\%$.

Clearly, the overall performance of Riggeti's 23 qubit device is lower than that for IBM's device. First, the effects of noise (measured in energy error per qubit) are stronger. Second, the mean error with error-mitigation is only around ≈ 5.6 times smaller than the error without error-mitigation (as opposed to factor over 22 for IBM's device). Third, the comparison to the uncorrelated noise model shows that the uncorrelated model performs not much worse than the correlated one.

Here we provide possible explanations of this poorer quality of experiments performed on Rigetti's device. Due to the limited availability of the Rigetti's device, we used a much lower number of samples to estimate Hamiltonian's energies in those experiments. Specifically, each energy estimator for Rigetti's experiments was calculated using only 1000 samples, while for IBM's experiments the number of samples was 40960. This should lead to statistical errors higher by a factor of roughly $\sqrt{41} \approx 6.4$ (and note that the errors in error-mitigated energy estimation in Rigetti's device are around 7.8 times higher than corresponding errors for IBM's device for the same class of Hamiltonians). Similarly, we used fewer measurements to perform DDOT – on Rigetti's device, we implemented 504 DDOT circuits sampled 1000 each, while on IBM's device we performed 749 circuits sampled 8192 each. Less DDOT circuits imply less balanced collection, hence, as already mentioned, some correlations might have been overestimated. In summary, our characterization of this device was in general less accurate than on IBM's device. This might be further amplified by the fact that single-qubit gates (which are used to implement DDOT circuits) were of lower quality for Rigetti's device. Finally, as illustrated in Fig. 4.2, we observed that correlations in measurement noise for Rigetti's device are much more complex than in the case of IBM's. As

mentioned in the Figure's description, to work around this we imposed locality constraints in the constructed noise model by disregarding the lowest correlations between qubits, which made the model less accurate.

We note that due to the limited availability of Rigetti's device, we did not perform the study of memory effects similar to that performed in IBM. The local noise model presented for this device originates from a separate uncorrelated noise characterization performed prior to the rest of the experiments (hence it is analogous to the "local 2" model in IBM's case).

To summarize, presented results suggest that in experiments on near-term quantum devices it will be indispensable to account for cross-talk effects in measurement noise. For both studied quantum devices we provided proof-of-principle experiments showing significant improvements in ground state energy estimation on the systems of sizes in the NISQ regime. Motivated by those results, we hope that the framework developed in this work will prove useful in the future, more complex experiments on even larger systems.

4.7 Error analysis for QAOA

In this section, we analyze the magnitude of errors resulting from our noise-mitigation scheme when applied to an energy estimation problem. Those errors result from two sources. First, from the fact that we use approximate correction matrices instead of the exact ones (see in Proposition 1). Secondly, by statistical errors due to the common practice of measuring *multiple* marginals *simultaneously* in a single run of the experiment. In the following, we will analyze both cases separately and then provide a bound that takes them both into account. We restrict our analysis to local Hamiltonians diagonal in the computational basis. A detailed derivation of the results below can be found in Appendices.

4.7.1 QAOA overview

Before starting, let us provide a short overview of the QAOA algorithm. In standard implementation [144], one initializes quantum system to be in $|+\rangle^{\otimes N}$ state, where $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. Then p-layer QAOA is performed via implementation of unitaries of the form

$$U_p(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_j^p U_{\alpha_j} U_{\beta_j}, \tag{4.14}$$

where $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ are the angles to-be-optimized. Unitary matrices are given by $U_{\alpha_j} := \exp(-i\,\alpha_j \mathcal{H}_D)$, and $U_{\beta_j} := \exp(-i\,\beta_j \mathcal{H}_O)$, where $\mathcal{H}_D$ is driver Hamiltonian (which we take to be $\mathcal{H}_D = \sum_k^N \sigma_x^k$), and $\mathcal{H}_O$ is objective Hamiltonian that one wishes to optimize (i.e., to find approximation for its ground state energy). The quantum state after p-th layer is $|\psi_p\rangle = U_p |+\rangle^{\otimes N}$ and the function which is passed to classical optimizer is the estimator of the expected value $\langle\psi_p| \mathcal{H}_O |\psi_p\rangle$ (note that this makes those estimators to effectively be a function

of parameters $\{\alpha_j\}, \{\beta_j\}$). The estimator is obtained by sampling from the distribution defined by the measurement of $|\psi_p\rangle$ in the computational basis, taking the relevant marginals, and calculating the expected value of $\mathcal{H}_O$ using values of those estimated marginals. Let us now proceed to the analysis of possible sources of errors while performing noise-mitigation on the level of marginals to estimate the energy of local Hamiltonians, such as those present in QAOA.

4.7.2 Approximation errors

We start by recalling that performing noise mitigation with the average noise matrix instead of the exact one subjects the estimation of each marginal to an error upper bounded by Eq. (4.10). It follows that the correction of multiple marginal distributions can lead to the accumulation of errors which for each marginal α (we label subset of qubits by α so that local term H_α acts non-trivially on qubits from α) take the form

$$\delta^\alpha := \frac{1}{2}||C_{av}^{\mathcal{S}_\alpha}||_{1\to 1} \max_{\mathbf{Y}_{\mathcal{N}(\mathcal{S}_\alpha)}} ||\mathrm{T}_{av}^{\mathcal{S}_\alpha} - \mathrm{T}^{\mathbf{Y}_{\mathcal{N}(\mathcal{S}_\alpha)}}||_{1\to 1} \tag{4.15}$$

where set $\mathcal{S}_\alpha = \cup_{\gamma\in\mathcal{A}} C_\gamma$, $\mathcal{A} = \{\chi \mid C_\chi \cap \alpha \neq \emptyset\}$, consists of clusters to which qubits from α belong, and $C_{av}^{\mathcal{S}_\alpha}$ is the average correction matrix for the marginal on that set. It is straightforward to show that the total possible deviation between the error-mitigated expected value $\langle\mathcal{H}^{\mathrm{corr}}\rangle$ and the noiseless one $\langle\mathcal{H}\rangle$ is upper bounded by

$$|\left\langle\mathcal{H}^{\mathrm{corr}}\right\rangle - \langle\mathcal{H}\rangle| \leq 2\sum_\alpha \delta^\alpha ||H_\alpha|| \quad \text{(additive approximation bound)} \ . \tag{4.16}$$

4.7.3 Additive statistical bound

Moving to the effect of measuring several marginals simultaneously, let us start by considering the simplest bound on the propagation of statistical deviations under our error-mitigation. In Appendix B.1 we derive that the Total-Variation Distance (TVD, Eq. (4.7)) between the estimator $\mathbf{p}_\alpha^{\mathrm{est}}$ and the actual local marginal $\mathbf{p}_\alpha$ is upper bounded by

$$\mathrm{TVD}\left(\mathbf{p}_\alpha^{\mathrm{est}}, \mathbf{p}_\alpha\right) \leq \tag{4.17}$$

$$\leq \epsilon^* := \sqrt{\frac{\log\left(2^{2^N}-2\right) + \log\left(\frac{1}{P_{\mathrm{err}}}\right) + \log\left(K\right)}{2s}}, \tag{4.18}$$

where N is the number of of qubits in the support of each local term (for simplicity we assume it to be the same for all H_α), K is the total number of local terms, s is the number of samples, and $1-P_{\mathrm{err}}$ is the confidence with which the above bound is stated. Importantly, the above bound is satisfied for each marginal *simultaneously*, hence the logarithmic overhead $log\,(K)$. Using Eq. (4.17) together with

standard norm inequalities one obtains the following bound for the total energy estimation

$$|\mathcal{H}^{\text{est}}_{\text{corr}} - \langle \mathcal{H} \rangle| \leq \sum_{\alpha} ||H^{\alpha}||\, ||\mathsf{A}_{\alpha}||_{1\to 1}\, \epsilon^{*}$$
$$(\text{additive statistical bound})\,.$$

Here $\mathcal{H}^{\text{est}}_{\text{corr}}$ denotes the estimator of the total energy with error-mitigation performed on each local term independently and A_{α} is the exact (not approximate) correction matrix on marginal α.

4.7.4 Joint approximation and statistical bound

The two bounds provided above took into account the two considered sources of errors independently. By using the triangle inequality (see Appendix B.1.4), we can now combine them to obtain

$$|\mathcal{H}^{\text{est}}_{\text{corr}} - \langle \mathcal{H} \rangle| \leq \tag{4.19}$$

$$2 \sum_{\alpha} ||H^{\alpha}|| \left(\underbrace{\epsilon^{*} ||\mathsf{A}^{\mathcal{S}_{\alpha}}_{av}||_{1\to 1}}_{\text{statistical errors}} + \underbrace{\delta_{\alpha}}_{\text{approximation errors}} \right). \tag{4.20}$$

It follows that the dominant scaling in the overall error are linear in the number of terms K caused by summing over all of them and the logarithmic overhead in ϵ^{*} added by the statistical errors.

4.7.5 Sampling complexity of energy estimation

While the additive bound from Eq. (4.19) could be tight *in principle*, we observed numerically on many occasions that in practice the statistical errors are much smaller (see Fig. 4.5 for exemplary results that are discussed in Section 4.7.8).

Here we will provide arguments that show that natural estimators of local energy terms H_{α} effectively behave as uncorrelated for a broad class of quantum states, hence leading to a significantly smaller total error than that obtained from an additive bound.

We start by describing in detail the natural strategy for energy estimation in the considered scenarios. In this work we are concerned with classical local Hamiltonians. This means that all local terms H_{α} can be measured simultaneously via a single computational basis measurement. The natural estimation procedure amounts to repeating s independent computational basis measurements on a quantum state ρ of interests. Outcomes of these measurements are then used to obtain local energy estimators $E^{\text{est}}_{\alpha} = \frac{1}{s}\sum_{i=1}^{s} E^{i}_{\alpha}$, where E^{i}_{α} are values of the local energy terms obtained in the i'th experimental run. Now to perform estimation of expected value of energy, $\langle \mathcal{H} \rangle$, we simply sum the local estimators E^{est}_{α}

$$\mathcal{H}^{\text{est}} = \sum_{\alpha} H^{\text{est}}_{\alpha} = \frac{1}{s} \sum_{i=1}^{s} \sum_{\alpha} E^{i}_{\alpha}\,. \tag{4.21}$$

It is clear that $\mathcal{H}^{\text{est}}$ is an unbiased estimator of $\langle \mathcal{H} \rangle$. Likewise E_α^{est} are unbiased estimators of $\langle H_\alpha \rangle$.

We would like to understand statistical properties of $\mathcal{H}^{\text{est}}$ (specifically its variance) as as a function of number of experimental runs (samples) s and the number of local terms in the Hamiltonian K. To this and we observe that random variables E_α^i, E_β^j are independent unless $i = j$ and therefore

$$\text{Var}\left(\mathcal{H}^{\text{est}}\right) = \frac{1}{s} \sum_{\alpha,\beta} \text{Cov}(E_\alpha^i, E_\beta^i) \,, \tag{4.22}$$

Assuming that measurements of the energy E_α^i are distributed according to the probability compatible with the Born rule allows us to write

$$\text{Cov}(E_\alpha^i, E_\beta^i) = \text{Cov}\left(H_\alpha, H_\beta\right) = \left\langle H_\alpha H_\beta \right\rangle - \langle H_\alpha \rangle \left\langle H_\beta \right\rangle \,. \tag{4.23}$$

Consequently we have

$$\text{Var}\left(\mathcal{H}^{\text{est}}\right) = \frac{1}{s}\text{Var}(\mathcal{H}) = \sum_{\alpha,\beta} \text{Cov}(H_\alpha, H_\beta). \tag{4.24}$$

The variance of $\mathcal{H}^{\text{est}}$ can be related to the *sample complexity* of the energy estimation. Let $\Delta E > 0$ be some positive number. Then using Chebyshev inequality we get

$$\text{Prob}\left(|\mathcal{H}^{\text{est}} - \langle \mathcal{H} \rangle| \geq \Delta E\right) \leq \frac{\text{Var}\left(\mathcal{H}^{\text{est}}\right)}{(\Delta E)^2} \,. \tag{4.25}$$

Choosing some parameter P_f as an upper bound bound on the RHS of this inequality, i.e., $\frac{\text{Var}\left(\mathcal{H}^{\text{est}}\right)}{(\Delta E)^2} \leq P_f$. Then we obtain using (4.24) that for the number of samples s satisfying

$$\frac{\text{Var}\left(\mathcal{H}\right)}{(\Delta E)^2 P_f} \leq s \,. \tag{4.26}$$

the estimator $\mathcal{H}^{\text{est}}$ will be within accuracy form the expectation value $\langle \mathcal{H} \rangle$ with probability at least $1 - P_f$. Now, if different Hamiltonian terms $\mathcal{H}_\alpha$ are correlated then according to the above bound the sample complexity grows like K^2, where K is the total number of terms in $\mathcal{H}$. Conversely, if $\text{Cov}(H_\alpha, H_\beta) \approx 0$ (for $\alpha \neq \beta$) then get sample complexity scaling linearly with K. Below, we will provide some arguments for making the approximation $\text{Cov}(H_\alpha, H_\beta) \approx 0$, i.e., that the variables are effectively uncorrelated.

The above consideration can be equivalently translated to the estimates of the confidence intervals associated with estimator $\mathcal{H}^{\text{est}}$ for a fixed value of samples s. Specifically, if local terms are uncorrelated, then the confidence interval (statistical error) will scale as square-root $\sqrt{K}$ of the number of Hamiltonian

terms, contrary to the pessimistic bound in Eq. (4.19) which is linear in K. We now show that one should expect that the sub-linear scaling of energy errors described above holds for a variety of quantum states, which in turn greatly reduces the sampling complexity compared to the pessimistic (linear) bound. We want to emphasize that our results are of immense practical importance for near-term devices. Our findings indicate a reduction (compared to the naive bounds) of sample complexity of energy estimation by orders of magnitude, even for relatively small systems ($K \approx 100$ and larger).

4.7.6 Generic 2-local Hamiltonians in QAOA

We start by considering states that appear at the beginning and at the end of QAOA. In recent work [171] it was shown that after p-th layer of QAOA optimization, for given two local terms H_α and H_β, there is no entanglement between qubits from α and qubits from β if they are further away from each other than $2p$ (on a graph corresponding to interactions present in a Hamiltonian). Therefore, for generic QAOA optimization, one can expect that for low p, i.e., at the beginning of the QAOA, the local Hamiltonian terms will be uncorrelated variables.

The following Proposition provides a more quantitative description of the variance behavior for generic 2-local Hamiltonians corresponding to random graphs.

Proposition 2. *Consider Hamiltonian with connectivity given by Erdös-Rényi random graph [171] in which each edge of the graph is added independently at random with some fixed probability. Assume that the probability of adding edge is chosen so the average degree of a node is equal to $q = \frac{K}{N}$, hence that a random graph has on average N nodes and K edges. For QAOA starting from product state, if the number of layers satisfies*

$$p < \frac{w \log(N)}{8 \log(2q/\ln(2))} - 1 \tag{4.27}$$

with $w < 1$, then with probability $1 - e^{-N^{a/2}}$ the variance of the Hamiltonian is bounded by

$$\mathrm{Var}\,(\mathcal{H}) \leq f_{\mathcal{H}}\; q\; N^{A+1} \tag{4.28}$$

where

$$\begin{aligned} f_{\mathcal{H}} &= \max_{\alpha,\beta} ||H_\alpha||\;||H_\beta||\,, \\ A &= w\,\frac{(2 + |\log_{2q}(\ln(2))|)}{(1 + |\log_{2q}(\ln(2))|)}\,, \\ a &= \frac{w}{3(1 + |\log_{2q}(\ln(2))|}\,, \end{aligned}$$

where maximization in first definition goes over all two-qubit local terms acting on subsets of qubits α and β. Since we can always choose $w < 1$ such

that $A < 1$, the variance thus scales sub-quadratically for shallow depth QAOA. Importantly, the parameter w can be chosen in such a way that A goes asymptotically to 0 (an exemplary choice is $w = O(\log(N)^{-\frac{1}{2}})$, which in turn means that in the large N regime the variance $\mathrm{Var}(\mathcal{H})$ will scale almost linearly with the number of terms.

The proof of the above Proposition 2 uses insights from Ref. [171] and is delegated to Appendix B.4.2.

4.7.7 Random quantum states

A simple argument can be made to show that for generic Haar-random pure sates, as well as random states appearing in random local quantum circuits, the variance of a local Hamiltonian $\mathcal{H}$ behaves as if different energy terms were independent. Let $|\psi\rangle$ be a pure state on $(\mathbb{C}^2)^{\otimes N}$. For a subset of qubits γ, let ρ_γ denote the reduced density matrix $|\psi\rangle\langle\psi|$ on qubits in γ, and let id_γ denote the normalized maximally mixed state on qubits in γ. Assume now that local Hamiltonian terms H_α, H_β have disjoint supports. It can be then shown (see Proposition 4 in Appendices for the proof) that

$$\mathrm{Cov}(H_\alpha, H_\beta) \leq 3\|H_\alpha\|\|H_\beta\| \, \|\rho_{\alpha\cup\beta} - \mathrm{id}_{\alpha\cup\beta}\|_1 \, . \tag{4.29}$$

Now, it is a well-known fact [172] that with overwhelming probability *all* few-body marginals ρ_γ of a Haar-random multiqubit states $|\psi\rangle$ are exponentially close to maximally mixed states. Therefore, assuming that $\max_{\alpha,\beta}$ does not scale with the system size, we have that high probability over the choice of $|\psi\rangle$, for every disjoint terms H_α, H_β in a local Hamiltonian $\mathcal{H}$

$$\mathrm{Cov}(H_\alpha, H_\beta) \approx 0 \, . \tag{4.30}$$

The above reasoning mimics the computation done in Theorem 1 of [173], where it was used to establish that generic Haar-random pure states attain only the so-called standard quantum limit in the paradigmatic interferometric scenarios (again the underlying argument was based on the fact that *all* few-body reduced density matrices of a generic pure state $|\psi\rangle$ are very close to maximally mixed states with overwhelming probability).

Analogous analysis can be carried out for typical states generated by local random quantum circuits. Such circuits are known to form approximate t-designs, i.e., capture properties of typical Haar-random unitaries captured by low-degree moments [174]. Specifically, a recent paper [175] considered evolution of local entropies for pure states $|\psi\rangle$ generated by shallow local random quantum circuits. From Theorem 1 of that work it directly follows that with probability greater than $1-\delta$ over the choice random states $|\psi\rangle$ generated by random local circuits in the brick-wall architecture of depth r, all marginals ρ_γ of size $|\gamma| = k$, satisfy $\|\rho_\gamma - \mathrm{id}_\gamma\|_1 \leq \epsilon$, where

$$\delta \leq \frac{\binom{N}{k}^2}{\epsilon^2}\left(2^{2k-N} + 2^k\left(\frac{4}{5}\right)^{2(r-1)}\right) \, . \tag{4.31}$$

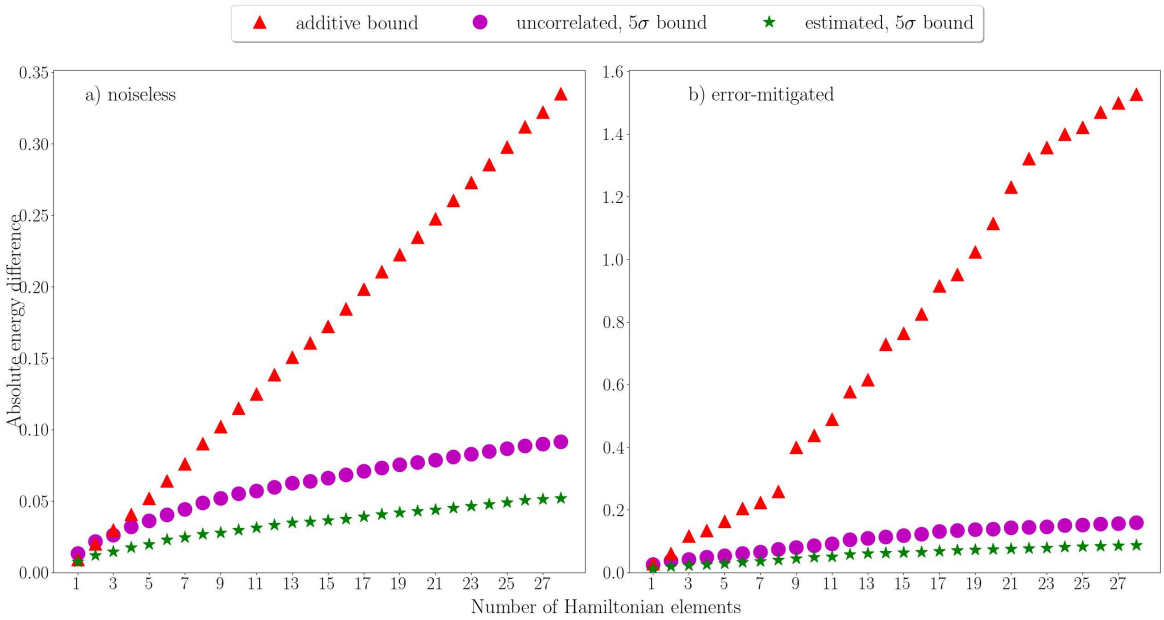


Figure 4.5: *Numerical results illustrating low statistical errors in the estimation of the energy of local Hamiltonians. We show the results for the energy of 8-qubit random Sherrington-Kirkpatrick Hamiltonians, calculated on quantum states obtained* after the third layer *of QAOA optimization. The horizontal axis shows the number of Hamiltonian elements (in a sense that for a given Hamiltonian one estimates only some local terms). The Y-axis shows the absolute difference energy error. Each data point is an average over* 92 *Hamiltonians. Red triangles correspond to additive error bound from Eq.* (4.19) *(with correction matrix norm set to 1 and approximation errors to* 0 *for noiseless scenario). Magenta circles show the confidence intervals on energy estimation corresponding to* 5σ*-confidence (via Chebyshev's bound) assuming* uncorrelated *variables for the sample size* $s \approx 10^4$ *(see Eq.* (4.26) *and description in the text). Green stars are analogous confidence intervals* estimated *in numerical simulations (for each Hamiltonian empirical standard deviations were obtained by repeating simulation 100 times). Plot a) shows results for the noiseless scenario, while b) shows results for the error-mitigated scenario with noise mitigation performed on marginals and noise model inspired by IBM's device's characterization.*

Clearly, if the size of the marginals k is fixed, setting $r = c\log(N/\epsilon)$, for a suitable constant c, allows us to conclude that for all γ such that $|\gamma| \leq k$ one has $\|\rho_\gamma - \mathrm{id}_\gamma\|_1 \leq \epsilon$ with probability approaching 1 with the increasing system size.

4.7.8 Effects of measurement noise

To conclude, let us provide some analysis of the effects of measurement noise mitigation on the above considerations. First, let us note that it is straightforward to generalize all of the above arguments to include the *uncorrelated* readout noise. Intuitively, if the measurement noise is not correlated, it cannot increase

the level of correlations in the energy estimators, and the same holds for noise-mitigation for such model (see Appendix B.4.3 for derivations).

For the correlated noise model, it is hard to provide analytical results, however, one can expect that if correlations in measurement noise are mild, then it should not drastically increase sample complexity. To test this hypothesis for small system sizes, we performed numerical simulations in the following way. Consider confidence intervals of the energy estimation. The most pessimistic bound on the error, as already explained, is given by Eq. (4.19) and is additive in the number of Hamiltonian terms K (if one considers the situation without measurement noise, then it suffices to set $\delta_\alpha = 0$ and $||C_{av}^{S_\alpha}||_{1\to 1} = 1$ and Eq. (4.19) still holds). To obtain confidence intervals expected for uncorrelated variables, we simply set $\mathrm{Var}(\mathcal{H}) = \sum_\alpha \mathrm{Var}(H_\alpha)$, and provide corresponding confidence interval by calculating LHS of Eq. (4.26). To test whether resulting bounds are close to what happens in practice, we numerically *estimate* the variance of $\mathcal{H}$ and calculate the resulting confidence intervals. Such estimation of variance can be done, for example, by performing multiple numerical experiments, each giving an empirical estimate of $\mathrm{Var}(\mathcal{H})$ (for a fixed number of samples s), and taking the mean of those estimates. Such comparison is plotted on Fig. 4.5 for the system of $N = 8$ qubits and Hamiltonians estimated on the states coming from 3-layer QAOA. Bound P_f on the probability (of energy estimator being outside the calculated confidence interval) in Chebyshev's inequality is set to $P_f = 0.04$ which corresponds to 5σ-confidence. Shown are results for the noiseless scenario, and for the error-mitigated estimators with the noise model inspired by IBM's device characterization. It is clear that for tested Hamiltonians the confidence intervals in the estimation behave roughly like for uncorrelated variables.

4.8 Effects of measurement errors on QAOA – numerical study

In this section we apply our noise characterization and mitigation strategies to numerically study the effects of correlated measurement errors on QAOA and how they can be reduced with our techniques. As test Hamiltonians, we choose those encoding random MAX-2SAT instances with clause density 4, the Hamiltonians corresponding to fully-connected graphs with random interactions and local fields with magnitude from $[-1, 1]$, and the Sherrington-Kirkpatrick (SK) [176] model in 2D (i.e. random Gaussian ZZ interactions on a square lattice).

For all models, we classically simulate a QAOA algorithm on an 8-qubit device with a number of layers ranging from $p = 3$ to 30. The parameter optimization is performed using Simultaneous perturbation stochastic approximation (SPSA) [177; 178; 156; 61] (see Appendix B.5.1 for details of optimization). As a correlated noise model, we adopt one inspired by our previous characterization of the IBM 15-qubit device.

We first simulated the result obtained by a QAOA algorithm where both the energy estimation and the gradient estimation used to guide the state evolution were affected by readout noise. For different numbers of gate layers, we compared the resulting energy estimators between an optimization guided by noisy

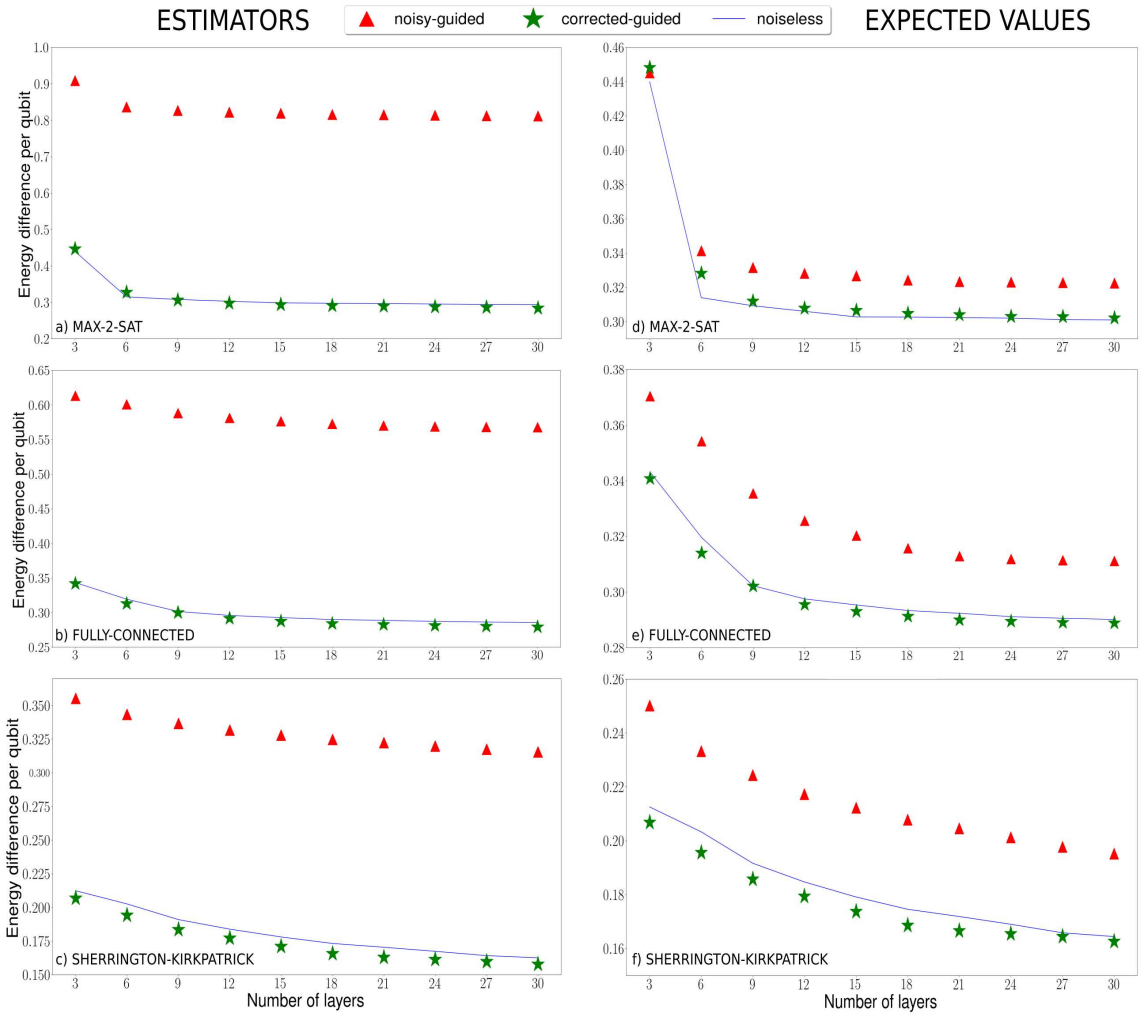


Figure 4.6: *Numerical study of the effects of readout noise on energy estimation for QAOA on an 8-qubit system. The algorithm is used to prepare the ground state of a,d) Hamiltonians encoding random MAX 2-SAT problems, b,e) Hamiltonian corresponding to a fully-connected graph with uniformly random (from* $[-1, 1]$*) interactions and local fields, and c,f) the* 2*D Sherrington-Kirkpatrick model. In each plot, the horizontal axis shows a number of layers in the QAOA optimization, while the vertical axis shows the absolute difference between obtained and theoretical energies per qubit. Each data point is the* average *over 96 Hamiltonians. Note that we made an offset on the y-axis in order to make differences visible. Red data-points indicate optimization guided by noisy function evaluations, green points indicate noise-mitigated evaluations, while blue lines correspond to noiseless optimization given for reference. The estimators were obtained from* $\approx 10^4$ *samples. Due to differences in spectra of various Hamiltonians over which the mean is calculated the fluctuations around the presented means are very high, therefore for clarity we decided not to include error bars.*

estimators (“noisy-guided” in Fig. 4.6) and the optimization guided by estimators on which error-mitigation was performed (“corrected-guided” in Fig. 4.6). The results of our numerical studies are presented in the first column of Fig. 4.6,

together with the noiseless case as a reference. Note that to make differences more visible, we set offset on the vertical axis. It is clear that for the considered Hamiltonians, the noise-mitigated estimators are much better than noisy ones. This suggests that our noise-mitigation scheme can be used to obtain an overall more reliable QAOA algorithm.

For our second analysis, we wanted to closely analyze the effects of noise (and its mitigation) on the parameter optimization only. To do so, we still compare the results of QAOA between the cases where the optimization is guided by a noiseless, noisy ("noisy-guided" in Fig. 4.6) and noise-mitigated ("corrected-guided" in Fig. 4.6) energy estimators. However, in order to isolate the effect of noise on the optimization procedure, instead of sampling the energy from the noisy probability distributions, we calculated its expectation value directly on the quantum state obtained at each layer of the QAOA circuit (the same circuits as in left column). As a theoretical comparison between the optimal parameters found by QAOA, we took the distance between the resulting average energies on the state, divided by the number of qubits. The plots in the second column of Fig. 4.6 show our numerical results. For both of the considered models, the noise-mitigated optimization leads to better parameters regions. We note that the difference is not high, yet it is systematic. Since this is the case already for 8 qubits, one can expect that for bigger systems, the relative improvement will be higher.

To conclude, let us stress that the purpose of this section was to illustrate the possible effects of mildly correlated, realistic readout noise on QAOA algorithm. The above results clearly indicate that correlated readout noise can potentially influence the optimization, as opposed to an identical and uncorrelated one (see Ref. [153]). We find that in a number of instances the error-mitigation strategy helps to land in better parameters regions, however, we also note that the effects of noise on the optimization are not dramatic in the studied cases. From the point of view of near-term applications, this should be viewed as a positive result – the noise does not seem to strongly affect optimization, yet even those mild effects can be reduced by performing our error-mitigation. Furthermore, we note that due to the stochastic nature of the SPSA optimizer, the results might differ from run to run (each data point presented in the plots comes from the results of the optimization run which was the better one amongst two performed independent optimizations, see Appendix B.5.1 for details). Finally, we note that to obtain accurate estimates of energy, the noise-mitigation, unsurprisingly, remains highly beneficial.

4.9 Summary

In the first part of this Chapter, we proposed an efficiently describable model of correlated measurement noise in quantum detectors. The basic idea of the model is to group qubits into strongly-correlated clusters that are mildly affected by their neighborhoods, which, provided that the size of those groups is bounded by a constant, allows to describe a global noise model by much smaller number

of parameters compared to the most generic situation. To characterize our noise model, we have introduced Diagonal Detector Overlapping Tomography, which is a procedure inspired by recently introduced Quantum Overlapping Tomography [151], tailored to the efficient characterization of the proposed measurement noise model. Similarly to the [151], proposed method can estimate k-qubit correlations in measurement noise affecting N-qubit device using $O\left(k2^k \log\left(N\right)\right)$ randomly chosen quantum circuits. We have shown that the measurement noise can be efficiently mitigated in problems that require estimation of multiple marginal probability distributions, an example of which is Quantum Approximate Optimization Algorithm [144]. Importantly, from the fact that noise-mitigation is performed on marginal distributions it follows that sampling-complexity of noise mitigation is similar to that of the original problem, provided that cross-talk in readout noise is of bounded locality. We proposed a benchmark of the noise model and error reduction, which we implemented in experiments on up to 15 qubits on IBM's *Melbourne* device and on 23 qubits on Rigetti's *Aspen-8* device, and concluded significant improvements compared to simple, uncorrelated noise model. Interestingly, additional experimental data have pointed at previously unreported memory effects in IBM's device that were demonstrated to not-negligibly change the results of experiments.

In the second part of the Chapter, we provided an analysis of the statistical errors one may expect when performing the simultaneous estimation of multiple local terms of Hamiltonian on various classes of states. We provided simple arguments why low sampling complexity (i.e., scaling of the variance as the square-root of a number of local Hamiltonian terms) should be expected from Haar-random quantum states, and for states generated by shallow random quantum circuits. Similarly, we gave some arguments based on [171], why for states appearing at the beginning and at the end of the QAOA, one may expect that estimated local terms will effectively behave as uncorrelated variables, reducing the sampling complexity. Furthermore, we have provided analytical results for Hamiltonians encoding random MAX-2-SAT instances.

In the last part of the Chapter, we have presented numerical results and extended discussion of the effects that correlated measurement noise can have on the performance of QAOA. We have demonstrated that already for 8 qubits the correlated measurement noise can alter the energy landscape in such a way, that the quantum-classical optimization leads to sub-optimal energy regions (compared to reference runs without noise). At the same time, we demonstrated that our noise-mitigation procedure can reduce those effects, improving optimization.

5. Discussion

5.0.1 Summary

In this Thesis, we studied, from multiple perspectives, how classical randomness can be useful for analysis of various quantum computing protocols. In Chapter 3, we introduced quantum average-case (AC) distances between quantum states, quantum measurements, and quantum channels. Those distances, expressed via simple functions of underlying objects, bound the average-case total-variation distances between corresponding probability distributions (recall Theorems 1, 2, and 3). Hence, the AC distances have sound operational interpretation of quantifying the statistical distinguishability of two objects using randomized protocols. Indeed, as discussed in Section 3.5, considering the fact that our bounds hold for approximate unitary 4-designs, it follows that, in principle, even very high-dimensional objects could be distinguished with randomized protocols involving circuits of very low (logarithmic in the number of qubits) depth. While this certainly is an interesting insight, "there is no free lunch", and, in general, the cost of the distinguishability is moved from circuit depth to potentially exponentially costly classical post-processing. Intuitively, this means that while we can always implement a random ensemble of shallow-depth circuits, we do not provide a recipe for how to use the measurement outcomes to actually distinguish between given two objects (and we expect it to be, in general, very costly).

While this finding doesn't thus provide a *practical* application of classical randomness for statistical distinguishably, our further analysis in Section 3.5 demonstrates that AC distances provide a flexible new tool for analyzing the average-case behavior of noisy quantum protocols. In particular, we have shown that, in simplified scenarios, we can calculate AC distances to obtain bounds on how, on average, the noisy probability distribution (e.g., resulting from imperfect realization of a quantum measurement) i) deviates from ideal distribution (resulting from theoretical, ideal quantum measurement); and ii) converges to useless, uniform distribution (equivalent to performing "trivial" measurement that returns outcomes completely at random). In this case, we have demonstrated that **the classical randomness can be viewed as a useful tool that allows to gain insights on the average-case behavior of noisy quantum protocols**.

On more technical side, we studied multiple properties of AC distances – we have shown that our distances indeed fulfill metric axioms, they are subaddi-

tive with respect to tensor products, have a joint convexity property, and are non-increasing under *unital* quantum channels. We analyzed in detail exemplary scenarios where we believed AC distance provided an insight into some aspect of quantum information science – such as consequences of noise accumulation for protocols involving random circuits in Section 3.5. Moreover, in Section 3.5.4, we performed extensive numerical studies that demonstrated the usefulness of AC distances as compared to worst-case distances. This includes practically-relevant ensembles of random quantum circuits with the structure of variational circuits, used in hybrid quantum-classical protocols such as Quantum Approximate Optimization Algorithm (QAOA) and Variational Quantum Eigensolver (VQE).

In Chapter 4, we have turned our attention to the problem of characterizing and reducing the effects of classical measurement noise. We have introduced Diagonal Detector Overlapping Tomography – a protocol that exploits classical randomness to characterize locally-correlated measurement noise. We have shown that the protocol is scalable, in the sense that it requires only $O\left(2^{(k)\log(n)}\right)$ circuits to characterize k-local noise on n-qubit device. Here, **the classical randomness turned out to be an indispensable resource allowing for efficient characterization of measurement noise in experiments**.

Moreover, in that Chapter, we proposed a simple noise model that aims to capture small-range correlations between qubits in a non-trivial way. We considered how additional classical post-processing, based on DDOT results, can be used to perform error mitigation on the level of marginal probability distributions. We further studied correlation decay in local observables of random graphs in the standard QAOA setting and presented bounds on the variance of generic Hamiltonians estimated on states resulting from the application of shallow QAOA circuits.

To support the practical relevance of our findings, we performed extensive numerical simulations that investigated the effects of correlated measurement noise on QAOA optimization (together with error mitigation). Importantly, we presented experimental results obtained on Rigetti's and IBM's superconducting quantum systems, that demonstrated the effectiveness of proposed error characterization and mitigation strategies. We obtained a reduction of errors by a factor as big as ~ 22 for 15-qubit experiments on IBM's device.

We believe that, across both main Chapters of this Thesis, we have provided multiple scenarios which demonstrated that

> Auxiliary classical resources can be used to assess and improve the quality of the implementation of noisy quantum measurements

As such, we are convinced that our findings provide strong arguments in favor of the principal hypothesis advanced in the Thesis. Let us now conclude the dissertation by discussing possible future research directions.

5.0.2 Future research directions

Average-case quantum distances

Our work presented in Chapter 3 leaves many interesting problems left for future research. The first important question one can ask is how to estimate quantum average-case distances in an easy-to-implement setting. A natural candidate seems to be randomized benchmarking types of experiments, as they also employ unitary designs [89; 79]. It would be also very interesting to connect quantum average-case distances with commonly used figures of merit used to assess the quality of quantum devices. Those include measures such as average fidelity [89] (which is perhaps the most widely used quality measure), unitarity of quantum channels [179; 180; 79], or partitioned trace distances [181]. We note that the partitioned trace distances share with the average-case distances the property of being non-increasing under unital channels, which might suggest a deeper connection between the two. Furthermore, one can ask whether similar results can be obtained for different functions of output probability distributions such us classical fidelity or f-divergences [182; 183]. Another natural direction to pursue is to obtain better constants that appear in bounds (for example by considering higher moments) relating the average TV distance with the quantum average-case distance. Another straightforward research direction is to check how the quantum average-case distances compare with worst-case distances for small subsets of qubits in actual quantum devices. For example, for pairs of qubits full Quantum Process Tomography [17] (or even Gate Set Tomography [184]) is possible, therefore one would be able to calculate the distances directly from objects in question. Such studies could provide some insight into what to expect from existing devices in worst and average-case scenarios.

Readout noise characterization and mitigation

We believe that our findings from Chapter 4 will prove useful in both near- and long-term applications as efficient methods of characterizing and reducing measurement noise. At the same time, we find that a number of future research directions opens. It is natural to ask how well the error mitigation will perform in actual multi-qubit experiments of QAOA, which can be tested only with more access to the quantum devices than is available for the public via cloud services. Similarly, testing noise mitigation in more general, Variational Quantum Eigensolver (VQE) scenarios is of great interest [185; 186].

The analysis of statistical errors in the VQE setup is a natural extension – we note that since the local terms in VQE Hamiltonians do not commute, it is less straightforward than in the QAOA scenario. Another interesting problem is to design more benchmarks for the noise model and mitigation than the one method described in this work. A natural extension of techniques presented in this work would be to develop more methods for inference of correlations structure and construction of noise model from DDOT data. Having a noise model for marginal probability distributions, it is desirable to test and compare techniques of noise mitigiation that go beyond a standard noise matrix inversion analyzed in this work. In particular, methods based on Bayesian inference [138], or on symmetrization

of noise via randomized twirling [122; 124], seem promising.

Moreover, it is tempting to ask whether error-mitigation performance and demonstrated memory effects are stable over time, a question which is often omitted (see however recent work [187] where systematic methods for studying time instabilities were developed and [188] which made an important contribution by studying the stability of various types of noise over time for IBM's devices). Another open problem is to find out whether further generalizations of Quantum Overlapping Tomography can be used to perform reliable characterization of more general types of noise – coherent measurement noise and the noise affecting quantum gates. For example, in the work [189] the authors consider an ansatz for the generic noisy quantum process which uses only 2-local noise processes – a setup which seems natural to benefit from similar techniques. In Refs. [190; 191] the authors develop methods of estimating generic Pauli channels, and it would be of great interest to assess whether those methods can benefit from using measurement error-mitigation techniques such as ours. A very important research problem is that of mitigating measurement noise in scenarios which include estimators obtained from a very few samples, such as those in Refs. [84; 192]. We note that in its current form our methods cannot be directly implemented in such scenarios because they operate on marginal probability distributions.

Finally, it would be extremely interesting to investigate whether our model of measurement noise is accurate for quantum devices based on architectures different than transmon qubits, such as flux qubits [193], trapped ions [194] or photonic quantum devices [195]. We intend to investigate some of the listed problems in future works.

Bibliography

[1] Filip B. Maciejewski, Zbigniew Puchała, and Michał Oszmaniec. Operational quantum average-case distances. *Quantum*, 7:1106, September 2023. URL: http://dx.doi.org/10.22331/q-2023-09-11-1106, doi:10.22331/q-2023-09-11-1106.

[2] Filip B. Maciejewski, Zbigniew Puchała, and Michał Oszmaniec. Exploring quantum average-case distances: Proofs, properties, and examples. *IEEE Transactions on Information Theory*, 69(7):4600–4619, 2023. doi:10.1109/TIT.2023.3250100.

[3] Filip B. Maciejewski, Flavio Baccari, Zoltá n Zimborás, and Michał Oszmaniec. Modeling and mitigation of cross-talk effects in readout noise with applications to the quantum approximate optimization algorithm. *Quantum*, 5:464, 6 2021. URL: https://doi.org/10.22331%2Fq-2021-06-01-464, doi:10.22331/q-2021-06-01-464.

[4] AK Wróblewski. Historia fizyki, od czasów najdawniejszych do współczesności, 2007.

[5] Scott Aaronson. *Quantum Computing since Democritus*. Cambridge University Press, Cambridge, 2013.

[6] P. Liang, T. Hashimoto, C. Re, R. Bommasani, and S.M. Xie. Stanford course "cs324 - large language models". URL: https://stanford-cs324.github.io/winter2022/.

[7] Seth Lloyd. Universal quantum simulators. *Science*, 273(5278):1073–1078, 1996. URL: https://www.science.org/doi/abs/10.1126/science.273.5278.1073, arXiv:https://www.science.org/doi/pdf/10.1126/science.273.5278.1073, doi:10.1126/science.273.5278.1073.

[8] Seth Lloyd. Ultimate physical limits to computation. *Nature*, 406(6799):1047–1054, Aug 2000. doi:10.1038/35023282.

[9] David Poulin, Angie Qarry, Rolando Somma, and Frank Verstraete. Quantum simulation of time-dependent hamiltonians and the convenient illusion of hilbert space. *Physical Review Letters*, 106(17), April 2011. URL: http://dx.doi.org/10.1103/PhysRevLett.106.170501, doi:10.1103/physrevlett.106.170501.

[10] Ludovico Lami and Bartosz Regula. No second law of entanglement manipulation after all. *Nature Physics*, 19(2):184–189, Feb 2023. doi:10.1038/s41567-022-01873-9.

[11] Julia Kempe, Alexei Kitaev, and Oded Regev. The complexity of the local hamiltonian problem, 2005. URL: https://arxiv.org/abs/quant-ph/0406180, arXiv:quant-ph/0406180.

[12] Toby Cubitt and Ashley Montanaro. Complexity classification of local hamiltonian problems. *SIAM Journal on Computing*, 45(2):268–316, 2016.

[13] Chi-Fang Chen, Hsin-Yuan Huang, John Preskill, and Leo Zhou. Local minima in quantum systems. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1323–1330, 2024.

[14] Matthew F. Pusey, Jonathan Barrett, and Terry Rudolph. On the reality of the quantum state. *Nature Physics*, 8(6):475–478, May 2012. URL: http://dx.doi.org/10.1038/nphys2309, doi:10.1038/nphys2309.

[15] Daniela Frauchiger and Renato Renner. Quantum theory cannot consistently describe the use of itself. *Nature communications*, 9(1):3711, 2018.

[16] Kok-Wei Bong, Aníbal Utreras-Alarcón, Farzad Ghafari, Yeong-Cherng Liang, Nora Tischler, Eric G. Cavalcanti, Geoff J. Pryde, and Howard M. Wiseman. A strong no-go theorem on the wigner's friend paradox. *Nature Physics*, 16(12):1199–1205, Dec 2020. doi:10.1038/s41567-020-0990-x.

[17] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. doi:10.1017/CBO9780511976667.

[18] Melanie Mitchell. *Complexity: A Guided Tour*. Oxford University Press, Inc., USA, 2009.

[19] Joschka Roffe. Quantum error correction: an introductory guide. *Contemporary Physics*, 60(3):226–245, July 2019. URL: http://dx.doi.org/10.1080/00107514.2019.1667078, doi:10.1080/00107514.2019.1667078.

[20] Emanuel Knill, Raymond Laflamme, and Lorenza Viola. Theory of quantum error correction for general noise. *Physical Review Letters*, 84(11):2525–2528, March 2000. URL: http://dx.doi.org/10.1103/PhysRevLett.84.2525, doi:10.1103/physrevlett.84.2525.

[21] Frank Arute et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 10 2019. doi:10.1038/s41586-019-1666-5.

[22] IBM. Ibm quantum website. URL: https://www.ibm.com/quantum.

[23] Google. Google quantum ai website. URL: https://quantumai.google/.

[24] Rigetti. Rigetti computing website. URL: https://www.rigetti.com/.

[25] IonQ. Ionq website. URL: https://ionq.com/.

[26] Quantinuum. Quantinuum website. URL: https://www.quantinuum.com/.

[27] Masoud Mohseni, Artur Scherer, K. Grace Johnson, Oded Wertheim, Matthew Otten, Navid Anjum Aadit, Kirk M. Bresniker, Kerem Y. Camsari, Barbara Chapman, Soumitra Chatterjee, Gebremedhin A. Dagnew, Aniello Esposito, Farah Fahim, Marco Fiorentino, Abdullah Khalid, Xiangzhou Kong, Bohdan Kulchytskyy, Ruoyu Li, P. Aaron Lott, Igor L. Markov, Robert F. McDermott, Giacomo Pedretti, Archit Gajjar, Allyson Silva, John Sorebo, Panagiotis Spentzouris, Ziv Steiner, Boyan Torosov, Davide Venturelli, Robert J. Visser, Zak Webb, Xin Zhan, Yonatan Cohen, Pooya Ronagh, Alan Ho, Raymond G. Beausoleil, and John M. Martinis. How to build a quantum supercomputer: Scaling challenges and opportunities, 2024. URL: https://arxiv.org/abs/2411.10406, arXiv:2411.10406.

[28] Sergey Bravyi, Andrew W. Cross, Jay M. Gambetta, Dmitri Maslov, Patrick Rall, and Theodore J. Yoder. High-threshold and low-overhead fault-tolerant quantum memory, 2024. arXiv:2308.07915.

[29] John Preskill. Quantum computing in the nisq era and beyond. *Quantum*, 2:79, August 2018. URL: http://dx.doi.org/10.22331/q-2018-08-06-79, doi:10.22331/q-2018-08-06-79.

[30] Rajeev Acharya, Igor Aleiner, Richard Allen, Trond I. Andersen, Markus Ansmann, Frank Arute, Kunal Arya, Abraham Asfaw, Juan Atalaya, Ryan Babbush, Dave Bacon, Joseph C. Bardin, Joao Basso, Andreas Bengtsson, Sergio Boixo, Gina Bortoli, Alexandre Bourassa, Jenna Bovaird, Leon Brill, Michael Broughton, Bob B. Buckley, David A. Buell, Tim Burger, Brian Burkett, Nicholas Bushnell, Yu Chen, Zijun Chen, Ben Chiaro, Josh Cogan, Roberto Collins, Paul Conner, William Courtney, Alexander L. Crook, Ben Curtin, Dripto M. Debroy, Alexander Del Toro Barba, Sean Demura,

Andrew Dunsworth, Daniel Eppens, Catherine Erickson, Lara Faoro, Edward Farhi, Reza Fatemi, Leslie Flores Burgos, Ebrahim Forati, Austin G. Fowler, Brooks Foxen, William Giang, Craig Gidney, Dar Gilboa, Marissa Giustina, Alejandro Grajales Dau, Jonathan A. Gross, Steve Habegger, Michael C. Hamilton, Matthew P. Harrigan, Sean D. Harrington, Oscar Higgott, Jeremy Hilton, Markus Hoffmann, Sabrina Hong, Trent Huang, Ashley Huff, William J. Huggins, Lev B. Ioffe, Sergei V. Isakov, Justin Iveland, Evan Jeffrey, Zhang Jiang, Cody Jones, Pavol Juhas, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Tanuj Khattar, Mostafa Khezri, Mária Kieferová, Seon Kim, Alexei Kitaev, Paul V. Klimov, Andrey R. Klots, Alexander N. Korotkov, Fedor Kostritsa, John Mark Kreikebaum, David Landhuis, Pavel Laptev, Kim-Ming Lau, Lily Laws, Joonho Lee, Kenny Lee, Brian J. Lester, Alexander Lill, Wayne Liu, Aditya Locharla, Erik Lucero, Fionn D. Malone, Jeffrey Marshall, Orion Martin, Jarrod R. McClean, Trevor McCourt, Matt McEwen, Anthony Megrant, Bernardo Meurer Costa, Xiao Mi, Kevin C. Miao, Masoud Mohseni, Shirin Montazeri, Alexis Morvan, Emily Mount, Wojciech Mruczkiewicz, Ofer Naaman, Matthew Neeley, Charles Neill, Ani Nersisyan, Hartmut Neven, Michael Newman, Jiun How Ng, Anthony Nguyen, Murray Nguyen, Murphy Yuezhen Niu, Thomas E. O'Brien, Alex Opremcak, John Platt, Andre Petukhov, Rebecca Potter, Leonid P. Pryadko, Chris Quintana, Pedram Roushan, Nicholas C. Rubin, Negar Saei, Daniel Sank, Kannan Sankaragomathi, Kevin J. Satzinger, Henry F. Schurkus, Christopher Schuster, Michael J. Shearn, Aaron Shorter, Vladimir Shvarts, Jindra Skruzny, Vadim Smelyanskiy, W. Clarke Smith, George Sterling, Doug Strain, Marco Szalay, Alfredo Torres, Guifre Vidal, Benjamin Villalonga, Catherine Vollgraff Heidweiller, Theodore White, Cheng Xing, Z. Jamie Yao, Ping Yeh, Juhwan Yoo, Grayson Young, Adam Zalcman, Yaxing Zhang, Ningfeng Zhu, and Google Quantum AI. Suppressing quantum errors by scaling a surface code logical qubit. *Nature*, 614(7949):676–681, Feb 2023. doi:10.1038/s41586-022-05434-1.

[31] Dolev Bluvstein, Simon J Evered, Alexandra A Geim, Sophie H Li, Hengyun Zhou, Tom Manovitz, Sepehr Ebadi, Madelyn Cain, Marcin Kalinowski, Dominik Hangleiter, et al. Logical quantum processor based on reconfigurable atom arrays. *Nature*, 626(7997):58–65, 2024.

[32] Ben W. Reichardt, David Aasen, Rui Chao, Alex Chernoguzov, Wim van Dam, John P. Gaebler, Dan Gresh, Dominic Lucchetti, Michael Mills, Steven A. Moses, Brian Neyenhuis, Adam Paetznick, Andres Paz, Peter E. Siegfried, Marcus P. da Silva, Krysta M. Svore, Zhenghan Wang, and Matt Zanner. Demonstration of quantum computation and error correction with a tesseract code, 2024. URL: https://arxiv.org/abs/2409.04628, arXiv:2409.04628.

[33] Rajeev Acharya, Dmitry A. Abanin, Laleh Aghababaie-Beni, Igor Aleiner,

Trond I. Andersen, Markus Ansmann, Frank Arute, Kunal Arya, Abraham Asfaw, Nikita Astrakhantsev, Juan Atalaya, Ryan Babbush, Dave Bacon, Brian Ballard, Joseph C. Bardin, Johannes Bausch, Andreas Bengtsson, Alexander Bilmes, Sam Blackwell, Sergio Boixo, Gina Bortoli, Alexandre Bourassa, Jenna Bovaird, Leon Brill, Michael Broughton, David A. Browne, Brett Buchea, Bob B. Buckley, David A. Buell, Tim Burger, Brian Burkett, Nicholas Bushnell, Anthony Cabrera, Juan Campero, Hung-Shen Chang, Yu Chen, Zijun Chen, Ben Chiaro, Desmond Chik, Charina Chou, Jahan Claes, Agnetta Y. Cleland, Josh Cogan, Roberto Collins, Paul Conner, William Courtney, Alexander L. Crook, Ben Curtin, Sayan Das, Alex Davies, Laura De Lorenzo, Dripto M. Debroy, Sean Demura, Michel Devoret, Agustin Di Paolo, Paul Donohoe, Ilya Drozdov, Andrew Dunsworth, Clint Earle, Thomas Edlich, Alec Eickbusch, Aviv Moshe Elbag, Mahmoud Elzouka, Catherine Erickson, Lara Faoro, Edward Farhi, Vinicius S. Ferreira, Leslie Flores Burgos, Ebrahim Forati, Austin G. Fowler, Brooks Foxen, Suhas Ganjam, Gonzalo Garcia, Robert Gasca, Élie Genois, William Giang, Craig Gidney, Dar Gilboa, Raja Gosula, Alejandro Grajales Dau, Dietrich Graumann, Alex Greene, Jonathan A. Gross, Steve Habegger, John Hall, Michael C. Hamilton, Monica Hansen, Matthew P. Harrigan, Sean D. Harrington, Francisco J. H. Heras, Stephen Heslin, Paula Heu, Oscar Higgott, Gordon Hill, Jeremy Hilton, George Holland, Sabrina Hong, Hsin-Yuan Huang, Ashley Huff, William J. Huggins, Lev B. Ioffe, Sergei V. Isakov, Justin Iveland, Evan Jeffrey, Zhang Jiang, Cody Jones, Stephen Jordan, Chaitali Joshi, Pavol Juhas, Dvir Kafri, Hui Kang, Amir H. Karamlou, Kostyantyn Kechedzhi, Julian Kelly, Trupti Khaire, Tanuj Khattar, Mostafa Khezri, Seon Kim, Paul V. Klimov, Andrey R. Klots, Bryce Kobrin, Pushmeet Kohli, Alexander N. Korotkov, Fedor Kostritsa, Robin Kothari, Borislav Kozlovskii, John Mark Kreikebaum, Vladislav D. Kurilovich, Nathan Lacroix, David Landhuis, Tiano Lange-Dei, Brandon W. Langley, Pavel Laptev, Kim-Ming Lau, Loïck Le Guevel, Justin Ledford, Joonho Lee, Kenny Lee, Yuri D. Lensky, Shannon Leon, Brian J. Lester, Wing Yan Li, Yin Li, Alexander T. Lill, Wayne Liu, William P. Livingston, Aditya Locharla, Erik Lucero, Daniel Lundahl, Aaron Lunt, Sid Madhuk, Fionn D. Malone, Ashley Maloney, Salvatore Mandrà, James Manyika, Leigh S. Martin, Orion Martin, Steven Martin, Cameron Maxfield, Jarrod R. McClean, Matt McEwen, Seneca Meeks, Anthony Megrant, Xiao Mi, Kevin C. Miao, Amanda Mieszala, Reza Molavi, Sebastian Molina, Shirin Montazeri, Alexis Morvan, Ramis Movassagh, Wojciech Mruczkiewicz, Ofer Naaman, Matthew Neeley, Charles Neill, Ani Nersisyan, Hartmut Neven, Michael Newman, Jiun How Ng, Anthony Nguyen, Murray Nguyen, Chia-Hung Ni, Murphy Yuezhen Niu, Thomas E. O'Brien, William D. Oliver, Alex Opremcak, Kristoffer Ottosson, Andre Petukhov, Alex Pizzuto, John Platt, Rebecca Potter, Orion Pritchard, Leonid P. Pryadko, Chris Quintana, Ganesh Ramachandran, Matthew J. Reagor, John Redding, David M. Rhodes, Gabrielle Roberts, Eliott Rosenberg, Emma Rosenfeld, Pedram Roushan,

Nicholas C. Rubin, Negar Saei, Daniel Sank, Kannan Sankaragomathi, Kevin J. Satzinger, Henry F. Schurkus, Christopher Schuster, Andrew W. Senior, Michael J. Shearn, Aaron Shorter, Noah Shutty, Vladimir Shvarts, Shraddha Singh, Volodymyr Sivak, Jindra Skruzny, Spencer Small, Vadim Smelyanskiy, W. Clarke Smith, Rolando D. Somma, Sofia Springer, George Sterling, Doug Strain, Jordan Suchard, Aaron Szasz, Alex Sztein, Douglas Thor, Alfredo Torres, M. Mert Torunbalci, Abeer Vaishnav, Justin Vargas, Sergey Vdovichev, Guifre Vidal, Benjamin Villalonga, Catherine Vollgraff Heidweiller, Steven Waltman, Shannon X. Wang, Brayden Ware, Kate Weber, Travis Weidel, Theodore White, Kristi Wong, Bryan W. K. Woo, Cheng Xing, Z. Jamie Yao, Ping Yeh, Bicheng Ying, Juhwan Yoo, Noureldin Yosri, Grayson Young, Adam Zalcman, Yaxing Zhang, Ningfeng Zhu, Nicholas Zobrist, Google Quantum AI, and Collaborators. Quantum error correction below the surface code threshold. *Nature*, Dec 2024. doi:10.1038/s41586-024-08449-y.

[34] Zhenyu Cai, Ryan Babbush, Simon C. Benjamin, Suguru Endo, William J. Huggins, Ying Li, Jarrod R. McClean, and Thomas E. O'Brien. Quantum error mitigation, 2023. arXiv:2210.00921.

[35] Yihui Quek, Daniel Stilck França, Sumeet Khatri, Johannes Jakob Meyer, and Jens Eisert. Exponentially tighter bounds on limitations of quantum error mitigation, 2023. arXiv:2210.11505.

[36] Yasunari Suzuki, Suguru Endo, Keisuke Fujii, and Yuuki Tokunaga. Quantum error mitigation as a universal error reduction technique: Applications from the nisq to the fault-tolerant quantum computing eras. *PRX Quantum*, 3(1):010345, 2022.

[37] Andreas Elben, Steven T. Flammia, Hsin-Yuan Huang, Richard Kueng, John Preskill, Benoît Vermersch, and Peter Zoller. The randomized measurement toolbox. *Nature Reviews Physics*, 5(1):9–24, December 2022. URL: http://dx.doi.org/10.1038/s42254-022-00535-2, doi:10.1038/s42254-022-00535-2.

[38] Michał Oszmaniec and Tanmoy Biswas. Operational relevance of resource theories of quantum measurements. *Quantum*, 3:133, 4 2019. doi:10.22331/q-2019-04-26-133.

[39] Asher Peres. *Quantum theory: Concepts and methods*, volume 57. Springer Science & Business Media, 2006. doi:10.1007/0-306-47120-5.

[40] G. Chiribella, G. M. D'Ariano, and P. Perinotti. Transforming quantum operations: Quantum supermaps. *EPL (Europhysics Letters)*, 83(3):30004, July 2008. URL: http://dx.doi.org/10.1209/0295-5075/83/30004, doi:10.1209/0295-5075/83/30004.

[41] Yanzhu Chen, Maziar Farahzad, Shinjae Yoo, and Tzu-Chieh Wei. Detector tomography on IBM quantum computers and mitigation of an imperfect measurement. *Physical Review A*, 100(5), 11 2019. URL: http://dx.doi.org/10.1103/PhysRevA.100.052315, doi:10.1103/physreva.100.052315.

[42] Filip B. Maciejewski, Zoltán Zimborás, and Michał Oszmaniec. Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography. *Quantum*, 4:257, 4 2020. doi:10.22331/q-2020-04-24-257.

[43] Carl W. Helstrom. *Quantum detection and estimation theory*, volume 1. 6 1969. doi:10.1007/BF01007479.

[44] Miguel Navascués and Sandu Popescu. How energy conservation limits our measurements. *Phys. Rev. Lett.*, 112:140502, 4 2014. URL: https://link.aps.org/doi/10.1103/PhysRevLett.112.140502, doi:10.1103/PhysRevLett.112.140502.

[45] Zbigniew Puchała, Łukasz Pawela, Aleksandra Krawiec, and Ryszard Kukulski. Strategies for optimal single-shot discrimination of quantum measurements. *Physical Review A*, 98(4), 10 2018. URL: http://dx.doi.org/10.1103/PhysRevA.98.042103, doi:10.1103/physreva.98.042103.

[46] Zbigniew Puchała, Łukasz Pawela, Aleksandra Krawiec, Ryszard Kukulski, and Michał Oszmaniec. Multiple-shot and unambiguous discrimination of von neumann measurements. *Quantum*, 5:425, 4 2021. URL: http://dx.doi.org/10.22331/q-2021-04-06-425, doi:10.22331/q-2021-04-06-425.

[47] John Watrous. Semidefinite programs for completely bounded norms. *Theory of Computing*, 5(11):217–238, 2009. URL: http://www.theoryofcomputing.org/articles/v005a011, doi:10.4086/toc.2009.v005a011.

[48] Antonio Anna Mele. Introduction to haar measure tools in quantum information: A beginneramp;apos;s tutorial. *Quantum*, 8:1340, May 2024. URL: http://dx.doi.org/10.22331/q-2024-05-08-1340, doi:10.22331/q-2024-05-08-1340.

[49] Andris Ambainis and Joseph Emerson. Quantum T-designs: T-wise independence in the quantum world. In *Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity*, CCC '07, page 129–140, USA, 2007. IEEE Computer Society. doi:10.1109/CCC.2007.26.

[50] Richard A. Low. *Pseudo-randomness and Learning in Quantum Computation*. PhD thesis, -, 6 2010.

[51] Fernando G. S. L. Brandão, Aram W. Harrow, and Michał Horodecki. Local random quantum circuits are approximate polynomial-designs. *Communications in Mathematical Physics*, 346(2):397–434, 9 2016. arXiv:1208.0692, doi:10.1007/s00220-016-2706-8.

[52] Michał Oszmaniec, Adam Sawicki, and Michał Horodecki. Epsilon-nets, unitary designs, and random quantum circuits. *IEEE Transactions on Information Theory*, 68(2):989–1015, 2022. doi:10.1109/TIT.2021.3128110.

[53] Jonas Haferkamp and Nicholas Hunter-Jones. Improved spectral gaps for random quantum circuits: Large local dimensions and all-to-all interactions. *Phys. Rev. A*, 104:022417, 8 2021. URL: https://link.aps.org/doi/10.1103/PhysRevA.104.022417, doi:10.1103/PhysRevA.104.022417.

[54] Thomas Schuster, Jonas Haferkamp, and Hsin-Yuan Huang. Random unitaries in extremely low depth, 2025. URL: https://arxiv.org/abs/2407.07754, arXiv:2407.07754.

[55] Fernando G.S.L. Brandão, Wissam Chemissany, Nicholas Hunter-Jones, Richard Kueng, and John Preskill. Models of quantum complexity growth. *PRX Quantum*, 2:030316, 7 2021. URL: https://link.aps.org/doi/10.1103/PRXQuantum.2.030316, doi:10.1103/PRXQuantum.2.030316.

[56] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and et al. Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644, 8 2021. URL: http://dx.doi.org/10.1038/s42254-021-00348-9, doi:10.1038/s42254-021-00348-9.

[57] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm, 2014. URL: https://arxiv.org/abs/1411.4028, arXiv:1411.4028.

[58] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Leo Zhou. The quantum approximate optimization algorithm and the sherrington-kirkpatrick model at infinite size. 2019. URL: https://arxiv.org/abs/1910.08187, arXiv:1910.08187.

[59] Matthew P. Harrigan, Kevin J. Sung, Matthew Neeley, Kevin J. Satzinger, Frank Arute, Kunal Arya, Juan Atalaya, Joseph C. Bardin, Rami Barends, Sergio Boixo, Michael Broughton, Bob B. Buckley, David A. Buell, Brian Burkett, Nicholas Bushnell, Yu Chen, Zijun Chen, Ben Chiaro, Roberto Collins, William Courtney, Sean Demura, Andrew Dunsworth, Daniel Eppens, Austin Fowler, Brooks Foxen, Craig Gidney, Marissa Giustina, Rob Graff, Steve Habegger, Alan Ho, Sabrina Hong, Trent Huang, L. B. Ioffe,

Sergei V. Isakov, Evan Jeffrey, Zhang Jiang, Cody Jones, Dvir Kafri, Kostyantyn Kechedzhi, Julian Kelly, Seon Kim, Paul V. Klimov, Alexander N. Korotkov, Fedor Kostritsa, David Landhuis, Pavel Laptev, Mike Lindmark, Martin Leib, Orion Martin, John M. Martinis, Jarrod R. McClean, Matt McEwen, Anthony Megrant, Xiao Mi, Masoud Mohseni, Wojciech Mruczkiewicz, Josh Mutus, Ofer Naaman, Charles Neill, Florian Neukart, Murphy Yuezhen Niu, Thomas E. O'Brien, Bryan O'Gorman, Eric Ostby, Andre Petukhov, Harald Putterman, Chris Quintana, Pedram Roushan, Nicholas C. Rubin, Daniel Sank, Andrea Skolik, Vadim Smelyanskiy, Doug Strain, Michael Streif, Marco Szalay, Amit Vainsencher, Theodore White, Z. Jamie Yao, Ping Yeh, Adam Zalcman, Leo Zhou, Hartmut Neven, Dave Bacon, Erik Lucero, Edward Farhi, and Ryan Babbush. Quantum approximate optimization of non-planar graph problems on a planar superconducting processor. *Nature Physics*, 17(3):332–336, feb 2021. URL: https://doi.org/10.1038%2Fs41567-020-01105-y, doi:10.1038/s41567-020-01105-y.

[60] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5(1), 7 2014. URL: http://dx.doi.org/10.1038/ncomms5213, doi:10.1038/ncomms5213.

[61] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. , 549:242–246, 9 2017. arXiv:1704.05018, doi:10.1038/nature23879.

[62] Robert M. Parrish, Edward G. Hohenstein, Peter L. McMahon, and Todd J. Martínez. Quantum computation of electronic transitions using a variational quantum eigensolver. *Phys. Rev. Lett.*, 122:230401, 6 2019. URL: https://link.aps.org/doi/10.1103/PhysRevLett.122.230401, doi:10.1103/PhysRevLett.122.230401.

[63] Cheng Xue, Zhao-Yun Chen, Yu-Chun Wu, and Guo-Ping Guo. Effects of quantum noise on quantum approximate optimization algorithm. *Chinese Physics Letters*, 38(3):030302, 3 2021. doi:10.1088/0256-307x/38/3/030302.

[64] Jeffrey Marshall, Filip Wudarski, Stuart Hadfield, and Tad Hogg. Characterizing local noise in QAOA circuits. *IOP SciNotes*, 1(2):025208, 8 2020. doi:10.1088/2633-1357/abb0d7.

[65] Filip B. Maciejewski, Flavio Baccari, Zoltán Zimborás, and Michał Oszmaniec. Modeling and mitigation of cross-talk effects in readout

noise with applications to the quantum approximate optimization algorithm. *Quantum*, 5:464, 6 2021. URL: http://dx.doi.org/10.22331/q-2021-06-01-464, doi:10.22331/q-2021-06-01-464.

[66] Daniel Stilck França and Raul García-Patrón. Limitations of optimization algorithms on noisy quantum devices. *Nature Physics*, 17(11):1221–1227, Nov 2021. doi:10.1038/s41567-021-01356-3.

[67] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9:4812, 11 2018. arXiv:1803.11173, doi:10.1038/s41467-018-07090-4.

[68] C.A. Fuchs and J. van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Transactions on Information Theory*, 45(4):1216–1227, 1999. doi:10.1109/18.761271.

[69] Ingemar Bengtsson and Karol Zyczkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press, 2006. doi:10.1017/CBO9780511535048.

[70] William Matthews, Stephanie Wehner, and Andreas Winter. Distinguishability of quantum states under restricted families of measurements with an application to quantum data hiding. *Communications in Mathematical Physics*, 291(3):813–843, 11 2009. doi:10.1007/s00220-009-0890-5.

[71] Cécilia Lancien and Andreas Winter. Distinguishing multi-partite states by local measurements. *Communications in Mathematical Physics*, 323(2):555–573, Oct 2013. doi:10.1007/s00220-013-1779-x.

[72] Giacomo De Palma, Milad Marvian, Dario Trevisan, and Seth Lloyd. The quantum wasserstein distance of order 1. *IEEE Transactions on Information Theory*, 67(10):6627–6643, 10 2021. URL: http://dx.doi.org/10.1109/TIT.2021.3076442, doi:10.1109/tit.2021.3076442.

[73] Shunlong Luo and Qiang Zhang. Informational distance on quantum-state space. *Phys. Rev. A*, 69:032106, 3 2004. URL: https://link.aps.org/doi/10.1103/PhysRevA.69.032106, doi:10.1103/PhysRevA.69.032106.

[74] A. Acín. Statistical distinguishability between unitary operations. *Physical Review Letters*, 87(17), 10 2001. URL: http://dx.doi.org/10.1103/PhysRevLett.87.177901, doi:10.1103/physrevlett.87.177901.

[75] Mark D. Bowdrey, Daniel K.L. Oi, Anthony J. Short, Konrad Banaszek, and Jonathan A. Jones. Fidelity of single qubit maps. *Physics Letters A*, 294(5-6):258–260, 3 2002. URL: http://dx.doi.org/10.1016/S0375-9601(02)00069-5, doi:10.1016/s0375-9601(02)00069-5.

[76] Michael A Nielsen. A simple formula for the average gate fidelity of a quantum dynamical operation. *Physics Letters A*, 303(4):249–252, 10 2002. URL: http://dx.doi.org/10.1016/S0375-9601(02)01272-0, doi:10.1016/s0375-9601(02)01272-0.

[77] Benjamin Schumacher. Sending entanglement through noisy quantum channels. *Phys. Rev. A*, 54:2614–2628, 10 1996. URL: https://link.aps.org/doi/10.1103/PhysRevA.54.2614, doi:10.1103/PhysRevA.54.2614.

[78] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. General teleportation channel, singlet fraction, and quasidistillation. *Phys. Rev. A*, 60:1888–1898, 9 1999. URL: https://link.aps.org/doi/10.1103/PhysRevA.60.1888, doi:10.1103/PhysRevA.60.1888.

[79] Yoshifumi Nakata, Da Zhao, Takayuki Okuda, Eiichi Bannai, Yasunari Suzuki, Shiro Tamiya, Kentaro Heya, Zhiguang Yan, Kun Zuo, Shuhei Tamate, Yutaka Tabuchi, and Yasunobu Nakamura. Quantum circuits for exact unitary t-designs and applications to higher-order randomized benchmarking. *PRX Quantum*, 2:030339, 9 2021. URL: https://link.aps.org/doi/10.1103/PRXQuantum.2.030339, doi:10.1103/PRXQuantum.2.030339.

[80] J. S. Lundeen, A. Feito, H. Coldenstrodt-Ronge, K. L. Pregnell, Ch. Silberhorn, T. C. Ralph, J. Eisert, M. B. Plenio, and I. A. Walmsley. Tomography of quantum detectors. *Nature Physics*, 5:27, 11 2008. URL: http://dx.doi.org/10.1038/nphys1133.

[81] Lijian Zhang, Animesh Datta, Hendrik B Coldenstrodt-Ronge, Xian-Min Jin, Jens Eisert, Martin B Plenio, and Ian A Walmsley. Recursive quantum detector tomography. *New Journal of Physics*, 14(11):115005, 11 2012. doi:10.1088/1367-2630/14/11/115005.

[82] Mamoru Endo, Tatsuki Sonoyama, Mikihisa Matsuyama, Fumiya Okamoto, Shigehito Miki, Masahiro Yabuno, Fumihiro China, Hirotaka Terai, and Akira Furusawa. Quantum detector tomography of a superconducting nanostrip photon-number-resolving detector. *Opt. Express*, 29(8):11728–11738, 4 2021. URL: http://www.osapublishing.org/oe/abstract.cfm?URI=oe-29-8-11728, doi:10.1364/OE.423142.

[83] Scott Aaronson. Shadow tomography of quantum states, 2018. arXiv:1711.01053.

[84] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, 6 2020. URL: http://dx.doi.org/10.1038/s41567-020-0932-7, doi:10.1038/s41567-020-0932-7.

[85] Charles Hadfield, Sergey Bravyi, Rudy Raymond, and Antonio Mezzacapo. Measurements of quantum hamiltonians with locally-biased classical shadows, 2020. arXiv:2006.15788.

[86] Senrui Chen, Wenjun Yu, Pei Zeng, and Steven T. Flammia. Robust shadow estimation. *PRX Quantum*, 2(3), 9 2021. URL: http://dx.doi.org/10.1103/PRXQuantum.2.030348, doi:10.1103/prxquantum.2.030348.

[87] Charles Hadfield. Adaptive pauli shadows for energy estimation, 2021. arXiv:2105.12207.

[88] Joseph Emerson, Robert Alicki, and Karol Życzkowski. Scalable noise estimation with random unitary operators. *Journal of Optics B: Quantum and Semiclassical Optics*, 7(10):S347–S352, 9 2005. URL: http://dx.doi.org/10.1088/1464-4266/7/10/021, doi:10.1088/1464-4266/7/10/021.

[89] Easwar Magesan, J. M. Gambetta, and Joseph Emerson. Scalable and robust randomized benchmarking of quantum processes. *Physical Review Letters*, 106(18), 5 2011. URL: http://dx.doi.org/10.1103/PhysRevLett.106.180504, doi:10.1103/physrevlett.106.180504.

[90] Easwar Magesan, Jay M. Gambetta, B. R. Johnson, Colm A. Ryan, Jerry M. Chow, Seth T. Merkel, Marcus P. da Silva, George A. Keefe, Mary B. Rothwell, Thomas A. Ohki, and et al. Efficient measurement of quantum gate error by interleaved randomized benchmarking. *Physical Review Letters*, 109(8), 8 2012. URL: http://dx.doi.org/10.1103/PhysRevLett.109.080505, doi:10.1103/physrevlett.109.080505.

[91] Jay M. Gambetta, A. D. Córcoles, S. T. Merkel, B. R. Johnson, John A. Smolin, Jerry M. Chow, Colm A. Ryan, Chad Rigetti, S. Poletto, Thomas A. Ohki, and et al. Characterization of addressability by simultaneous randomized benchmarking. *Physical Review Letters*, 109(24), 12 2012. URL: http://dx.doi.org/10.1103/PhysRevLett.109.240504, doi:10.1103/physrevlett.109.240504.

[92] Jonas Helsen, Xiao Xue, Lieven M. K. Vandersypen, and Stephanie Wehner. A new class of efficient randomized benchmarking protocols, 2019. arXiv:1806.02048.

[93] Steven T. Flammia. Averaged circuit eigenvalue sampling, 2021. arXiv:2108.05803.

[94] Jonas Helsen, Marios Ioannou, Ingo Roth, Jonas Kitzinger, Emilio Onorati, Albert H. Werner, and Jens Eisert. Estimating gate-set properties from random sequences. *arXiv e-prints*, page arXiv:2110.13178, 10 2021. arXiv:2110.13178.

[95] Jaikumar Radhakrishnan, Martin Rötteler, and Pranab Sen. Random measurement bases, quantum state distinction and applications to the hidden subgroup problem. *Algorithmica*, 55(3):490–516, Nov 2009. doi:10.1007/s00453-008-9231-x.

[96] Jinhyoung Lee, M. S. Kim, and Časlav Brukner. Operationally invariant measure of the distance between quantum states by complementary measurements. *Physical Review Letters*, 91(8), 8 2003. URL: http://dx.doi.org/10.1103/PhysRevLett.91.087902, doi:10.1103/physrevlett.91.087902.

[97] Stephen Brierley, Stefan Weigert, and Ingemar Bengtsson. All mutually unbiased bases in dimensions two to five, 2010. URL: https://arxiv.org/abs/0907.4097, arXiv:0907.4097.

[98] Qingling Zhu, Sirui Cao, Fusheng Chen, Ming-Cheng Chen, Xiawei Chen, Tung-Hsun Chung, Hui Deng, Yajie Du, Daojin Fan, Ming Gong, et al. Quantum computational advantage via 60-qubit 24-cycle random circuit sampling. *arXiv preprint arXiv:2109.03494*, 2021.

[99] Dorit Aharonov, Jordan Cotler, and Xiao-Liang Qi. Quantum algorithmic measurement. *arXiv e-prints*, page arXiv:2101.04634, 1 2021. arXiv:2101.04634.

[100] Michał Oszmaniec, Leonardo Guerini, Peter Wittek, and Antonio Acín. Simulating positive-operator-valued measures with projective measurements. *Phys. Rev. Lett.*, 119:190501, 11 2017. URL: https://link.aps.org/doi/10.1103/PhysRevLett.119.190501, doi:10.1103/PhysRevLett.119.190501.

[101] John Preskill. Lecture notes for physics 219/computer science 219: Quantum computation. http://theory.caltech.edu/~preskill/ph229/. California Institute of Technology. Accessed: 2025-01, year = 1998.

[102] J. L. W. V. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30(none):175 – 193, 1906. doi:10.1007/BF02418571.

[103] Bonnie Berger. The fourth moment method. *SIAM Journal on Computing*, 26(4):1188–1207, 1997. arXiv:https://doi.org/10.1137/S0097539792240005, doi:10.1137/S0097539792240005.

[104] Aram W. Harrow. The church of the symmetric subspace, 2013. arXiv:1308.6595.

[105] Richard Kueng, Huangjun Zhu, and David Gross. Distinguishing quantum states using clifford orbits. *arXiv e-prints*, page arXiv:1609.08595, 9 2016. arXiv:1609.08595.

[106] Aram Harrow and Saeed Mehraban. Approximate unitary t-designs by short random quantum circuits using nearest-neighbor and long-range gates. *arXiv e-prints*, page arXiv:1809.06957, 9 2018. arXiv:1809.06957.

[107] Sergio Boixo, Sergei V. Isakov, Vadim N. Smelyanskiy, Ryan Babbush, Nan Ding, Zhang Jiang, Michael J. Bremner, John M. Martinis, and Hartmut Neven. Characterizing quantum supremacy in near-term devices. *Nature Physics*, 14(6):595–600, 6 2018. doi:10.1038/s41567-018-0124-x.

[108] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Information-theoretic bounds on quantum advantage in machine learning. , 126(19):190505, 5 2021. arXiv:2101.02464, doi:10.1103/PhysRevLett.126.190505.

[109] Pierre Hansen and Brigitte Jaumard. Algorithms for the maximum satisfiability problem. *Computing*, 44(4):279–303, 12 1990. doi:10.1007/BF02241270.

[110] Guillermo García-Pérez, Matteo A.C. Rossi, Boris Sokolov, Francesco Tacchino, Panagiotis Kl. Barkoutsos, Guglielmo Mazzola, Ivano Tavernelli, and Sabrina Maniscalco. Learning to measure: Adaptive informationally complete generalized measurements for quantum algorithms. *PRX Quantum*, 2(4), 11 2021. URL: http://dx.doi.org/10.1103/PRXQuantum.2.040342, doi:10.1103/prxquantum.2.040342.

[111] Alexei Gilchrist, Nathan K. Langford, and Michael A. Nielsen. Distance measures to compare real and ideal quantum processes. *Physical Review A*, 71(6), 6 2005. URL: http://dx.doi.org/10.1103/PhysRevA.71.062310, doi:10.1103/physreva.71.062310.

[112] Peter M Alberti and Armin Uhlmann. *Stochasticity and partial order*. Deutscher Verlag der Wissenschaften Berlin, 1982.

[113] Zbigniew Puchała, Łukasz Rudnicki, and Karol Życzkowski. Majorization entropic uncertainty relations. *Journal of Physics A: Mathematical and Theoretical*, 46(27):272002, 2013.

[114] Francesco Buscemi, Michael Keyl, Giacomo Mauro D'Ariano, Paolo Perinotti, and Reinhard F. Werner. Clean positive operator valued measures. *Journal of Mathematical Physics*, 46(8):082109, 8 2005. arXiv:quant-ph/0505095, doi:10.1063/1.2008996.

[115] Wojciech Roga, Zbigniew Puchała, Łukasz Rudnicki, and Karol Życzkowski. Entropic trade-off relations for quantum operations. *Physical Review A*, 87(3):032308, 2013.

[116] Anna Jenčová and Martin Plávala. Conditions for optimal input states for discrimination of quantum channels. *Journal of Mathematical Physics*, 57(12):122203, 12 2016. URL: https://doi.org/10.1063%2F1.4972286, doi:10.1063/1.4972286.

[117] Ion Nechita, Zbigniew Puchała, Łukasz Pawela, and Karol Życzkowski. Almost all quantum channels are equidistant. *Journal of Mathematical Physics*, 59(5):052201, 2018.

[118] Zbigniew Puchała, Łukasz Pawela, Aleksandra Krawiec, and Ryszard Kukulski. Strategies for optimal single-shot discrimination of quantum measurements. *Physical Review A*, 98(4), 10 2018. URL: http://dx.doi.org/10.1103/PhysRevA.98.042103, doi:10.1103/physreva.98.042103.

[119] Michael R Geller and Mingyu Sun. Toward efficient correction of multiqubit measurement errors: pair correlation method. *Quantum Science and Technology*, 6(2):025009, 2 2021. doi:10.1088/2058-9565/abd5c9.

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

[121] Michał Oszmaniec, Filip B. Maciejewski, and Zbigniew Puchała. Simulating all quantum measurements using only projective measurements and postselection. *Phys. Rev. A*, 100:012351, 7 2019. URL: https://link.aps.org/doi/10.1103/PhysRevA.100.012351, doi:10.1103/PhysRevA.100.012351.

[122] Ewout van den Berg, Zlatko K. Minev, and Kristan Temme. Model-free readout-error mitigation for quantum expectation values. *Physical Review A*, 105(3), 3 2022. URL: https://doi.org/10.1103%2Fphysreva.105.032620, doi:10.1103/physreva.105.032620.

[123] Alistair W. R. Smith, Kiran E. Khosla, Chris N. Self, and M. S. Kim. Qubit readout error mitigation with bit-flip averaging. *Science Advances*, 7(47), 11 2021. URL: https://doi.org/10.1126%2Fsciadv.abi8009, doi:10.1126/sciadv.abi8009.

[124] Shuanghong Tang, Congcong Zheng, and Kun Wang. Detecting and eliminating quantum noise of quantum measurements, 2022. URL: https://arxiv.org/abs/2206.13743, doi:10.48550/ARXIV.2206.13743.

[125] Frank Arute et al. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, 10 2019. doi:10.1038/s41586-019-1666-5.

[126] Benjamin Villalonga, Dmitry Lyakh, Sergio Boixo, Hartmut Neven, Travis S Humble, Rupak Biswas, Eleanor G Rieffel, Alan Ho, and Salvatore Mandrà. Establishing the quantum supremacy frontier with a 281 pflop/s simulation. *Quantum Science and Technology*, 5(3):034003, 4 2020. URL: http://dx.doi.org/10.1088/2058-9565/ab7eeb, doi:10.1088/2058-9565/ab7eeb.

[127] Edward Farhi and Aram W Harrow. Quantum supremacy through the quantum approximate optimization algorithm. *arXiv e-prints*, page arXiv:1602.07674, 2 2016. URL: https://arxiv.org/abs/1602.07674, arXiv:1602.07674.

[128] Nikolaj Moll, Panagiotis Barkoutsos, Lev S. Bishop, Jerry M. Chow, Andrew Cross, Daniel J. Egger, Stefan Filipp, Andreas Fuhrer, Jay M. Gambetta, Marc Ganzhorn, Abhinav Kandala, Antonio Mezzacapo, Peter Müller, Walter Riess, Gian Salis, John Smolin, Ivano Tavernelli, and Kristan Temme. Quantum optimization using variational algorithms on near-term quantum devices. *Quantum Science and Technology*, 3(3):030503, 7 2018. URL: https://arxiv.org/abs/1710.01022v2, arXiv:1710.01022, doi:10.1088/2058-9565/aab822.

[129] John Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, 2:79, 8 2018. doi:10.22331/q-2018-08-06-79.

[130] Joel J. Wallman and Joseph Emerson. Noise tailoring for scalable quantum computation via randomized compiling. *Physical Review A*, 94(5), 11 2016. URL: http://dx.doi.org/10.1103/PhysRevA.94.052325, doi:10.1103/physreva.94.052325.

[131] Ying Li and Simon C. Benjamin. Efficient variational quantum simulator incorporating active error minimization. *Physical Review X*, 7(2), 6 2017. URL: http://dx.doi.org/10.1103/PhysRevX.7.021050, doi:10.1103/physrevx.7.021050.

[132] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. Error mitigation for short-depth quantum circuits. *Physical Review Letters*, 119(18), 11 2017. URL: http://dx.doi.org/10.1103/PhysRevLett.119.180509, doi:10.1103/physrevlett.119.180509.

[133] Suguru Endo, Simon C. Benjamin, and Ying Li. Practical quantum error mitigation for near-future applications. *Physical Review X*, 8:031027, 7 2018. arXiv:1712.09271, doi:10.1103/PhysRevX.8.031027.

[134] Abhinav Kandala, Kristan Temme, Antonio D. Córcoles, Antonio Mezzacapo, Jerry M. Chow, and Jay M. Gambetta. Error mitigation extends the computational reach of a noisy quantum processor. *Nature*, 567(7749):491–495, 3 2019. URL: http://dx.doi.org/10.1038/s41586-019-1040-7, doi:10.1038/s41586-019-1040-7.

[135] Jinzhao Sun, Xiao Yuan, Takahiro Tsunoda, Vlatko Vedral, Simon C. Benjamin, and Suguru Endo. Mitigating realistic noise in practical noisy intermediate-scale quantum devices. *Physical Review Applied*, 15(3), 3 2021. URL: http://dx.doi.org/10.1103/PhysRevApplied.15.034026, doi:10.1103/physrevapplied.15.034026.

[136] William J. Huggins, Sam McArdle, Thomas E. O'Brien, Joonho Lee, Nicholas C. Rubin, Sergio Boixo, K. Birgitta Whaley, Ryan Babbush, and Jarrod R. McClean. Virtual distillation for quantum error mitigation. 2020. URL: https://arxiv.org/abs/2011.07064, arXiv:2011.07064.

[137] Michael R Geller. Rigorous measurement error correction. *Quantum Science and Technology*, 5(3):03LT01, 6 2020. URL: http://dx.doi.org/10.1088/2058-9565/ab9591, doi:10.1088/2058-9565/ab9591.

[138] Benjamin Nachman, Miroslav Urbanek, Wibe A. de Jong, and Christian W. Bauer. Unfolding quantum computer readout noise. *npj Quantum Information*, 6(1):84, Sep 2020. doi:10.1038/s41534-020-00309-7.

[139] Hyeokjea Kwon and Joonwoo Bae. A hybrid quantum-classical approach to mitigating measurement errors in quantum algorithms. *IEEE Transactions on Computers*, page 1–1, 2020. URL: http://dx.doi.org/10.1109/TC.2020.3009664, doi:10.1109/tc.2020.3009664.

[140] Kathleen E. Hamilton, Tyler Kharazi, Titus Morris, Alexander J. McCaskey, Ryan S. Bennink, and Raphael C. Pooser. Scalable quantum processor noise characterization. In *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)*, pages 430–440, 2020. URL: https://arxiv.org/abs/2006.01805, doi:10.1109/QCE49297.2020.00060.

[141] Megan L. Dahlhauser and Travis S. Humble. Modeling noisy quantum circuits using experimental characterization. *Phys. Rev. A*, 103:042603, 4 2021. URL: https://link.aps.org/doi/10.1103/PhysRevA.103.042603, doi:10.1103/PhysRevA.103.042603.

[142] Lena Funcke, Tobias Hartung, Karl Jansen, Stefan Kühn, Paolo Stornati, and Xiaoyang Wang. Measurement error mitigation in quantum computers through classical bit-flip correction. 2020. URL: https://arxiv.org/abs/2007.03663, arXiv:2007.03663.

[143] Muqing Zheng, Ang Li, Tamás Terlaky, and Xiu Yang. A bayesian approach for characterizing and mitigating gate and measurement errors. 2020. URL: https://arxiv.org/abs/2010.09188, arXiv:2010.09188.

[144] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. 2014. URL: https://arxiv.org/abs/1411.4028, arXiv:1411.4028.

[145] M Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and P. J. Coles. Variational quantum algorithms. 2020. URL: https://arxiv.org/abs/2012.09265, arXiv:2012.09265.

[146] Stuart Hadfield, Zhihui Wang, Bryan O'Gorman, Eleanor Rieffel, Davide Venturelli, and Rupak Biswas. From the quantum approximate optimization algorithm to a quantum alternating operator ansatz. *Algorithms*, 12(2):34, 2 2019. URL: http://dx.doi.org/10.3390/a12020034, doi:10.3390/a12020034.

[147] Seth Lloyd. Quantum approximate optimization is computationally universal. 2018. URL: https://arxiv.org/abs/1812.11075, arXiv:1812.11075.

[148] M. E. S. Morales, J. D. Biamonte, and Z. Zimborás. On the universality of the quantum approximate optimization algorithm. *Quantum Information Processing*, 19(9):291, 8 2020. doi:10.1007/s11128-020-02748-9.

[149] G. G. Guerreschi and A. Y. Matsuura. QAOA for max-cut requires hundreds of qubits for quantum speed-up. *Scientific Reports*, 9(1):6903, 5 2019. doi:10.1038/s41598-019-43176-9.

[150] Dmitry Panchenko. The sherrington-kirkpatrick model: An overview. *Journal of Statistical Physics*, 149(2):362–383, 9 2012. URL: http://dx.doi.org/10.1007/s10955-012-0586-7, doi:10.1007/s10955-012-0586-7.

[151] Jordan Cotler and Frank Wilczek. Quantum overlapping tomography. *Physical Review Letters*, 124(10), 3 2020. URL: http://dx.doi.org/10.1103/PhysRevLett.124.100401, doi:10.1103/physrevlett.124.100401.

[152] Jens Koch, Terri M. Yu, Jay Gambetta, A. A. Houck, D. I. Schuster, J. Majer, Alexandre Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf. Charge-insensitive qubit design derived from the cooper pair box. *Phys. Rev. A*, 76:042319, 10 2007. URL: https://link.aps.org/doi/10.1103/PhysRevA.76.042319, doi:10.1103/PhysRevA.76.042319.

[153] Cheng Xue, Zhao-Yun Chen, Yu-Chun Wu, and Guo-Ping Guo. Effects of quantum noise on quantum approximate optimization algorithm. *Chinese Physics Letters*, 38(3):030302, 3 2021. doi:10.1088/0256-307x/38/3/030302.

[154] Jeffrey Marshall, Filip Wudarski, Stuart Hadfield, and Tad Hogg. Characterizing local noise in QAOA circuits. *IOP SciNotes*, 1(2):025208, 8 2020. URL: http://dx.doi.org/10.1088/2633-1357/abb0d7, doi:10.1088/2633-1357/abb0d7.

[155] Mahabubul Alam, Abdullah Ash-Saki, and Swaroop Ghosh. Analysis of quantum approximate optimization algorithm under realistic noise in superconducting qubits. 2019. URL: https://arxiv.org/abs/1907.09631, arXiv:1907.09631.

[156] Ashley Montanaro and Stasja Stanisic. Compressed variational quantum eigensolver for the fermi-hubbard model. 2020. URL: https://arxiv.org/abs/2006.01179, arXiv:2006.01179.

[157] Pranav Gokhale, Ali Javadi-Abhari, Nathan Earnest, Yunong Shi, and Frederic T. Chong. Optimized quantum compilation for near-term algorithms with OpenPulse. 2020. URL: https://arxiv.org/abs/2004.11205, arXiv:2004.11205.

[158] Zdeněk Hradil, Jaroslav Řeháček, Jaromír Fiurášek, and Miroslav Ježek. *3 Maximum-Likelihood Methodsin Quantum Mechanics*, pages 59–112. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004. doi:10.1007/978-3-540-44481-7_3.

[159] Jaromír Fiurášek. Maximum-likelihood estimation of quantum measurement. *Physical Review A*, 64:024102, 8 2001. arXiv:quant-ph/0101027, doi:10.1103/PhysRevA.64.024102.

[160] I. Gianani, Y.S. Teo, V. Cimini, H. Jeong, G. Leuchs, M. Barbieri, and L.L. Sánchez-Soto. Compressively certifying quantum measurements. *PRX Quantum*, 1(2), 10 2020. URL: http://dx.doi.org/10.1103/PRXQuantum.1.020307, doi:10.1103/prxquantum.1.020307.

[161] Tim J. Evans, Robin Harper, and Steven T. Flammia. Scalable bayesian hamiltonian learning. 2019. URL: https://arxiv.org/abs/1912.07636, arXiv:1912.07636.

[162] Nengkun Yu. Sample efficient tomography via Pauli measurements. 2020. URL: https://arxiv.org/abs/2009.04610, arXiv:2009.04610.

[163] B. S. Majewski, N. C. Wormald, G. Havas, and Z. J. Czech. A family of perfect hashing methods. *The Computer Journal*, 39(6):547–554, 01 1996. doi:10.1093/comjnl/39.6.547.

[164] D. R. Stinson, R. Wei, and L. Zhu. New constructions for perfect hash families and related structures using combinatorial designs and codes. *Journal of Combinatorial Designs*, 8(3):189–200, 2000. doi:10.1002/(SICI)1520-6610(2000)8:3<189::AID-JCD4>3.0.CO;2-A.

[165] Simon R. Blackburn. Perfect hash families: Probabilistic methods and explicit constructions. *Journal of Combinatorial Theory, Series A*, 92(1):54 – 60, 2000. URL: https://www.sciencedirect.com/science/article/pii/S0097316599930509, doi:10.1006/jcta.1999.3050.

[166] Noga Alon and Shai Gutner. Balanced families of perfect hash functions and their applications. 5 2008. URL: https://arxiv.org/abs/0805.4300.

[167] Pierre Hansen and Brigitte Jaumard. Algorithms for the maximum satisfiability problem. *Computing*, 44(4):279–303, 1990.

[168] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. Quantum computation by adiabatic evolution. *arXiv preprint quant-ph/0001106*, 2000.

[169] Siddhartha Santra, Gregory Quiroz, Greg Ver Steeg, and Daniel A Lidar. Max 2-SAT with up to 108 qubits. *New Journal of Physics*, 16(4):045006, 4 2014. doi:10.1088/1367-2630/16/4/045006.

[170] Kenneth Rudinger, Timothy Proctor, Dylan Langharst, Mohan Sarovar, Kevin Young, and Robin Blume-Kohout. Probing context-dependent errors in quantum processors. *Physical Review X*, 9(2), 6 2019. URL: http://dx.doi.org/10.1103/PhysRevX.9.021045, doi:10.1103/physrevx.9.021045.

[171] Edward Farhi, David Gamarnik, and Sam Gutmann. The quantum approximate optimization algorithm needs to see the whole graph: A typical case. 2020. URL: https://arxiv.org/abs/2004.09002, arXiv:2004.09002.

[172] Sandu Popescu, Anthony J. Short, and Andreas Winter. Entanglement and the foundations of statistical mechanics. *Nature Physics*, 2(11):754–758, 11 2006. arXiv:quant-ph/0511225, doi:10.1038/nphys444.

[173] M. Oszmaniec, R. Augusiak, C. Gogolin, J. Kołodyński, A. Acín, and M. Lewenstein. Random bosonic states for robust quantum metrology. *Phys. Rev. X*, 6:041044, 12 2016. URL: https://link.aps.org/doi/10.1103/PhysRevX.6.041044, doi:10.1103/PhysRevX.6.041044.

[174] Fernando G. S. L. Brandão, Aram W. Harrow, and Michał Horodecki. Local random quantum circuits are approximate polynomial-designs. *Communications in Mathematical Physics*, 346(2):397–434, 9 2016. arXiv:1208.0692, doi:10.1007/s00220-016-2706-8.

[175] Jordan Cotler, Nicholas Hunter-Jones, and Daniel Ranard. Fluctuations of subsystem entropies at late times. 10 2020. URL: https://arxiv.org/abs/2010.11922.

[176] David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. *Physical review letters*, 35(26):1792, 1975.

[177] J. Spall. AN OVERVIEW OF THE SIMULTANEOUS PERTURBATION METHOD FOR EFFICIENT OPTIMIZATION. *Johns Hopkins Apl Technical Digest*, 19:482–492, 1998. URL: https://www.jhuapl.edu/Content/techdigest/pdf/V19-N04/19-04-Spall.pdf.

[178] Chris Cade, Lana Mineh, Ashley Montanaro, and Stasja Stanisic. Strategies for solving the fermi-hubbard model on near-term quantum computers. *Phys. Rev. B*, 102:235122, 12 2020. URL: https://link.aps.org/doi/10.1103/PhysRevB.102.235122, doi:10.1103/PhysRevB.102.235122.

[179] Bas Dirkse, Jonas Helsen, and Stephanie Wehner. Efficient unitarity randomized benchmarking of few-qubit clifford gates. *Physical Review A*, 99(1), 1 2019. URL: http://dx.doi.org/10.1103/PhysRevA.99.012315, doi:10.1103/physreva.99.012315.

[180] Matthew Girling, Cristina Cîrstoiu, and David Jennings. Estimation of correlations and nonseparability in quantum channels via unitarity benchmarking. *Phys. Rev. Res.*, 4:023041, 4 2022. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.4.023041, doi:10.1103/PhysRevResearch.4.023041.

[181] Alexey E. Rastegin. Partitioned trace distances. *Quantum Information Processing*, 9(1):61–73, Feb 2010. doi:10.1007/s11128-009-0128-7.

[182] Dénes Petz. Quasi-entropies for finite quantum systems. *Reports on Mathematical Physics*, 23(1):57–65, 1986. URL: https://www.sciencedirect.com/science/article/pii/0034487786900674, doi:10.1016/0034-4877(86)90067-4.

[183] Marcin Jarzyna and Jan Kolodynski. Geometric approach to quantum statistical inference. *IEEE Journal on Selected Areas in Information Theory*, 1(2):367–386, 8 2020. URL: http://dx.doi.org/10.1109/JSAIT.2020.3017469, doi:10.1109/jsait.2020.3017469.

[184] Erik Nielsen, John King Gamble, Kenneth Rudinger, Travis Scholten, Kevin Young, and Robin Blume-Kohout. Gate set tomography. *Quantum*, 5:557, 10 2021. URL: http://dx.doi.org/10.22331/q-2021-10-05-557, doi:10.22331/q-2021-10-05-557.

[185] William J. Huggins, Jarrod R. McClean, Nicholas C. Rubin, Zhang Jiang, Nathan Wiebe, K. Birgitta Whaley, and Ryan Babbush. Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers. *npj Quantum Information*, 7(1), February 2021. URL: http://dx.doi.org/10.1038/s41534-020-00341-7, doi:10.1038/s41534-020-00341-7.

[186] George S. Barron and Christopher J. Wood. Measurement error mitigation for variational quantum algorithms. 2020. URL: https://arxiv.org/abs/2010.08520, arXiv:2010.08520.

[187] Timothy Proctor, Melissa Revelle, Erik Nielsen, Kenneth Rudinger, Daniel Lobser, Peter Maunz, Robin Blume-Kohout, and Kevin Young. Detecting and tracking drift in quantum information processors. *Nature Communications*, 11(1):5396, 10 2020. doi:10.1038/s41467-020-19074-4.

[188] Samudra Dasgupta and Travis S. Humble. Characterizing the stability of NISQ devices. In *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)*, pages 419–429, 2020. doi:10.1109/QCE49297.2020.00059.

[189] L. C. G. Govia, G. J. Ribeill, D. Ristè, M. Ware, and H. Krovi. Bootstrapping quantum process tomography via a perturbative ansatz. *Nature Communications*, 11(1):1084, 2 2020. doi:10.1038/s41467-020-14873-1.

[190] Steven T. Flammia and Joel J. Wallman. Efficient estimation of Pauli channels. *ACM Transactions on Quantum Computing*, 1(1):1–32, 12 2020. URL: http://dx.doi.org/10.1145/3408039, doi:10.1145/3408039.

[191] Robin Harper, Steven T. Flammia, and Joel J. Wallman. Efficient learning of quantum noise. *Nature Physics*, 16(12):1184–1188, 8 2020. URL: http://dx.doi.org/10.1038/s41567-020-0992-8, doi:10.1038/s41567-020-0992-8.

[192] Senrui Chen, Wenjun Yu, Pei Zeng, and Steven T. Flammia. Robust shadow estimation. 2020. URL: https://arxiv.org/abs/2011.09636, arXiv:2011.09636.

[193] Kelly Boothby, Paul Bunyk, Jack Raymond, and Aidan Roy. Next-generation topology of D-wave quantum processors. 2020. URL: https://arxiv.org/abs/2003.00133, arXiv:2003.00133.

[194] Colin D. Bruzewicz, John Chiaverini, Robert McConnell, and Jeremy M. Sage. Trapped-ion quantum computing: Progress and challenges. *Applied Physics Reviews*, 6(2):021314, 6 2019. URL: http://dx.doi.org/10.1063/1.5088164, doi:10.1063/1.5088164.

[195] Jianwei Wang, Fabio Sciarrino, Anthony Laing, and Mark G. Thompson. Integrated photonic quantum technologies. *Nature Photonics*, 14(5):273–284, 10 2019. URL: http://dx.doi.org/10.1038/s41566-019-0532-1, doi:10.1038/s41566-019-0532-1.

[196] Tsachy Weissman, Erik Ordentlich, Gadiel Seroussi, Sergio Verdu1, and Marcelo J. Weinberger. Inequalities for the L1 deviation of the empirical distribution. *Technical Report HPL-2003-97R1, Hewlett-Packard*

Labs, 08 2003. URL: https://www.hpl.hp.com/techreports/2003/HPL-2003-97R1.pdf?origin=publicationDetail.

[197] F. B. Maciejewski, T. Rybotycki, and M. Oszmaniec. Quantum readout errors mitigation (QREM) – open source GitHub repository, 2020. URL: https://github.com/fbm2718/QREM.

[198] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. doi:10.1017/CBO9780511976667.

[199] V. Akshay, H. Philathong, M. E. S. Morales, and J. D. Biamonte. Reachability deficits in quantum approximate optimization. *Phys. Rev. Lett.*, 124:090504, 3 2020. URL: https://link.aps.org/doi/10.1103/PhysRevLett.124.090504, doi:10.1103/PhysRevLett.124.090504.

A. Operational Average-Case Distances

Here we provide proofs of more technical results of Chapter 3 – proof of Lemma 5 (Appendix A.1) and proofs of main Theorems 1, 2, 3 for *approximate* 4-designs (Appendix A.2).

A.1 Proof of Lemma 5

In what follows we prove Lemma 5, which we repeat here for Reader's convenience.

Lemma 29 (Repeated Lemma 5). *Let $X, Y \in \mathrm{Herm}(\mathcal{H})$ be Hermitian operators acting on $\mathcal{H} \simeq \mathcal{H}$. Let $\mathbb{P}^{(k)}_{\mathrm{sym}}$ denotes the orthogonal projector onto k-fold symmetrization of $\mathcal{H}^{(k)}_{\mathrm{sym}} \subset \mathcal{H}^{\otimes k}$. We then have the following inequality*

$$\mathrm{tr}\left(X^{\otimes 2} \otimes Y^{\otimes 2}\, \mathbb{P}^{(4)}_{\mathrm{sym}}\right) \leq C\, \mathrm{tr}\left(X^{\otimes 2}\, \mathbb{P}^{(2)}_{\mathrm{sym}}\right) \mathrm{tr}\left(Y^{\otimes 2}\, \mathbb{P}^{(2)}_{\mathrm{sym}}\right), \text{ where } C = \frac{13}{6}. \quad \text{(A.1)}$$

Proof. We begin by noting that, for Hermitian matrices A and B we have

$$\mathrm{tr}\left((A^{\otimes 2} \otimes B^{\otimes 2})(\mathbb{P}^{(2)}_{\mathrm{sym}} \otimes \mathbb{P}^{(2)}_{\mathrm{sym}})\right) = \frac{1}{4}(\mathrm{tr}(A^2) + (\mathrm{tr}(A))^2)(\mathrm{tr}(B^2) + (\mathrm{tr}(B))^2). \quad \text{(A.2)}$$

We also have

$$\begin{aligned} 4!\ \mathrm{tr}\left((A^{\otimes 2}B^{\otimes 2})\mathbb{P}^{(4)}_{\mathrm{sym}}\right) &= ((\mathrm{tr}(A))^2 + \mathrm{tr}(A^2))((\mathrm{tr}(B))^2 + \mathrm{tr}(B^2)) \\ &\quad + 4\,\mathrm{tr}(A)\,\mathrm{tr}(B)\,\mathrm{tr}(AB) + 4\,\mathrm{tr}(A)\,\mathrm{tr}(AB^2) + 4\,\mathrm{tr}(B)\,\mathrm{tr}(A^2B) \\ &\quad + 2(\mathrm{tr}(AB))^2 + 2\,\mathrm{tr}(A^2B^2) + 2\,\mathrm{tr}(ABAB). \end{aligned} \quad \text{(A.3)}$$

Now we consider the following difference for, with arbitrary scalar parameter c

$$\begin{aligned} &c\ \mathrm{tr}\left((A^{\otimes 2} \otimes B^{\otimes 2})(\mathbb{P}^{(2)}_{\mathrm{sym}} \otimes \mathbb{P}^{(2)}_{\mathrm{sym}})\right) - \mathrm{tr}\left((A^{\otimes 2} \otimes B^{\otimes 2})\mathbb{P}^{(4)}_{\mathrm{sym}}\right) = \\ &= \frac{1}{4!}\Big((6c-1)(\mathrm{tr}(A^2) + (\mathrm{tr}(A))^2)(\mathrm{tr}(B^2) + (\mathrm{tr}(B))^2) \\ &- 4\,\mathrm{tr}(A)\,\mathrm{tr}(B)\,\mathrm{tr}(AB) - 4\,\mathrm{tr}(A)\,\mathrm{tr}(AB^2) - 4\,\mathrm{tr}(B)\,\mathrm{tr}(A^2B) \\ &- 2(\mathrm{tr}(AB))^2 - 2\,\mathrm{tr}(A^2B^2) - 2\,\mathrm{tr}(ABAB)\Big). \end{aligned} \quad \text{(A.4)}$$

Now we will bound the terms which occur above using standard inequalities, to get

$$\begin{aligned}
-4\operatorname{tr}(A)\operatorname{tr}(B)\operatorname{tr}(AB) &\geq -4|\operatorname{tr}(A)||\operatorname{tr}(B)|\sqrt{\operatorname{tr}(A^2)}\sqrt{\operatorname{tr}(B^2)} \geq \\
&\geq -2(|\operatorname{tr}(A)|^2|\operatorname{tr}(B)|^2 + \operatorname{tr}(A^2)\operatorname{tr}(B^2)); \\
-4\operatorname{tr}(A)\operatorname{tr}(AB^2) &\geq -4|\operatorname{tr}(A)|\sqrt{\operatorname{tr}(A^2)}\sqrt{\operatorname{tr}(B^4)} \geq -4|\operatorname{tr}(A)|\sqrt{\operatorname{tr}(A^2)}\operatorname{tr}(B^2) \geq \\
&\geq -2(|\operatorname{tr}(A)|^2 + \operatorname{tr}(A^2))\operatorname{tr}(B^2); \\
-4\operatorname{tr}(B)\operatorname{tr}(A^2B) &\geq -4|\operatorname{tr}(B)|\sqrt{\operatorname{tr}(B^2)}\sqrt{\operatorname{tr}(A^4)} \geq -4|\operatorname{tr}(B)|\sqrt{\operatorname{tr}(B^2)}\operatorname{tr}(A^2) \geq \\
&\geq -2(|\operatorname{tr}(B)|^2 + \operatorname{tr}(B^2))\operatorname{tr}(A^2); \\
-2(\operatorname{tr}(AB))^2 &\geq -2\operatorname{tr}(A^2)\operatorname{tr}(B^2); \\
-2\operatorname{tr}(A^2B^2) &\geq -2\operatorname{tr}(A^2)\operatorname{tr}(B^2); \\
-2\operatorname{tr}(ABAB) &\geq -2\operatorname{tr}(A^2)\operatorname{tr}(B^2).
\end{aligned} \tag{A.5}$$

Combining above inequalities, we will determine the value of parameter c, for which (A.4) is non-negative

$$\begin{aligned}
&c\ \operatorname{tr}(A^{\otimes 2}\otimes B^{\otimes 2})(\mathbb{P}^{(2)}_{\mathrm{sym}}\otimes\mathbb{P}^{(2)}_{\mathrm{sym}}) - \operatorname{tr}(A^{\otimes 2}\otimes B^{\otimes 2})\mathbb{P}^{(4)}_{\mathrm{sym}} \geq \\
&= \frac{1}{4!}\Big((6c-1)(\operatorname{tr}A^2 + (\operatorname{tr}A)^2)(\operatorname{tr}B^2 + (\operatorname{tr}B)^2) - 12\operatorname{tr}A^2\operatorname{tr}B^2 \\
&- 2|\operatorname{tr}A|^2|\operatorname{tr}B|^2 - 2|\operatorname{tr}A|^2\operatorname{tr}B^2 - 2|\operatorname{tr}B|^2\operatorname{tr}A^2\Big) \\
&= \frac{1}{4!}\Big((6c-13)\operatorname{tr}A^2\operatorname{tr}B^2 + (6c-3)(\operatorname{tr}A^2(\operatorname{tr}B)^2 + (\operatorname{tr}A)^2\operatorname{tr}B^2 + (\operatorname{tr}A)^2(\operatorname{tr}B)^2)\Big).
\end{aligned} \tag{A.6}$$

Note, that the above is larger than 0 for $c \geq 13/6$. □

A.2 Proofs of main theorems for δ-approximate 4-designs

Here we outline the extension of proofs of Theorems 1, 2, 3 for approximate 4-designs.

A.2.1 Quantum states and measurements

We will start with quantum states. Let us consider δ-approximate 4-design ν (recall Section 2.5), i.e., we have

$$\left\|\mathcal{T}_{4,\nu} - \mathcal{T}_{4,\mu}\right\|_\diamond \leq \delta\,, \tag{A.7}$$

where μ is the Haar measure in $\mathsf{U}(\mathcal{H}_\mathrm{d})$ and $\mathcal{T}_{4,\nu}$ is the quantum channel acting on $\mathcal{H}_\mathrm{d}^{\otimes 4}$ defined as $\mathcal{T}_{4,\nu}(A) = \int_{\mathsf{U}(\mathcal{H}_\mathrm{d})} d\nu(U) U^{\otimes 4} A (U^\dagger)^{\otimes 4}$. For a measure $\nu = \{\nu_\alpha, U_\alpha\}$

on $\mathsf{U}(\mathcal{H}_d)$ let $\tilde{\nu}$ denote a measure supported on 'inverted gates' i.e. $\nu = \{\nu_\alpha, U_\alpha^\dagger\}$ (the generalization to non-discrete measures is straightforward). From the definition of the diamond norm and the identity $\mu = \tilde{\mu}$ it follows that

$$\left\|\mathcal{T}_{k,\nu} - \mathcal{T}_{k,\mu}\right\|_\diamond = \left\|\mathcal{T}_{k,\tilde{\nu}} - \mathcal{T}_{k,\mu}\right\|_\diamond . \tag{A.8}$$

Denote $X_{i,U} = \mathrm{tr}(|i\rangle\langle i| U\Delta U^\dagger)$, where $\Delta = \rho - \sigma$ for two quantum states $\rho, \sigma \in \mathsf{D}(\mathcal{H}_d)$ which we wish to compare. Using Berger's inequality (cf. Lemma 4) for every summand in the expression for the TV distance $\mathsf{TV}(\mathbf{p}^{\rho,U}, \mathbf{p}^{\sigma,U})$, where $U \sim \nu$ we get

$$\mathop{\mathbb{E}}_{U\sim\nu} |X_{i,U}| \geq \frac{\left(\mathop{\mathbb{E}}_{U\sim\nu} X_{i,U}^2\right)^{3/2}}{\left(\mathop{\mathbb{E}}_{U\sim\nu} X_{i,U}^4\right)^{1/2}} . \tag{A.9}$$

Our goal is to compare the right-hand side of the above expression with its counterpart evaluated using the Haar measure μ i.e. : $\left(\mathop{\mathbb{E}}_{U\sim\mu} X_{i,U}^2\right)^{\frac{3}{2}} \left(\mathop{\mathbb{E}}_{U\sim\mu} X_{i,U}^4\right)^{-\frac{1}{2}}$. We begin with the lower bound for the numerator of (A.9).

$$\begin{aligned}
\mathop{\mathbb{E}}_{U\sim\nu} X_{i,U}^2 &\geq \mathop{\mathbb{E}}_{U\sim\mu} X_{i,U}^2 - \left|\mathrm{tr}(\mathcal{T}_{2,\mu} - \mathcal{T}_{2,\tilde{\nu}})[|i\rangle\langle i|^{\otimes 2}]\Delta^{\otimes 2}\right| \\
&\geq \frac{1}{d(d+1)}\mathrm{tr}(\Delta^{\otimes 2}) - \left\|\mathrm{tr}(\mathcal{T}_{2,\mu} - \mathcal{T}_{2,\tilde{\nu}})[|i\rangle\langle i|^{\otimes 2}]\right\|_1 \|\Delta\|_\infty^2 \\
&\geq \frac{1}{d(d+1)}\mathrm{tr}(\Delta^{\otimes 2}) - \delta\|\Delta\|_\infty^2 \\
&\geq \frac{1}{d(d+1)}\mathrm{tr}(\Delta^{\otimes 2})(1 - d(d+1)\delta).
\end{aligned} \tag{A.10}$$

where we used standard inequalities $|\mathrm{tr}(AB)| \leq ||A||_1 \, ||B||_\infty$, $||A||_1 \leq ||A||_\diamond$, $||A||_\infty^2 \leq ||A||_{\mathrm{HS}}^2$, the definition of the diamond norm and (A.7).

Next, we bound denominator from above using Lemma 4 and reasoning analogous as before

$$\begin{aligned}
\mathop{\mathbb{E}}_{U\sim\nu} X_{i,U}^4 &\leq \mathop{\mathbb{E}}_{U\sim\mu} X_{i,U}^4 + \left|\mathrm{tr}(\mathcal{T}_{4,\mu} - \mathcal{T}_{4,\tilde{\nu}})[|i\rangle\langle i|^{\otimes 4}]\Delta^{\otimes 4}\right| \\
&\leq C\left(\mathop{\mathbb{E}}_{U\sim\mu} X_{i,U}^2\right)^2 + \delta\|\Delta\|_\infty^4 \\
&\leq C\frac{\mathrm{tr}(\Delta^2)^2}{(d(d+1))^2}\left(1 + \frac{(d(d+1))^4\delta}{C}\right) ,
\end{aligned} \tag{A.11}$$

with $C = 10.1$.

Combining above inequalities, we obtain that for δ approximate 4-desing, we have

$$\frac{\left(\mathop{\mathbb{E}}_{U\sim\nu} X_{i,U}^2\right)^{3/2}}{\left(\mathop{\mathbb{E}}_{U\sim\nu} X_{i,U}^4\right)^{1/2}} \geq \tilde{\ell}(\delta)\frac{\left(\mathop{\mathbb{E}}_{U\sim\mu} X_{i,U}^2\right)^{3/2}}{\left(\mathop{\mathbb{E}}_{U\sim\mu} X_{i,U}^4\right)^{1/2}} \tag{A.12}$$

with

$$\tilde{\ell}(\delta) = \frac{(1-d(d+1)\delta)^{3/2}}{\left(1+\frac{\delta(d(d+1))^2}{C}\right)^{1/2}} \geq \frac{(1-2d^2\delta)^{3/2}}{\left(1+\frac{4d^4\delta}{C}\right)^{1/2}} \geq \frac{(1-2d^2\delta)^{3/2}}{\left(1+2d^4\delta\right)^{1/2}} \tag{A.13}$$

where we used the fact that $x(x+1) \leq 2x^2$ and $x^2(x+1)^2 \leq 4x^4$ for any $x \geq 1$, and $\frac{2}{C} = \frac{2}{10.1} < 1$.

By setting $\delta = \frac{\delta'}{2d^4}$, we obtain

$$\tilde{l}(\delta') \geq \sqrt{\frac{(1-\frac{\delta'}{d^2})^3}{1+\delta'}} =: \ell(\delta') \,. \tag{A.14}$$

Using analogous reasoning for bounding $\mathop{\mathbb{E}}_{U\sim\nu} X_{i,U}^2$ from above, we obtain upper bound

$$\mathop{\mathbb{E}}_{U\sim\nu} |X_{i,U}| \leq \tilde{u}(\delta) \mathop{\mathbb{E}}_{U\sim\mu} |X_{i,U}| \,, \tag{A.15}$$

with

$$\tilde{u}(\delta) = (1+d(d+1)\delta)^{1/2} \leq (1+2\delta d^2)^{1/2} = (1+\frac{\delta'}{d^2})^{1/2} =: u(\delta'), \tag{A.16}$$

where we used the fact that $x(x+1) \leq 2x^2$ for any $x \geq 1$. This concludes the proof for quantum states.

For quantum measurements, we follow the analogous technique of proof. For POVMs $\mathbf{M}$ and $\mathbf{N}$, each $\Delta_i = M_i - N_i$ will play a role of previous Δ. The only difference will be that the second moment is equal to

$$\mathop{\mathbb{E}}_{U\sim\nu} X_{i,U}^2 = \frac{1}{d(d+1)}(\mathrm{tr}(\Delta_i)^2 + \mathrm{tr}(\Delta_i^2)), \tag{A.17}$$

because operators Δ_i are generally not traceless.

A.2.2 Quantum channels

Let us now proceed to the proof of Theorem 3 for channels. Denote $X_{i,V,U} = \mathrm{tr}(|i\rangle\langle i|\, U\Delta(V\psi_0 V^\dagger)U^\dagger)$ where $\Delta = \Lambda - \Gamma$ with two quantum channels $\Lambda, \Gamma \in \mathrm{CPTP}(\mathcal{H}_d)$ that we are comparing. From the reasoning given in the preceding section (i.e. the proof of Theorem 1 for approximate 4-designs) we have that for $\delta = \delta'/(2d^4)$

$$\ell(\delta')\,\frac{a}{2}\|\Delta[\psi_V]\|_{\mathrm{HS}} \leq \mathop{\mathbb{E}}_{U\sim\nu} \mathrm{TV}(\mathbf{p}^{\Lambda,\psi_V,U}, \mathbf{p}^{\Gamma,\psi_V,U}) \leq u(\delta')\,\frac{A}{2}\|\Delta[\psi_V]\|_{\mathrm{HS}}\,. \tag{A.18}$$

We will use the above bounds together with Jensen's and Berger's inequality (applied for the function $Y_V := \|\Delta[\psi_V]\|_{\mathrm{HS}}$ and $V \sim \nu$) to establish the desired result. We start by re-expressing the second and fourth moment of Y_V in a convenient form :

$$\mathop{\mathbb{E}}_{V\sim\nu} Y_V^2 = \mathop{\mathbb{E}}_{V\sim\nu} \mathrm{tr}\left(\mathbb{S}\Delta^{\otimes 2}[\psi_V^{\otimes 2}]\right) = 2\,\mathrm{tr}\left(\mathbb{P}^{(2)}_{\mathrm{sym}}\Delta^{\otimes 2}[\mathcal{T}_{2,\nu}(\psi_0^{\otimes 2})]\right)\,, \tag{A.19}$$

$$\mathop{\mathbb{E}}_{V\sim\nu} Y_V^4 = \mathop{\mathbb{E}}_{V\sim\nu} \mathrm{tr}\left(\mathbb{S}\Delta^{\otimes 2}[\psi_V^{\otimes 2}]\right)^2 = 4\,\mathrm{tr}\left(\mathbb{P}^{(2)}_{\mathrm{sym}}\otimes\mathbb{P}^{(2)}_{\mathrm{sym}}\Delta^{\otimes 4}[\mathcal{T}_{4,\nu}(\psi_0^{\otimes 4})]\right)\,, \tag{A.20}$$

where we have used the 'swap trick': $\mathrm{tr}(AB) = \mathrm{tr}(A\otimes B\mathbb{S})$, and the fact that $\mathrm{tr}(\Delta[\psi_V]) = 0$. We start by using Eq. (A.19) to derive an *upper bound* on $\mathop{\mathbb{E}}_{V\sim\nu} Y_V^2$. From above for δ-approximate 4-design ν

$$\begin{aligned}
\mathop{\mathbb{E}}_{V\sim\nu} Y_V^2 &\leq \mathop{\mathbb{E}}_{V\sim\mu} Y_V^2 + 2\left|\mathrm{tr}\left(\mathbb{P}^{(2)}_{\mathrm{sym}}\Delta^{\otimes 2}[(\mathcal{T}_{2,\nu}-\mathcal{T}_{2,\mu})(\psi_0^{\otimes 2})]\right)\right| && \text{(A.21)}\\
&= \mathop{\mathbb{E}}_{V\sim\mu} Y_V^2 + 2\left|\mathrm{tr}\left((\Delta^\dagger)^{\otimes 2}\left[\mathbb{P}^{(2)}_{\mathrm{sym}}\right](\mathcal{T}_{2,\nu}-\mathcal{T}_{2,\mu})(\psi_0^{\otimes 2})\right)\right| && \text{(A.22)}\\
&\leq \mathop{\mathbb{E}}_{V\sim\mu} Y_V^2 + 2\left\|(\Delta^\dagger)^{\otimes 2}\left[\mathbb{P}^{(2)}_{\mathrm{sym}}\right]\right\|_\infty \delta\,, && \text{(A.23)}
\end{aligned}$$

where we have used the definition of the dual of a super operator and utilized that δ-approximate 4-design ν is also δ-approximate 2-design. We proceed with bounding the operator norm of $(\Delta^\dagger)^{\otimes 2}\left[\mathbb{P}^{(2)}_{\mathrm{sym}}\right]$ in terms of HS norm of the Jamiołkowski-Choi state $\mathcal{J}_\Delta$:

$$\begin{aligned}
\left\|(\Delta^\dagger)^{\otimes 2}\left[\mathbb{P}^{(2)}_{\mathrm{sym}}\right]\right\|_\infty &= \binom{d+1}{2}\left\|(\Delta^\dagger)^{\otimes 2}\left[\mathop{\mathbb{E}}_{U\sim\mu}\psi_U^{\otimes 2}\right]\right\|_\infty \leq && \text{(A.24)}\\
&\leq \binom{d+1}{2}\max_{\psi\in\mathcal{S}(\mathcal{H})}\left\|(\Delta^\dagger)^{\otimes 2}\left[\psi^{\otimes 2}\right]\right\|_\infty = \binom{d+1}{2}\left(\max_{\psi,\phi\in\mathcal{S}(\mathcal{H})}\mathrm{tr}(\phi\Delta(\psi))\right)^2. && \text{(A.25)}
\end{aligned}$$

The result of double maximization can be upper bounded as follows:

$$\max_{\psi,\phi\in\mathcal{S}(\mathcal{H}_d)}\mathrm{tr}(\phi\Delta(\psi)) = \max_{\psi,\phi\in\mathcal{S}(\mathcal{H}_d)} d\,\mathrm{tr}(\mathcal{J}_\Delta\phi\otimes\psi^T) \leq d\|\mathcal{J}_\Delta\|_{\mathrm{HS}}\,. \tag{A.26}$$

Inserting this to (A.23) and recalling that $\mathop{\mathbb{E}}_{V\sim\mu} Y_V^2 = (A^{\mathrm{ch}})^2(\|\mathcal{J}_\Delta\|_{\mathrm{HS}}^2 + \mathrm{tr}(\Delta(\tau_{\mathrm{d}})^2)$ with $A^{\mathrm{ch}} = \frac{d}{d+1}$(cf. Eq. (A.18)) we get:

$$\mathop{\mathbb{E}}_{V\sim\nu} Y_V^2 \leq (A^{\mathrm{ch}})^2(\|\mathcal{J}_\Delta\|_{\mathrm{HS}}^2+\mathrm{tr}(\Delta(\tau_{\mathrm{d}})^2)+d^5(d+1)\|\mathcal{J}_\Delta\|_{\mathrm{HS}}^2\delta \leq \mathop{\mathbb{E}}_{V\sim\mu} Y_V^2\left(1+\frac{d^5(d+1)}{(A^{\mathrm{ch}})^2}\delta\right). \tag{A.27}$$

Integrating both sides of the upper bound in Eq. (A.18) , using Jensen's inequality, and noting that $\frac{1}{2}\sqrt{\mathop{\mathbb{E}}_{V\sim\mu} Y_V^2} = \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Gamma)$ yields

$$\mathop{\mathbb{E}}_{V\sim\nu}\mathop{\mathbb{E}}_{U\sim\nu}\mathrm{TV}(\mathbf{p}^{\Lambda,\psi_V,U},\mathbf{p}^{\Gamma,\psi_V,U}) \leq u(2d^4\,\delta)\sqrt{1+\frac{d^5(d+1)}{(A^{\mathrm{ch}})^2}\delta}\,\frac{A^{\mathrm{ch}}}{2}\sqrt{\mathop{\mathbb{E}}_{V\sim\mu} Y_V^2}\,, \tag{A.28}$$

$$= u(2d^4\,\delta)\sqrt{1+d^3(d+1)^3\delta}\;A^{\mathrm{ch}}\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Gamma)\,, \tag{A.29}$$

$$= \tilde{u}^{\mathrm{ch}}(\delta)\;A^{\mathrm{ch}}\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Gamma)\,, \tag{A.30}$$

where we defined

$$\tilde{u}^{\mathrm{ch}}(\delta) := u(2d^4\,\delta)\,\sqrt{1+d^3(d+1)^3\delta} \leq u(2d^4\,\delta)\,\sqrt{1+8d^6\delta}\,. \tag{A.31}$$

where in second step we used inequality $(x+a) \leq x(1+a)$ for $x, a \geq 1$,

To get the lower bound, we will integrate LHS of Eq. (A.18) and apply berger inequality. Proceeding analogously as before we obtain

$$\mathop{\mathbb{E}}_{V\sim\nu} Y_V^2 \geq \mathop{\mathbb{E}}_{V\sim\mu} Y_V^2 - 2\left\|(\Delta^\dagger)^{\otimes 2}\left[\mathbb{P}_{\mathrm{sym}}^{(2)}\right]\right\|_\infty\delta \;\geq\; \mathop{\mathbb{E}}_{V\sim\mu} Y_V^2(1-d^3(d+1)^3\delta)\,. \tag{A.32}$$

$$\mathop{\mathbb{E}}_{V\sim\nu} Y_V^4 \leq \mathop{\mathbb{E}}_{V\sim\mu} Y_V^4 + 4\left\|(\Delta^\dagger)^{\otimes 2}\left[\mathbb{P}_{\mathrm{sym}}^{(2)}\right]\right\|_\infty^2\delta \;\leq \mathop{\mathbb{E}}_{V\sim\mu} Y_V^4 + d^4(d+1)^4\|\mathcal{J}_\Delta\|_{\mathrm{HS}}^4\delta \tag{A.33}$$

We now recall that fourth moment w.r.p. to Haar measure is bounded by (see Eq. (3.39))

$$\mathop{\mathbb{E}}_{V\sim\mu} Y_V^4 \leq \nu\cdot\left(\mathop{\mathbb{E}}_{V\sim\mu} Y_V^2\right)^2 = \nu\cdot(A^{\mathrm{ch}})^4\,(2\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Gamma))^4\,, \tag{A.34}$$

with $\nu = \frac{\frac{13}{6}\binom{d+1}{2}^2}{\binom{d+3}{4}}$, where we used the fact that $\mathop{\mathbb{E}}_{V\sim\mu} Y_V^2 = (A^{\mathrm{ch}})^2\,(\|\mathcal{J}_\Delta\|_{\mathrm{HS}}^2+\mathrm{tr}(\Delta(\tau_{\mathrm{d}})^2) = (A^{\mathrm{ch}})^2\,(2\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Gamma))^2$.

We now note that $\|\mathcal{J}_\Delta\|_{\mathrm{HS}}^4 \leq \left(\|\mathcal{J}_\Delta\|_{\mathrm{HS}}^2 + \mathrm{tr}\left(\Delta\left(\tau_{\mathrm{d}}\right)^2\right)\right)^2 = (2\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Gamma))^4$, and combine it with Eq. (A.34) and Eq. (A.33) to obtain

$$\mathop{\mathbb{E}}_{V\sim\nu} Y_V^4 \leq \nu\,(A^{\mathrm{ch}})^4\,(2\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Gamma))^4\,\left(1-\delta\,\frac{d^4(d+1)^4}{\nu\,(A^{\mathrm{ch}})^4}\right) \tag{A.35}$$

$$= \nu\,(A^{\mathrm{ch}})^4\,(2\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Gamma))^4\,\left(1+\delta\,\frac{(d+1)^7(d+2)(d+3)}{13d}\right). \tag{A.36}$$

Similarly, we get

$$\mathop{\mathbb{E}}_{V\sim\nu} Y_V^2 \geq (A^{\mathrm{ch}})^2 \, (2\mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Gamma))^2 \, \left(1-\delta\, d^3(d+1)^3\right) \tag{A.37}$$

Inserting the above to Berger's inequality yields

$$\frac{\left(\mathop{\mathbb{E}}_{V\sim\nu} Y_V^2\right)^{3/2}}{\left(\mathop{\mathbb{E}}_{V\sim\nu} Y_V^4\right)^{1/2}} \geq 2\, \tilde{b}_{\mathrm{d}}\, \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Gamma)\, \frac{\left(1-\delta\, d^3(d+1)^3\right)^{3/2}}{\left(1+\delta\, \frac{(d+1)^7(d+2)(d+3)}{13d\,,}\right)^{1/2}} \tag{A.38}$$

where $\tilde{b}_{\mathrm{d}} = \frac{d}{d+1}\sqrt{\frac{(d+2)(d+3)}{13d(d+1)}}$.

Now we integrate Eq. (A.18) and combine with the above to obtain

$$\mathop{\mathbb{E}}_{U\sim\nu} \mathrm{TV}(\mathbf{p}^{\Lambda,\psi_V,U}, \mathbf{p}^{\Gamma,\psi_V,U}) \;\geq a^{\mathrm{ch}}\, \mathrm{d}_{\mathrm{av}}^{\mathrm{ch}}(\Lambda,\Gamma)\, \tilde{I}^{\mathrm{ch}}(\delta)\,, \tag{A.39}$$

where

$$\tilde{I}^{\mathrm{ch}}(\delta) = \ell(2d^4\,\delta)\, \frac{\left(1-\delta\, d^3(d+1)^3\right)^{3/2}}{\left(1+\delta\, \frac{(d+1)^7(d+2)(d+3)}{13d\,,}\right)^{1/2}} \tag{A.40}$$

$$\geq \ell(2d^4\,\delta)\, \frac{\left(1-\delta\, 8\, d^6\right)^{3/2}}{\left(1+\delta\, \frac{d^8\cdot 2^7\cdot 12}{13}\right)^{1/2}} \tag{A.41}$$

$$\geq \ell(2d^4\,\delta)\, \frac{\left(1-\delta\, 2^8\, d^6\right)^{3/2}}{\left(1+\delta\, (2d)^8\right)^{1/2}}\,, \tag{A.42}$$

where we used inequality $(x+a) \leq x(1+a)$ for $x, a \geq 1$.

Now we set $\delta := \frac{\delta'}{(2d)^8}$ and get

$$\tilde{I}^{\mathrm{ch}}(\delta') \geq \ell(\frac{\delta'}{2^7 d^4})\, \frac{\left(1-\frac{\delta'}{d^2}\right)^{3/2}}{(1+\delta')^{1/2}} \geq \ell(\delta')\, \frac{\left(1-\frac{\delta'}{d^2}\right)^{3/2}}{(1+\delta')^{1/2}} \geq \frac{\left(1-\frac{\delta'}{d^2}\right)^{3}}{1+\delta'} =: \ell^{\mathrm{ch}}(\delta')\,, \tag{A.43}$$

where in second inequality we used the fact that $\ell(x)$ is a decreasing function of $x \geq 0$.

With set δ, we also rewrite upper bound from Eq. (A.31) as

$$\tilde{u}^{\mathrm{ch}}(\delta') \leq u(\frac{\delta'}{2^7 d^4})\sqrt{1+\frac{\delta'}{2^5 d^2}} \leq u(\delta)\sqrt{1+\frac{\delta'}{d^2}} \leq 1+\frac{\delta'}{d^2} =: u^{\mathrm{ch}}(\delta')\,, \tag{A.44}$$

where we used the fact that $u(x)$ is an increasing function of $x \geq 0$.

B. Modeling and mitigation of correlated readout noise

We collect here technical results that are used in the Chapter 4. Some of the results stated here can be of independent interest for further works on quantum error mitigation. In Appendix B.1 we discuss the details of error mitigation on marginals for correlated readout noise models, while Appendix B.2 provides details of our noise characterization procedure based on the Diagonal Detector Overlapping Tomography technique. In Appendix B.3 we give a short overview of the whole noise characterization scheme in a step-by-step manner. Results concerning the sample complexity of energy estimation are discussed in Appendix B.4, and finally, in Appendix B.5 we provide some additional experimental data and details of the numerical simulations.

B.1 Correlated readout noise model and its usage for error-mitigation on marginals

In this section, we provide some details on how correlated measurement noise affects marginal probability distributions. We start by providing in Appendix B.1.1 explicit relation between ideal and noisy marginals, when the global probability distribution is affected by readout noise given by our model. Then in Appendix B.1.2, we discuss the error-mitigation on the level of marginals, including proof of Proposition I from the main text. We finish by analyzing errors in simultaneous estimation of multiple marginal distributions (Appendix B.1.3) and estimation of multiple expected values of local terms (Appendix B.1.4).

B.1.1 Noise model for marginal distributions

In this subsection, we show how to translate a noise model for the full probability distribution given in Eq. (4.3) into a simple noise model for marginal distributions. Our results can be summarised in the following Proposition 3.

Proposition 3. *Let $p(Y_1Y_2\ldots,Y_N)$ and $\tilde{p}(X_1X_2,\ldots,X_N)$ be the ideal distribution and the one obtained on noisy detector respectively. Define by $p(\mathbf{Y}_S)$ and $\tilde{p}(\mathbf{X}_S)$ their corresponding marginal distributions for qubits in some subset $\mathcal{S}$. If the full distributions are connected by a stochastic matrix of the form Eq.* (4.3) *by means*

of

$$\tilde{p}\left(X_1 \ldots X_N\right) = \sum_{Y_1 Y_2 \ldots Y_N} p\left(Y_1 Y_2 \ldots Y_N\right) \mathsf{T}_{X_1 X_2 \ldots X_N | Y_1 Y_2 \ldots Y_N} \,, \tag{B.1}$$

then there exists a left-stochastic matrix $\mathsf{T}^{(\mathcal{S})}$ *that connects the noisy marginal to the ideal one, namely*

$$\tilde{p}(\mathbf{X}_{\mathcal{S}}) = \sum_{\mathbf{Y}_{\mathcal{S}}} \mathsf{T}^{(\mathcal{S})}_{\mathbf{X}_{\mathcal{S}}|\mathbf{Y}_{\mathcal{S}}} \, p(\mathbf{Y}_{\mathcal{S}}) \,. \tag{B.2}$$

Moreover, the form of the marginal noise matrix can be explicitly derived from the decomposition Eq. (4.3)*.*

We will now proceed to prove the above Proposition by deriving the concrete expression of $\mathsf{T}^{(\mathcal{S})}_{\mathbf{X}_{\mathcal{S}}|\mathbf{Y}_{\mathcal{S}}}$ as a function of the matrices in the given noise model. Let us start by computing the marginal of the noisy distribution on general $\mathcal{S}$. Let us denote by $C\left(\mathcal{S}\right) = \left\{C_\chi\right\}_\chi$ set of clusters which contain qubits from $\mathcal{S}$ (note that we might be interested in taking marginal over some qubits from clusters to which qubits from $\mathcal{S}$ belong), and by $L\left(\mathcal{S}\right) = \{\chi\}_\chi$ the corresponding set of labels of those clusters. Following from Eq. (B.1) with global noise noise map given by Eq. (4.3) we have

$$\begin{aligned}
\tilde{p}\left(\mathbf{X}_{\mathcal{S}}\right) &= \sum_{X \notin \mathbf{X}_{\mathcal{S}}} \tilde{p}\left(X_1 X_2 \ldots X_N\right) \\
&= \sum_{\substack{X \in C(\mathcal{S}) \\ X \notin \mathbf{X}_{\mathcal{S}}}} \sum_{X \notin C(\mathcal{S})} \sum_{Y_1 Y_2 \ldots Y_N} p\left(Y_1 Y_2 \ldots Y_N\right) \prod_\chi \mathsf{T}^{\mathbf{Y}_{\mathbf{N}_\chi}}_{\mathbf{X}_{C_\chi}|\mathbf{Y}_{C_\chi}|} \; = \\
&= \sum_{\substack{X \in C(\mathcal{S}) \\ X \notin \mathbf{X}_{\mathcal{S}}}} \sum_{Y_1 Y_2 \ldots Y_N} p\left(Y_1 Y_2 \ldots Y_N\right) \prod_{\chi \in L(\mathcal{S})} \mathsf{T}^{\mathbf{Y}_{\mathbf{N}_\chi}}_{\mathbf{X}_{C_\chi}|\mathbf{Y}_{C_\chi}|} \prod_{\chi \notin L(\mathcal{S})} \sum_{X \notin C(\mathcal{S})} \mathsf{T}^{\mathbf{Y}_{\mathbf{N}_\chi}}_{\mathbf{X}_{C_\chi}|\mathbf{Y}_{C_\chi}} \; = \\
&= \sum_{\substack{X \in C(\mathcal{S}) \\ X \notin \mathbf{X}_{\mathcal{S}}}} \sum_{Y_1 Y_2 \ldots Y_N} p\left(Y_1 Y_2 \ldots Y_N\right) \prod_{\chi \in L(\mathcal{S})} \mathsf{T}^{\mathbf{Y}_{\mathbf{N}_\chi}}_{\mathbf{X}_{C_\chi}|\mathbf{Y}_{C_\chi}|}, = \\
&= \sum_{\substack{X \in C(\mathcal{S}) \\ X \notin \mathbf{X}_{\mathcal{S}}}} \sum_{Y \in \mathbf{Y}_{\mathbf{N}(\mathcal{S}) \cup C(\mathcal{S})}} p\left(\mathbf{Y}_{\mathbf{N}(\mathcal{S}) \cup C(\mathcal{S})}\right) \prod_{\chi \in L(\mathcal{S})} \mathsf{T}^{\mathbf{Y}_{\mathbf{N}_\chi}}_{\mathbf{X}_{C_\chi}|\mathbf{Y}_{C_\chi}|} \,,
\end{aligned}$$

where to obtain the above simplifications, we have exploited the fact that the noise matrices $\mathsf{T}^{\mathbf{Y}_{\mathbf{N}_\chi}}$ are all left-stochastic (for any fixed $Y_{\mathbf{N}_\chi}$) and we have defined $p\left(\mathbf{Y}_{\mathbf{N}(\mathcal{S}) \cup C(\mathcal{S})}\right)$ as the marginal of the ideal distribution on the qubits belonging both the clusters and the neighborhoods of qubits from $\mathcal{S}$. By $\mathbf{N}\left(\mathcal{S}\right)$ we denote set of qubits from neighbourhoods of clusters $C\left(\mathcal{S}\right)$ but without including the qubits from $\mathcal{S}$, i.e., $\mathbf{N}\left(\mathcal{S}\right) = \cup_{\chi \in L(\mathcal{S})} \left(\mathbf{N}_\chi / \mathcal{S}\right)$. Note that this additional requirement is necessary since qubits which are neighbors of some clusters in $C\left(\mathcal{S}\right)$ might belong to some other clusters in $C\left(\mathcal{S}\right)$.

Now by means of the chain rule for probability we decompose

$$p\left(\mathbf{Y}_{\mathbf{N}(\mathcal{S})\cup C(\mathcal{S})}\right) = p\left(\mathbf{Y}_{\mathcal{S}}\right) p\left(\mathbf{Y}_{\{\mathbf{N}(\mathcal{S})\cup C(\mathcal{S})\}/\{\mathcal{S}\}}|\mathbf{Y}_{\mathcal{S}}\right) \tag{B.3}$$

which after substituting to the Eq. (B.3) gives

$$\tilde{p}\left(\mathbf{X}_{\mathcal{S}}\right) = \sum_{Y\in\mathbf{Y}_{\mathcal{S}}} p\left(\mathbf{Y}_{\mathcal{S}}\right) \sum_{\substack{X\in C(\mathcal{S})\\ X\notin\mathbf{X}_{\mathcal{S}}}} \sum_{\substack{Y\in\mathbf{Y}_{\mathbf{N}(\mathcal{S})\cup C(\mathcal{S})}\\ Y\notin\mathbf{Y}_{\mathcal{S}}}} p\left(\mathbf{Y}_{\{\mathbf{N}(\mathcal{S})\cup C(\mathcal{S})\}/\{\mathcal{S}\}}|\mathbf{Y}_{\mathcal{S}}\right) \prod_{\chi\in L(\mathcal{S})} \mathsf{T}^{\mathbf{Y}_{\mathbf{N}_\chi}}_{\mathbf{X}_{C_\chi}|\mathbf{Y}_{C_\chi}|} \,. \tag{B.4}$$

Now notice that Eq. (B.4) coincides exactly with Eq. (B.2) if we identify the marginal noise matrix elements as

$$\mathsf{T}^{(\mathcal{S})}_{\mathbf{X}_{\mathcal{S}}|Y_{\mathcal{S}}} := \sum_{\substack{X\in C(\mathcal{S})\\ X\notin\mathbf{X}_{\mathcal{S}}}} \sum_{\substack{Y\in\mathbf{Y}_{\mathbf{N}(\mathcal{S})\cup C(\mathcal{S})}\\ Y\notin\mathbf{Y}_{\mathcal{S}}}} p\left(\mathbf{Y}_{\{\mathbf{N}(\mathcal{S})\cup C(\mathcal{S})\}/\{\mathcal{S}\}}|\mathbf{Y}_{\mathcal{S}}\right) \prod_{\chi\in L(\mathcal{S})} \mathsf{T}^{\mathbf{Y}_{\mathbf{N}_\chi}}_{\mathbf{X}_{C_\chi}|\mathbf{Y}_{C_\chi}|} \,, \tag{B.5}$$

which shows, unsurprisingly, that the noise matrix acting on the marginal is state-dependent (via marginal distribution $p\left(\mathbf{Y}_{\{\mathbf{N}(\mathcal{S})\cup C(\mathcal{S})\}/\{\mathcal{S}\}}|\mathbf{Y}_{\mathcal{S}}\right)$). Note that the above equation *does not* include simple matrix multiplication of noise matrices for neighborhoods $\mathbf{N}\left(\mathcal{S}\right)$. In other words, the generic marginal noise matrix *is not* a convex combination of the matrices from the set

$$\cup_{\chi\in L(\mathcal{S})} \left\{\mathsf{T}^{\mathbf{Y}_{\mathbf{N}_\chi}}\right\}_{\mathbf{Y}_{\mathbf{N}_\chi}} \tag{B.6}$$

even if the neighborhoods do not overlap.

It is now left to show that the matrix defined in Eq. (B.5) is left-stochastic. To this aim, let us sum over string $\mathbf{X}_{\mathcal{S}}$ and obtain

$$\begin{aligned}
\sum_{\mathbf{X}_{\mathcal{S}}} \mathsf{T}^{(\mathcal{S})}_{\mathbf{X}_{\mathcal{S}}|Y_{\mathcal{S}}} &:= \sum_{\mathbf{X}_{\mathcal{S}}} \sum_{\substack{X\in C(\mathcal{S})\\ X\notin\mathbf{X}_{\mathcal{S}}}} \sum_{\substack{Y\in\mathbf{Y}_{\mathbf{N}(\mathcal{S})\cup C(\mathcal{S})}\\ Y\notin\mathbf{Y}_{\mathcal{S}}}} p\left(\mathbf{Y}_{\{\mathbf{N}(\mathcal{S})\cup C(\mathcal{S})\}/\{\mathcal{S}\}}|\mathbf{Y}_{\mathcal{S}}\right) \prod_{\chi\in L(\mathcal{S})} \mathsf{T}^{\mathbf{Y}_{\mathbf{N}_\chi}}_{\mathbf{X}_{C_\chi}|\mathbf{Y}_{C_\chi}} = \\
&= \sum_{\substack{Y\in\mathbf{Y}_{\mathbf{N}(\mathcal{S})\cup C(\mathcal{S})}\\ Y\notin\mathbf{Y}_{\mathcal{S}}}} p\left(\mathbf{Y}_{\{\mathbf{N}(\mathcal{S})\cup C(\mathcal{S})\}/\{\mathcal{S}\}}|\mathbf{Y}_{\mathcal{S}}\right) \sum_{\mathbf{X}_{\mathcal{S}}} \prod_{\chi\in L(\mathcal{S})} \mathsf{T}^{\mathbf{Y}_{\mathbf{N}_\chi}}_{\mathbf{X}_{C_\chi}|\mathbf{Y}_{C_\chi}} = \\
&= \sum_{\substack{Y\in\mathbf{Y}_{\mathbf{N}(\mathcal{S})\cup C(\mathcal{S})}\\ Y\notin\mathbf{Y}_{\mathcal{S}}}} p\left(\mathbf{Y}_{\{\mathbf{N}(\mathcal{S})\cup C(\mathcal{S})\}/\{\mathcal{S}\}}|\mathbf{Y}_{\mathcal{S}}\right) = 1 \,,
\end{aligned}$$

where we simply used the fact that the cluster noise matrices are left-stochastic and that the conditional distribution $p\left(\mathbf{Y}_{\{\mathbf{N}(\mathcal{S})\cup C(\mathcal{S})\}/\{\mathcal{S}\}}|\mathbf{Y}_{\mathcal{S}}\right)$ is normalised.

B.1.2 Noise mitigation for marginal distributions – proof of Proposition 1

Here we analyze in detail the noise mitigation strategy outlined in the main text, focusing on its scalability and effectiveness to recover the ideal marginal distribution. Again let us assume that we are interested in marginal on qubits from some subset $\mathcal{S}$.

Let us start by outlining the reasoning behind our choice of the mitigation strategy. Recall that, as proved in the previous section, the noisy marginals can always be related to ideal ones by means of a marginal stochastic matrix as for Eq. (B.2). Hence, taking the inverse of the marginal noise matrix would the natural strategy for correct noise mitigation. However, recall that, as shown in Eq. (B.5), the form of the marginal noise matrix depends on the ideal distribution itself, which is generally unknown. Therefore one needs an ansatz for such a distribution, which hopefully will work in the mitigation generic case of arbitrary conditional distribution. As indicated in the main text, the natural choice for such ansatz is a uniform distribution (inserted in place of $p\left(\mathbf{Y}_{\{\mathbf{N}(\mathcal{S})\cup C(\mathcal{S})\}\setminus\{\mathcal{S}\}}|\mathbf{Y}_{\mathcal{S}}\right)$ in Eq. (B.5)).

Now, let us make the following important observation. We are interested in mitigation of the noise of a $|\mathcal{S}|$-qubit marginal distribution of qubits generally belonging to some set of distinct clusters $C\left(\mathcal{S}\right)$. Note that, by definition, a cluster is a set of qubits with correlations in errors so big, that one can not consider the measurement outcomes on them separately. We argue that, if one is interested in correcting the marginal distribution of only *parts* of the clusters, say $\tilde{p}(\mathbf{X}_{\mathcal{S}})$, it is still a better idea *first* correct the marginal distribution on the whole variables in clusters, namely

$$\tilde{p}(\mathbf{X}_{C(\mathcal{S})}) = \sum_{\mathbf{Y}_{C(\mathcal{S})}} p(\mathbf{Y}_{C(\mathcal{S})})\, \mathsf{T}^{(C(\mathcal{S}))}_{\mathbf{X}_{C(\mathcal{S})}|\mathbf{Y}_{C(\mathcal{S})}} \,. \tag{B.7}$$

Notice that the corrected cluster marginal distribution can always be post-processed to obtain the two-body marginal distribution of interest. By proceeding in a similar manner as in the previous section, one can obtain the following expression for the cluster noise matrix

$$\mathsf{T}^{(C(\mathcal{S}))}_{\mathbf{X}_{C(\mathcal{S})}|\mathbf{Y}_{C(\mathcal{S})}} = \sum_{\mathbf{Y}_{\mathbf{N}(\mathcal{S})}} p\left(\mathbf{Y}_{\mathbf{N}(\mathcal{S})}|\mathbf{Y}_{C(\mathcal{S})}\right) \prod_{\chi\in L(\mathcal{S})} \mathsf{T}^{\mathbf{Y}_{\mathbf{N}_\chi}}_{\mathbf{X}_{C_\chi}|\mathbf{Y}_{C_\chi}} \,, \tag{B.8}$$

where $L\left(\mathcal{S}\right)$ is a set of labels of the clusters $C\left(\mathcal{S}\right)$ (note that those are also labels for neighbours of those clusters). Recall that $\mathbf{N}\left(\mathcal{S}\right)$ was defined as set of qubits from neighbourhoods of clusters $C\left(\mathcal{S}\right)$ but without including the qubits from $\mathcal{S}$, i.e., $\mathbf{N}\left(\mathcal{S}\right) = \cup_{\chi\in L(\mathcal{S})}\left(\mathbf{N}_\chi \setminus \mathcal{S}\right)$, so in the above definition we do not average over some qubits in clusters. As mentioned before, since the exact form of such a matrix requires access to the unknown information of $p\left(\mathbf{Y}_{\mathbf{N}(\mathcal{S})}|\mathbf{Y}_{C(\mathcal{S})}\right)$, we propose instead to invert the average cluster matrix

$$\mathsf{T}^{C(\mathcal{S})}_{av} := \frac{1}{2^{|\mathbf{N}(\mathcal{S})|}} \sum_{\mathbf{Y}_{\mathbf{N}(\mathcal{S})}} \mathsf{T}^{\mathbf{Y}_{\mathbf{N}(\mathcal{S})}} \,, \tag{B.9}$$

and use it to perform the error mitigation. From our numerical analysis, the replacement of the cluster noise matrix with its average version yields a much better correction than making a similar replacement at the level of the $|\mathcal{S}|$-body marginals (however we tested in only for the case of 2-qubit marginal distributions). This provides a clear indication in favor of performing the error mitigation

at the level of the clusters instead of the single variables. Moreover, notice that the computational cost of computing the two noise matrices in Eq. (B.8) and (B.5) is comparable. Indeed, both require access to the collections of matrices $\left\{\mathsf{T}^{\mathbf{Y}_{\mathbf{N}(\mathcal{S})}}\right\}_{\mathbf{Y}_{\mathbf{N}(\mathcal{S})}}$. Hence, as long as the size of the involved clusters and neighborhoods is reasonably small (particularly, it does not scale with the system size), computing the matrix in Eq. (B.9) is efficient.

Due to the above, the following discussion will concern the mitigation on the level of marginals on the level of clusters $\mathcal{C}(\mathcal{S})$. For clarity, let us from now on use slightly changed notation

$$\mathcal{S} \to \mathcal{S} = \cup_\chi C_\chi \,, \tag{B.10}$$

to indicate that we are interested in error-mitigation on the subset of qubits $\mathcal{S}$ which consists of full clusters, i.e., prior to performing error-mitigation we do not wish to take marginal over qubits belonging to the same cluster (while later one can of course marginalize the corrected distributions).

Having obtained $\mathsf{T}^{\mathcal{S}}_{av}$, one can use its inverse as a correction matrix which can reduce the noise on the marginal distribution, by defining the corrected distribution as

$$\mathbf{p}^{\text{corr}}_{\mathcal{S}} = \left(\mathsf{T}^{\mathcal{S}}_{av}\right)^{-1} \tilde{\mathbf{p}}_{\mathcal{S}} \,, \tag{B.11}$$

where we used vector notation for clarity. $\tilde{\mathbf{p}}_{\mathcal{S}}$ is noisy distribution on qubits from $\mathcal{S}$. Similarly, the vector without tilde symbol $\mathbf{p}_{\mathcal{S}}$ will denote corresponding ideal distribution, i.e., we have $\tilde{\mathbf{p}}_{\mathcal{S}} = \mathsf{T}^{(\mathcal{S})}\mathbf{p}_{\mathcal{S}}$. Error reduction via average matrix is perfect only in the case [1] of infinite-statistics and under the assumption that the actual conditional distribution in Eq. (B.8) is uniform (because then $\mathsf{T}^{\mathcal{S}} = \mathsf{T}^{\mathcal{S}}_{av}$). Since this scenario is not realistic, in practice one can hope only for partial noise mitigation, and not its complete reverse. The errors related to statistical noise when correcting probability distribution were thoroughly analyzed in work [42] (we also repeat a similar analysis in the context of expected values of Hamiltonians in next sections). Here we will analyze the errors which arise due to the fact that we use the inverse of the average cluster matrix $\mathsf{T}^{(\mathcal{S})}_{av}$ instead of inverse of the exact matrix $\mathsf{T}^{(\mathcal{S})}$, while keeping the assumption of infinite statistics. We will denote effective correction matrix as $\mathsf{A}^{(\mathcal{S})}_{av} := \left(\mathsf{T}^{(\mathcal{S})}_{av}\right)^{-1}$ (as in Eq. (4.9)).

To quantify the errors, we make use the Total-Variation-Distance (see Eq. (B.56)) to measure the distance between probability distributions. For later purposes, let us also recall that the operator norm induced by the vector L1 norm is

$$||A||_{1\to 1} = \sup_{|||v\rangle||_1=1} ||A\,|v\rangle\,||_1, \tag{B.12}$$

where A is any linear operator acting on given vector space.

[1] Clearly, the implicit assumption is that our model of noise is exact.

Now we will proceed to proving Proposition 1 which states that the maximum error due to using average correction matrix is given by

$$\mathrm{TVD}\left(\mathbf{p}_{\mathcal{S}}^{\mathrm{corr}}, \mathbf{p}^{\mathcal{S}}\right) \leq \frac{1}{2}||\mathsf{A}_{av}^{\mathcal{S}}||_{1\to 1} \max_{\mathbf{Y}_{\mathbf{N}(\mathcal{S})}} ||\mathsf{T}_{av}^{\mathcal{S}} - \mathsf{T}^{\mathbf{Y}_{\mathbf{N}(\mathcal{S})}}||_{1\to 1}\,, \tag{B.13}$$

As a starting point for the proof, let us make following decomposition $\mathsf{T}^{(\mathcal{S})} = \mathsf{T}_{av}^{(\mathcal{S})} + \left(\mathsf{T}^{(\mathcal{S})} - \mathsf{T}_{av}^{(\mathcal{S})}\right)$ so to replace the left-hand-side of Eq. (B.13) with

$$\begin{aligned} ||\mathsf{A}_{av}^{(\mathcal{S})}\left(\mathsf{T}^{(\mathcal{S})}\tilde{\mathbf{p}}_{\mathcal{S}}\right) - \mathbf{p}_{\mathcal{S}}|_1 &= ||\mathsf{A}_{av}^{(\mathcal{S})}\left(\mathsf{T}^{(\mathcal{S})} - \mathsf{T}_{av}^{(\mathcal{S})}\right)\mathbf{p}_{\mathcal{S}}||_1 \\ &\leq \sup_{\mathbf{p}_{\mathcal{S}}} ||\mathsf{A}_{av}^{(\mathcal{S})}\left(\mathsf{T}^{(\mathcal{S})} - \mathsf{T}_{av}^{(\mathcal{S})}\right)\mathbf{p}_{\mathcal{S}}||_1 \\ &\leq ||\mathsf{A}_{av}^{(\mathcal{S})}||_{1\to 1}\, ||\mathsf{T}^{(\mathcal{S})} - \mathsf{T}_{av}^{(\mathcal{S})}||_{1\to 1}\,, \end{aligned} \tag{B.14}$$

where the first inequality indicates the maximization over all possible probability distributions over the set $\mathcal{S}$ variables and the second inequality follows from Eq. (B.12) and the sub-multiplicativity of the L1 norm. Now, let us bound the second term of the above equation as follows

$$\begin{aligned} ||\mathsf{T}^{(\mathcal{S})} - \mathsf{T}_{av}^{(\mathcal{S})}||_{1\to 1} &= ||\left(\sum_{\mathbf{Y}_{\mathbf{N}(\mathcal{S})}} p\left(\mathbf{Y}_{\mathbf{N}(\mathcal{S})}|\mathbf{Y}_{\mathcal{S}}\right) \prod_{\chi\in L(\mathcal{S})} \mathsf{T}_{\mathbf{X}_{C_\chi}|\mathbf{Y}_{C_\chi}}^{\mathbf{Y}_{\mathbf{N}_\chi}}\right) - \mathsf{T}_{av}^{(\mathcal{S})}||_{1\to 1} \\ &\leq \sup_{\{p(\mathbf{Y}_{\mathbf{N}(\mathcal{S})}|\mathbf{Y}_{\mathcal{S}})\}} ||\left(\sum_{\mathbf{Y}_{\mathbf{N}(\mathcal{S})}} p\left(\mathbf{Y}_{\mathbf{N}(\mathcal{S})}|\mathbf{Y}_{\mathcal{S}}\right) \prod_{\chi\in L(\mathcal{S})} \mathsf{T}_{\mathbf{X}_{C_\chi}|\mathbf{Y}_{C_\chi}}^{\mathbf{Y}_{\mathbf{N}_\chi}}\right) - \mathsf{T}_{av}^{(\mathcal{S})}||_{1\to 1}\,, \end{aligned} \tag{B.15}$$

where the above sequence of inequalities removes the dependence on the unknown conditional distribution $p\left(\mathbf{Y}_{\mathbf{N}(\mathcal{S})}|\mathbf{Y}_{\mathcal{S}}\right)$.

The maximization in Eq. (B.15) corresponds to the maximal distance of the average matrix $\mathsf{T}_{av}^{(\mathcal{S})}$ from the convex hull of the set of matrices $\left\{\mathsf{T}^{\mathbf{Y}_{\mathbf{N}(\mathcal{S})}}\right\}_{\mathbf{Y}_{\mathbf{N}(\mathcal{S})}}$ (hence over a polytope with extremal points being each of the matrices $\mathsf{T}^{\mathbf{Y}_{\mathbf{N}(\mathcal{S})}}$). It is straightforward to see that this function is convex, due to the triangle inequality and absolute homogeneity of the L1 norm. From this fact, it follows that the maximum in Eq. (B.15) is attained for one of the extremal points of the convex hull, i.e., a particular $\mathsf{T}^{\mathbf{Y}_{\mathbf{N}_i}}\mathsf{T}^{\mathbf{Y}_{\mathbf{N}(\mathcal{S})}}$. Hence, combining this last bound with Eq. (B.14) one obtains exactly Eq. (B.13), which proves Proposition 1.

B.1.3 Statistical error bounds

The idea behind the energy estimation routine in variational algorithms such as QAOA is to estimate Hamiltonian with K-terms by estimating those terms separately and then adding them up. However, if one provides statistical error bounds for each of those terms, one needs to take into account the probabilistic nature

of such bounds. Let us say that we want to estimate some distribution $\mathbf{p}$ with N outcomes by sampling from it s number of times. It is well known [196] that the probability of estimated distribution being ϵ-close to the true distribution $\mathbf{p}$ in terms of TVD (Eq. (4.7) is vanishing exponentially in number of sample

$$\Pr\left(\mathrm{TVD}\left(\mathbf{p}, \mathbf{p}^{est}\right) \geq \epsilon\right) \leq (2^n - 2) \exp\left(-2s\epsilon^2\right) . \tag{B.16}$$

Now it is often convenient to set the probability fixed and consider the confidence intervals, i.e., the bound for TVD $\left(\mathbf{p}, \mathbf{p}^{est}\right)$. Then we can rewrite the above equation as a function $\epsilon^*\left(n, s, P_{1,\mathrm{err}}, 1\right)$ of three fixed parameters – number of outcomes n, number of samples s and probability of the upper bound being incorrect $P_{1,\mathrm{err}}$. Then the basic manipulations of Eq.(B.16) give

$$\mathrm{TVD}\left(\mathbf{p}, \mathbf{p}^{est}\right) \leq \epsilon^*\left(n, s, P_{1,\mathrm{err}}, 1\right) = \sqrt{\frac{\log\left(2^n - 2\right) - \log\left(P_{1,\mathrm{err}}\right)}{2s}} . \tag{B.17}$$

However, since in estimation of the K-term Hamiltonian one combines the upper bounds of the form (B.17) for particular terms into upper bound $\epsilon^*\left(n, s, P_{1,\mathrm{err}}, K\right)$ for the whole Hamiltonian, one needs to make sure that all upper bounds for particular terms are true at the same time with the desired probability. The union bound states that the probability of at least one event from some set occurring is no greater than the sum of probabilities of particular events. In our case, the interesting set consists of events of the type "one of the bounds of type (B.17) is not satisfied". Hence the probability of at least one bound being wrong is upper bounded by

$$P_{K,\mathrm{err}} \leq K P_{1,\mathrm{err}} . \tag{B.18}$$

Therefore, if we wish to ensure that probability of all the bounds being right at the same time is fixed and equal to $P_{1,\mathrm{err}}$ for fixed number of samples s, we need to effectively increase the upper bound to

$$\begin{aligned}
&\epsilon^*\left(n, s, P_{1,\mathrm{err}}, K\right) = \\
&= \sqrt{\frac{\log\left(2^n - 2\right) - \log\left(\frac{P_{1,\mathrm{err}}}{K}\right)}{2s}} = \\
&= \sqrt{\frac{\log\left(2^n - 2\right) - \log\left(P_{1,\mathrm{err}}\right) + \log\left(K\right)}{2s}}
\end{aligned}$$

with additional term $\log K$ under the square root. In other words, we effectively lower the probability of error occurring in estimation of particular marginal distributions by a factor of K, which ensures that simultaneous estimation of K marginal distributions has the precision from Eq. (B.19) with probability not lower than the initial $P_{1,\mathrm{err}}$ (by the virtue of union bound).

It is instructive to look at this from the perspective of sampling complexity. Let's say that we would like to increase number of samples

$$s \longrightarrow \tilde{s} =: s\,(1 + f_{\text{oh}})$$

in such a way, that upper bound remains fixed, i.e.,

$$\epsilon^* \left(n, \tilde{s}, P_{1,\text{err}}, K\right) = \epsilon^* \left(n, s, P_{1,\text{err}}, 1\right) \ .$$

Simple calculations show that in this case we would need to (multiplicatively) increase the number of samples s by the overhead equal to

$$1 + f_{\text{oh}} = 1 + C \log\left(K\right)$$

where parameter C is equal to

$$\left(\log\left(\frac{2^n - 2}{P_{1,\text{err}}}\right)\right)^{-1} \ . \tag{B.19}$$

Hence we see that for fixed dimension (fixed number of outcomes) simultaneous estimation of K Hamiltonian terms leads to sampling overhead logarithmic in K.

Note that if one fixes the initial sampling size s, probability of error $P_{1,\text{err}}$ and number of simultaneously estimated marginals K, the sampling overhead is, perhaps counterintuitively, decaying in the number of outcomes. This dependence is roughly linear in the number of outcomes, hence it is exponential in the number of qubits.

B.1.4 Proofs of energy error bounds

Here we will provide proofs of worst-case error bounds for energy estimation given in the main text. Note that in the following discussion, for clarity of notation we won't add a special index to marginal probability distributions indicating that they are marginals. This should be clear from the context. On the contrary, local Hamiltonian terms, noise matrices, and correction matrices will have an additional label (α) indicating their locality.

Approximation errors

We will start by providing proof of Eq. (4.19) which gives additive bound for possible deviations of error-mitigated energy from ideal one when the source of deviations is an approximation used in constructing noise model (recall Eq. (4.8)). Consider estimation of the local term of Hamiltonian H_α. Recall that locality of the terms means that H_α acts non-trivially only on the subset of qubits. Now let us denote by $\mathbf{p}$ the marginal probability distribution on the qubits belonging to the relevant subset. Then expected value of the local term can be written as

$$\langle H_\alpha \rangle = \sum_t \mathsf{T}^\alpha_t p_t = \langle \mathsf{T}^\alpha | \mathbf{p} \rangle, \tag{B.20}$$

where by T^{α} we denoted vector of eigenvalues of local term H_{α}, and we used convenient braket notation to denote scalar product. For example, if $H_{\alpha} = \sigma^{z} \otimes \sigma^{z}$, then $\mathsf{T}^{\alpha} = (1, -1, -1, 1)$. Now we want to consider two different estimators of the same local term – one from ideal distribution $\mathbf{p}^{\text{ideal}}$ and second from the error-mitigated noisy distribution $\mathbf{p}^{\text{corr}} = C_{av}\mathbf{p}^{noisy}$ (recall Proposition 1). Now we want to upper bound the difference between energy estimators based on those two marginal distributions. From Eq. (B.20) we can write

$$|\left\langle H_{\alpha}^{\text{corr}}\right\rangle - \left\langle H_{\alpha}^{\text{ideal}}\right\rangle| = |\langle \mathsf{T}|\left(\left|\mathbf{p}^{\text{corr}}\right\rangle - \left|\mathbf{p}^{\text{ideal}}\right\rangle\right)| \leq \underbrace{\max_{t}|\mathsf{T}_{t}|}_{||H_{\alpha}||}\underbrace{\langle \mathbb{I}|\left|\left(\left|\mathbf{p}^{\text{corr}}\right\rangle - \left|\mathbf{p}^{\text{ideal}}\right\rangle\right)\right|}_{2\text{TVD}\left(\mathbf{p}^{\text{corr}},\mathbf{p}^{\text{ideal}}\right)} \leq 2||H_{\alpha}||\delta_{\alpha}\,, \tag{B.21}$$

where δ^{α} is approximation error defined in Eq. (4.15), $\langle \mathbb{I}|$ is a vector of ones and the last inequality follows from Lemma 1. Factor two comes from the fact that Total-Variation Distance is defined with $\frac{1}{2}$ factor. The additive error bound from Eq. (4.19) is just a multiple application of triangle inequality

$$|\left\langle \mathcal{H}^{\text{corr}}\right\rangle - \langle \mathcal{H}\rangle| = |\sum_{\alpha}\left\langle H_{\alpha}^{\text{corr}}\right\rangle - \left\langle H_{\alpha}^{\text{ideal}}\right\rangle| \leq \sum_{\alpha}|\left\langle H_{\alpha}^{\text{corr}}\right\rangle - \left\langle H_{\alpha}^{\text{ideal}}\right\rangle| \leq 2\sum_{\alpha}\delta^{\alpha}||H_{\alpha}||\,. \tag{B.22}$$

Statistical errors

Now we will proceed to prove Eq. (4.19) which bounds the effects of statistical noise on the error-mitigation. First, let us assume that the noise matrix acting on the marginal of interest is known exactly (not approximately as in the above derivations). In that case, in the lack of presence of the statistical errors, the correction is done exactly. However, in reality the statistics are finite, and the *estimator* of the noisy marginal distribution $\mathbf{p}^{noisy}$ from which we sample on imperfect detector can be formally written as

$$\underbrace{\mathbf{p}^{noisy}}_{\text{true}} \rightarrow \underbrace{\mathbf{p}^{noisy} + }_{estimated}\,. \tag{B.23}$$

Now we want to bound the error in the energy estimated after acting by correction matrix T_{α}^{-1} on the estimated distribution above. In analogy to Eq. (B.21) we can write

$$|H_{\alpha}^{\text{corr, est}} - \left\langle H_{\alpha}^{\text{ideal}}\right\rangle| = |\langle \mathsf{T}|\left(\underbrace{\mathsf{T}_{\alpha}^{-1}\left|\mathbf{p}^{noisy}\right\rangle}_{\left|\mathbf{p}^{\text{ideal}}\right\rangle} + \mathsf{T}_{\alpha}^{-1}|\rangle - \left|\mathbf{p}^{\text{ideal}}\right\rangle\right)| = |\langle \mathsf{T}|\mathsf{T}_{\alpha}^{-1}|\rangle\rangle| \leq \tag{B.24}$$

$$\leq ||H_{\alpha}||\,|\langle \mathbb{I}|\,\mathsf{T}_{\alpha}^{-1}|\rangle| \leq ||H_{\alpha}||\,||\mathsf{T}_{\alpha}^{-1}||_{1\to 1}\,||\,|\rangle\,||_{1} \leq \tag{B.25}$$

$$\leq 2||H_{\alpha}||\,||\mathsf{T}_{\alpha}^{-1}||_{1\to 1}\epsilon^{*}\,, \tag{B.26}$$

where first inequality follows from definition of operator norm, second inequality follows from a definition of any induced operator norm and the last inequality follows from analysis in previous section with ϵ^* being bound given by Eq. (B.19). In analogy to Eq. (B.22) by applying multiple times triangle inequality we obtain Eq. (4.19).

Approximation and statistical errors

Now we will combine two previous bounds by using a triangle inequality. Let us denote by $\mathbf{p}^{\text{noisy, est}}$ the estimator of noisy probability distribution $\mathbf{p}^{\text{noisy}}$ on the marginal α (hence the RHS of Eq. (B.23)), and by C^{α}_{av} the average correction matrix used to correct that marginal. We want to bound the distance between corrected estimator of noisy distribution $C^{\alpha}_{av}\mathbf{p}^{\text{noisy, est}}$ and the ideal distribution $\mathbf{p}^{\text{ideal}}$. For particular marginal we have triangle inequality for Total-Variation Distance

$$\text{TVD}\left(C^{\alpha}_{av}\mathbf{p}^{\text{noisy, est}}, \mathbf{p}^{\text{ideal}}\right) \leq \underbrace{\text{TVD}\left(C_{av}\mathbf{p}^{\text{noisy}}, C_{av}\mathbf{p}^{\text{noisy, est}}\right)}_{\leq 2||C_{av}||_{1\to 1}\epsilon^*} + \underbrace{\text{TVD}\left(C_{av}\mathbf{p}^{\text{noisy}}, \mathbf{p}^{\text{ideal}}\right)}_{2\delta_{alpha}} \tag{B.27}$$

$$\leq 2||C^{\alpha}_{av}||_{1\to 1}\epsilon^* + 2\delta_{\alpha}\,, \tag{B.28}$$

where ϵ^* is statistical bound given by Eq. (B.19), and δ_{α} is approximation error defined in Eq. (4.15). First underlined inequality follows from properties of induced operator norm and from statistical errors bound proved in the previous section (this step is analogous to the one in the second line of Eq. (B.24)). The second underlined inequality follows from Proposition 1. To translate the above result to expected values of local Hamiltonian one just repeats the reasoning given in proofs of Eq. (B.22) and Eq. (B.24) for marginal distributions of the form $C^{\alpha}_{av}\mathbf{p}^{\text{noisy, est}}$ for which we have bounds of the form given by Eq. (B.27).

To finish this section, let us note that the reasoning presented in this section is analogous to the one given in Ref. [42] where analysis of effects of statistical noise and non-classical noise on the error-mitigation performed on global distributions was presented.

B.2 Details of Diagonal Detector Overlapping Tomography

In this section, we give more details regarding our noise characterization procedure using DDOT. We start by providing efficient way of construction of DDOT circuits in Section. B.2.1. In Appendix B.2.2 we show how to use results of DDOT to infer the correlations structure of readout noise in a device. Then we discuss the construction of Diagonal Detector Overlapping Tomography circuits which are balanced (Appendix B.2.3) and perfect (Appendix B.2.4), together with proofs fro scaling of required number of random circuits. In Appendix B.2.5, we discuss heuristic procedures of making DDOT circuits more balanced. Finally, in Appendix B.2.6 we explain the effect of overestimating correlations which can happen if the implemented DDOT collection is not balanced.

B.2.1 Construction of Diagonal Detector Overlapping Tomography circuits

Here we will present in detail our building block for efficient characterization of correlations in a measurement device: Diagonal Detector Overlapping Tomography (DDOT). In analogy to Quantum Overlapping Tomography (QOT) [151], a collection of (N, k) DDOT circuits allows one to reconstruct the noise matrices for the readout process of any group of k qubits in an N-qubit device.

More precisely, we define a (N, k) perfect collection of DDOT circuits as a collection of N-qubit quantum circuits constructed only from $\mathbb{I}$ and X gates, with the property that for every subset of qubits of the size k, each of the computational basis states on that subset is prepared at least once. For example, we will call a collection of circuits $(N, 2)$ perfect if for all pairs of qubits, each of the states from $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ is prepared at least once in the whole collection.

One way to construct DDOT collection is to follow Ref. [151] and make use of the notion of hash functions (here by a hash function we mean every function $[N] \to [k]$ with $k < N$). To construct (N, k) perfect collection of DDOT circuits using hash functions one can use Algorithm 1, which encapsulates the idea of Quantum Overlapping Tomography from Ref. [151] translated to the construction of DDOT circuits. Specifically, this method corresponds to QOT with two "bases" which are preparation of state $|0\rangle$ or $|1\rangle$. Each hash function assigns each qubit a label from $[k]$. For a given function, qubits are assigned to k disjoint batches based on the value of the function value. For fixed assignment of batches, qubits belonging to a batch are initialized in the same state ($|0\rangle$ or $|1\rangle$), independently from the qubits in other batches. In this way, 2^k circuits are specified, which independently implement all computational basis states on all the batches. For example, in the case of $k = 2$ and $N = 6$ qubits, some specific hash functions could result in initializing the following states – $\{000000, 000111, 111000, 111111\}$, which implement all two-qubit computational basis states on the pairs of qubits from left and right parts of the register (in this example the two batches are $\{Q0, Q1, Q2\}$ and $\{Q3, Q4, Q5\}$). The DDOT circuits collection constructed in this way is perfect if the underlying collection of hash functions is perfect. This means that for *each* k-qubit subset in N-qubit device, there exists at least one hash function in the collection which assigns each qubit from that subset a distinct number from $[k]$. Note that if a given function is indeed injective on some k-qubit subset, this means that qubits from those subsets belong to distinct batches and therefore all computational-basis states will be implemented on them. Hence if this holds for all subsets, the collection is perfect. In Algorithm 1 we generate *random* hash functions to create a DDOT collection, therefore there is no deterministic guarantee that the collection constructed in this way will indeed be perfect. However, it follows directly from arguments in [151] (specifically, Section III) that if we use Algorithm 1 to generate DDOT collection, then if we wish the collection to be perfect with probability at least $1 - \delta$, the needed number circuits κ is at least of the order

$$\kappa > (2e)^k \left(\log(N) + \frac{1}{k} \log\left(\frac{1}{\delta}\right) \right), \tag{B.29}$$

see Appendix B.2.4 for detailed bounds.

On the other hand, one can also consider construction which uses random circuits without referring to the notion of hash functions. Generation of random bitstrings and using them as definitions of quantum circuits is used to construct DDOT collection in Algorithm 2. Similarly to Algorithm 2, since we use randomness to generate the collection, there is no deterministic guarantee that the resulting collection will be perfect. However, in Appendix B.2.4 we show that if we wish the collection to be perfect with probability at least $1 - \delta$, then the needed number of circuits κ is at least of the order

$$\kappa > 2^k \left(k \log (2N) + \log \frac{1}{\delta} \right), \tag{B.30}$$

which looks similar to Eq. (B.29), however in practice exhibits better scaling due to specific factors (see Appendix B.2.4 for more detailed bounds). Hence we expect that for higher system sizes the random circuits algorithm should perform better in practice compared to the one which uses hash functions. Of course, when the collection with desired properties has been constructed, it does not matter what method was used to create it.

As mentioned in the main text, during the implementation of DDOT, different circuits will in general be sampled a different number of times, this may cause some correlations to be overestimated (this effect can be reduced by proper post-processing of the data – see Appendix B.2.6). Hence it is beneficial for the perfect collection of DDOT circuits to be *approximately balanced*. Here by 'balanced' we mean that all k-qubit computational basis states are sampled the same number of times, and by 'approximately balanced' that they are sampled approximately the same number of times (see related notion for hash functions, for example in Ref [166]). For the construction using random circuits, we prove in Appendix B.2.3 that one can expect that with high probability the collection will also have this property with the required number of circuits scaling similarly to Eq. (B.30).

To conclude, let us point out that for a given pair of numbers (N, k) it suffices to generate the DDOT collection only once, and it can be used in the design of future experiments. We will make a number of pre-generated collections publicly available in our GitHub repository QREM (Quantum Readout Errors Mitigation) [197]. We note that the described techniques are suitable for noise characterization not only for the noise model we proposed but also for other models with bounded locality of correlations, for example with two-qubit correlations considered in Ref. [120].

B.2.2 Inferring the structure of clusters and neighborhoods

After the implementation of a (N, k) perfect collection of DDOT circuits, one has potential access to a lot of information about the measurement noise in a device. Here we provide a method of using data from such implementation to infer the correlations in measurement noise. As a starting point, let us note that one

can use the output of the DDOT circuits to construct all the possible two-qubit noise matrices $\mathsf{T}_{X_iX_j|Y_iY_j}$. In particular, whenever $k \geq 2$, there are subsets of circuits that implement all computational-basis states for each two-qubit subsystem. This allows to gather all two-qubit marginal probability distributions of the form

$$\left\{p\left(X_iX_j|Y_iY_j\right)\right\}_{i,j} \tag{B.31}$$

where X_i (X_j) is the measured state of ith (jth) qubit, and the Y_i (Y_j) is the input state of the corresponding qubit (i.e., the state that is supposed to be implemented by the quantum circuit from the DDOT collection). In the following Example 1 we give explicitly calculated single-qubit noise-matrices for exemplary pair of qubits $i = 1$ and $j = 2$.

Example 1. *For two-qubit set $S = \{1, 2\}$, assuming no dependence on the state of neighbors, generic left-stochastic map acting on its detector has a form*

$$\mathsf{T}^S = \begin{pmatrix} p\,(00|00) & p\,(00|01) & p\,(00|10) & p\,(00|11) \\ p\,(01|00) & p\,(01|01) & p\,(01|10) & p\,(01|11) \\ p\,(10|00) & p\,(10|01) & p\,(10|10) & p\,(10|11) \\ p\,(11|00) & p\,(11|01) & p\,(11|10) & p\,(11|11) \end{pmatrix}. \tag{B.32}$$

It follows that single-qubit noise matrices from Eq. (4.6) *for first qubit have form*

$$\mathsf{T}^{Y_2='0'} = \frac{1}{2}\begin{pmatrix} p\,(00|00) + p\,(01|00) & p\,(00|10) + p\,(01|10) \\ p\,(10|00) + p\,(11|00) & p\,(10|10) + p\,(11|10) \end{pmatrix}, \tag{B.33}$$

$$\mathsf{T}^{Y_2='1'} = \frac{1}{2}\begin{pmatrix} p\,(00|01) + p\,(01|01) & p\,(00|11) + p\,(01|11) \\ p\,(10|01) + p\,(11|01) & p\,(10|11) + p\,(11|11) \end{pmatrix} \tag{B.34}$$

$$\tag{B.35}$$

while for second qubit they are

$$\mathsf{T}^{Y_1='0'} = \frac{1}{2}\begin{pmatrix} p\,(00|00) + p\,(10|00) & p\,(00|01) + p\,(10|01) \\ p\,(01|00) + p\,(11|00) & p\,(01|01) + p\,(11|01) \end{pmatrix}, \tag{B.36}$$

$$\mathsf{T}^{Y_1='1'} = \frac{1}{2}\begin{pmatrix} p\,(00|10) + p\,(10|10) & p\,(00|11) + p\,(10|11) \\ p\,(01|10) + p\,(11|10) & p\,(01|11) + p\,(11|11) \end{pmatrix}. \tag{B.37}$$

Note that the idea is to simply fix the state of 'neighbouring' qubit to be either '0' or '1' and calculate according conditional marginals.

Now, we propose to use the information about the two-qubit noise matrices to calculate the correlations (Eq. (4.6)) between a given pair of qubits. For the i-th qubit, it's the dependence from the j-th qubit can be calculated by constructing

two single-qubit matrices on i-th qubit – the first one with the condition that j-th qubit was initialized in $|0\rangle$ state, and the second in $|1\rangle$ state.

Those matrices can be used to calculate the parameter $c_{j\to i}$ as the norm of a difference of those matrices (Eq. (4.6)). We then propose to infer the structure of the readout correlations according to the magnitude of the parameters $c_{j\to i}$ as follows. One specifies threshold parameters δ_{clust} and δ_{neighb} for the level of correlations between qubits, and then assigns qubit j to the neighborhood or to the cluster of qubit i, if the parameter $c_{j\to i}$ is greater than the respective threshold. In general, we advise to set those thresholds to be higher than the likely effects of statistical deviations. We outline the whole procedure in Algorithm 3, which takes as input the conditional single-qubit noise matrices $\left\{\mathsf{T}^{Y_j}_{Q_i}\right\}_{i\neq j}$ together with a set of thresholds, and outputs the structure of the clusters and neighborhoods in a device.

Finally, let us note that the above-described inference of correlations from two-qubit marginal distributions works under the assumption that the correlations do not vanish under taking marginals over other (than given pair) qubits. This does not need to be true in practice, and one can consider generalizations of Algorithm 3. The analogous set of single-qubit matrices depending on states of t neighboring qubits would be created in a fully analogous manner to that of Example 1 but now one would need to fix the state of t qubits. Note that if one implemented DDOT (N, k) collection, the data to create such matrices for $t = k-1$ is available from the experiments. In the future, we intend to investigate more elaborate methods of inferring correlations structure using DDOT, and we will accordingly expand our repository [197].

Once a model for the above structure is obtained, one can use the rest of the data obtained in the DDOT procedure to reconstruct the cluster noise matrices as a function of the state of the neighbors and consequently construct a global noise model (Eq. (4.3)), as well as the correction matrices for the marginals (Eqs. (4.8),(4.9)). From the definition of a perfect (N, k) collection, it follows that one can reconstruct only cluster noise matrices involving a number $t = |C_\chi|+|\mathbf{N}_\chi|$ of qubits which does not exceed k. If it happens that in fact $t > k$, we propose to implement one of two following solutions:

1. Neglect correlations between some qubits in order to enforce that $t = k$. For example, in Algorithm 3 if the number of qubits assigned to the neighborhood of some cluster exceeds the limit, one could decide to *not assign* to that neighborhood the qubits with the lowest values of $c_{j\to i}$ parameters. This will result in an imperfect model, which might nevertheless be accurate enough for the purposes of error mitigation.

2. Refine the noise model by performing additional experiments of standard Diagonal Detector Tomography on a chosen subset.

We note that for the refinement of the noise model, there also exists the alternative possibility of constructing a *restricted* DDOT collection that implements the characterization on a *specific* set of t-length subsets (as opposed to all such

Algorithm 1 Generation of a perfect collection of (N, k) DDOT quantum circuits using random hash functions

Input:
L: number of hash functions
N: number of all qubits
k: size of the marginal

=Start collection by creating two circuits which prepare all qubits in $|0\rangle$ state and in $|1\rangle$ state. Generate L random hash functions $\{f_1, \ldots, f_L\}$, i.e., random mappings $f_l : [N] \rightarrow [k]$. Define k-qubit sub-register as a set of all bit-strings $\left\{\mathbf{X}^1, \ldots, \mathbf{X}^{2^k}\right\}$ of length k. **For** each function f_l, **do**: **For** each bitstring $\mathbf{X}^i$ in k-qubit register, **do**: Define string $\mathbf{Y}$ of length N in the following way:

$$(\mathbf{Y})_{\mathrm{j}} = \left(\mathbf{X}^i\right)_{f_l(\mathrm{j})}$$

Save $\mathbf{Y}$ as definition of one quantum circuit in the DDOT collection – each symbol '0' corresponds to identity gate, each '1' symbol corresponds to NOT gate. Check if generated set of $s = 2 + L(2^k - 2)$ bitstrings $\{\mathbf{Y}\}$ contains all combinations of k-length subsets of symbols '0' and '1'.**If** yes: family $\{\mathbf{Y}\}$ defines perfect DDOT collection,**If** not: set $L \rightarrow L + 1$, generate new random hash function, add it to the collection and perform steps 2, 3 and 4.

Algorithm 2 Generation of (N, k) perfect collection of DDOT quantum circuits using random circuits

Input:
s: number of circuits
N: number of all qubits
k: size of the marginal

=Start collection by creating two circuits which prepare all qubits in $|0\rangle$ state and in $|1\rangle$ state. Generate s random bitstrings of size N. Each bitstring is a definition of quantum circuits – symbol '0' corresponds to identity gate, and symbol '1' symbol corresponds to NOT gate. Check if generated set of s bitstrings $\{\mathbf{Y}\}$ contains all combinations of k-length subsets of symbols '0' and '1'.**If** yes: family $\{\mathbf{Y}\}$ defines perfect DDOT collection,**If** not: set $s \rightarrow s + 1$, generate new random bitstring, add it to the collection and check new collection.

Algorithm 3 Assignment of qubits to clusters and neighborhoods

Input:
$\mathcal{S}_{\text{pairs}}$: set of all pair indices without repetitions of indices
δ_{clust}: threshold of correlations for assignment to clusters
δ_{neighb}: threshold of correlations for assignment to neighborhoods
$\left\{\mathrm{T}^{Y_j}_{X_i|Y_i}\right\}_{i,j}$: collection of single-qubit noise matrices depending on state of single neighbour for all pairs.

for (i,j) in $\mathcal{S}_{\text{pairs}}$ **do**
Calculate $c_{j\to i}$ and $c_{i\to j}$ as

$$c_{j\to i} = \frac{1}{2}||\mathrm{T}^{Y_j='0'}_{Q_i} - \mathrm{T}^{Y_j='1'}_{Q_i}||_{1\to 1}\,, \quad c_{i\to j} = \frac{1}{2}||\mathrm{T}^{Y_i='0'}_{Q_j} - \mathrm{T}^{Y_i='1'}_{Q_j}||_{1\to 1}\,,$$

where Y_j denotes input state of qubit j (see Eq. (4.6)).

if $c_{j\to i} > \delta_{\text{clust}}$ **or** $c_{i\to j} > \delta_{\text{clust}}$ **then**
Assign qubits i and j to the same cluster.

else
if $c_{j\to i} > \delta_{\text{neighb}}$ **then**
Assign qubit j to the neighborhood of qubit i.
if $c_{i\to j} > \delta_{\text{neighb}}$ **then**
Assign qubit i to the neighborhood of qubit j.

subsets). We will analyze this problem in more detail in future works, as well as during the development of our GitHub repository [197].

To finish this subsection, we note that Algorithm 3, while being straightforward and efficient, is not a flawless method. For example, it might happen that the noise on qubit i highly depends on the joint state of j and l, but this dependence is much lower if one considers qubit j and qubit l separately. In that case, Algorithm 3 might not assign the three qubits to the same cluster/neighborhood, even though they are correlated. A way to overcome the above limitation is to consider s generalizations of Algorithm 3 that makes use of three-qubit parameters "$c_{jl\to i}$", hence requiring a $k \geq 3$ DDOT procedure. This might be particularly useful for refining dependencies between clusters which were reported by the original Algorithm 3.

B.2.3 Constructing balanced DDOT collections

Interestingly, the tools utilized in the previous sections for the analysis of statistical deviations can be used to estimate the probability that a given collection of DDOT circuits will be approximately balanced. Recall that a perfect collection of (N, k) DDOT circuits is a collection of N-qubit quantum circuits consisting of $\mathbb{I}$ and X gates with a property that for each k-qubit subset every k-qubit computational basis state is implemented at least once in the whole collection. If the collection is *balanced*, it means that additionally each basis state is sampled the same number of times.

Consider the randomized construction of such collection – take circuits which are uniformly random combinations of 0 and 1 symbols ('0' corresponding to the identity gate and '1' to the NOT gate). This can be constructed efficiently since it suffices to choose each of the N bits at random independently. Such construction can be viewed as sampling from a 2^N-dimensional uniform distribution (corresponding to all possible N-bit strings describing possible circuits). The k-bit marginal distributions obtained from this distribution correspond to the distribution of local circuits on k-qubit subsets (i.e., the noise characterization by implementing computational-basis states on k-qubit subsets). Having this perspective, we can formulate the problem of approximate balancing of DDOT family in the following way: What is the probability that, when sampling from the global 2^N-dimensional uniform distribution, all of the k-bit marginals (corresponding to the subsets of interest in DDOT) will be at most ϵ-distant from the uniform distributions (on 2^k-dimensional space)?

Now we can use tools of statistical analysis presented in the previous sections. Our single marginal distribution has 2^k outcomes, hence from Eq. (B.16) we get that the probability of a single marginal being *at least* ϵ-distant in TVD is bounded by

$$P_1 \leq 2^{2^k} \exp\left(-2s\epsilon^2\right) , \tag{B.38}$$

where s is here the number of random circuits in the collection (viewed as samples from the global uniform distribution of bitstrings). Since there is $\binom{N}{k}$ number

of k-bit marginals, applying the union bound (as in the derivations in the previous sections) gives the upper bound for the probability $P_{N,k}$ that at least one of the $\binom{N}{k}$ marginals is at least ϵ-distant from uniform distributions as

$$P_{N,k} \leq \binom{N}{k} 2^{2^k} \exp\left(-2s\epsilon^2\right) . \tag{B.39}$$

Hence with probability *at least* $1-\binom{N}{k}2^{2^k}\exp\left(-2s\epsilon^2\right)$ all marginals are *at most* ϵ-distant from uniform distribution. Note that $\epsilon = 0$ corresponds to a perfectly balanced family of circuits, and small but non-zero ϵ will correspond to the approximately balanced family.

Let us now choose some parameter δ as an upper bound for probability Eq. (B.39) and find the bound for the required number of random circuits s (viewed as samples from a global uniform distribution) which are needed to obtain a family for which each marginal is distant from the uniform distribution by at most ϵ. After basic manipulations of Eq. (B.39) we obtain that

$$s \geq \frac{2^k \log 2 + \log\binom{N}{k} + \log\left(\frac{1}{\delta}\right)}{2\epsilon^2} \approx \frac{2^k + k\log N + \log\frac{1}{\delta}}{2\epsilon^2} . \tag{B.40}$$

B.2.4 The efficiency of the random construction of DDOT collection

Random circuits

Adopting the perspective from the previous section, we can in a simple manner tackle the problem of bounding the required number of random circuits which are needed to obtain a perfect collection of $\binom{N}{k}$ DDOT circuits. Consider randomly sampling s number of bit-string of length N. Now for a fixed k-element subset, the probability of a particular k-element combination *not appearing* is $1-\frac{1}{2^k}$. Hence after s samples, there is $\left(1-\frac{1}{2^k}\right)^s$ probability that this particular combination did not appear. Since there are $2^k\binom{N}{k}$ combinations of interest (i.e., 2^k small k-length bitstrings for all $\binom{N}{k}$ subsets), we can use the union bound to obtain

$$2^k\binom{N}{k}\left(1-\frac{1}{2^k}\right)^s \approx 2^k\binom{N}{k}\exp\left(-\frac{s}{2^k}\right) \tag{B.41}$$

as the upper bound for the probability that *at least* one k-length bit-string did not appear after s samples. This means that with probability of at least $1-2^k\binom{N}{k}\exp\left(-\frac{s}{2^k}\right)$, all of the k-length bitstrings appeared.

Let us now choose some parameter δ as the upper bound for Eq. (B.41) and calculate the resulting bound for the number of samples (i.e., random circuits). After basic manipulations we obtain

$$s > 2^k\left(k\log 2 + \log\binom{N}{k} + \log\frac{1}{\delta}\right) \approx 2^k\left(k\log 2N + \log\frac{1}{\delta}\right) . \tag{B.42}$$

Random hash functions

Using simple combinatorial arguments, in Ref. [151] it is shown that the probability of the collection of $\binom{N}{k}$ *random hash functions* not being perfect (for our purposes perfect hash function collection means that it can be used to construct a perfect DDOT collection) is upper bounded by

$$\binom{N}{k}\left(1-\frac{k!}{k^k}\right)^L, \tag{B.43}$$

where L is the number of generated random hash functions. Then the authors simplify the use of the above fact to derive a bound on the needed number of hash functions for the collection to be perfect. Let us repeat the derivation from Ref. [151] with paying attention to specific factors. From Algorithm 1 it follows that the number of circuits s obtained from a given hash function collection is equal to $s = 2 + \left(2^k - 2\right) L$. Now by choosing the parameter δ as upper bound on Eq. (B.43), we obtain that required number of circuits is lower bounded by

$$\begin{aligned}
s &\geq 2 + \left(2^k - 2\right) \frac{-\log\binom{N}{k} + \log\delta}{\log\left(1 - \sqrt{2\pi k}\, e^{-k}\right)} \\
&\geq 2 + \left(2^k - 2\right) \frac{-\log\binom{N}{k} + \log\delta}{-\sqrt{2\pi k}\, e^{-k} - 2\pi k e^{-2k}} \\
&\geq 2 + \left(2^k - 2\right) \frac{-\log\binom{N}{k} + \log\delta}{-\sqrt{2\pi}\, k\, e^{-k}} \\
&\approx 2^k e^k \left(\frac{1}{k}\log\frac{1}{\delta} + \log N\right),
\end{aligned}$$

where we used that $\log(1-x) \geq -x - x^2$. Let us note that Eq. (B.44) is a reiterated result from Ref. [151].

In above derivations we assumed that $\left(-\log\binom{N}{k} + \log\delta\right) < 0.$, then we utilized the fact that

$$\sqrt{2\pi}\, k^{k+\frac{1}{2}} e^{-k} \leq k! \leq e\, k^{k+\frac{1}{2}} e^{-k}, \tag{B.44}$$

and we used approximation $\binom{N}{k} \approx N^k$.

B.2.5 Heuristic balancing of DDOT collection

After generating a perfect DDOT collection, one can be interested in making it more balanced. There are various possible figures of merit that can be used to quantify the "balancing" of the family. For example, in previous sections, we used TVD between a uniform distribution and generated sample, when viewing obtained circuits in the collection as samples from the uniform distribution. Another possibility is to calculate a number of appearances of each marginal term

in the whole collection (for example, for 2-qubit subsets the number that each of 00, 01, 10, and 11 appeared, for every 2-qubit subset), and then take an empirical standard deviation $\sigma_{n,k}$ of this quantity. A perfectly balanced family would have 0 standard deviation defined in that way.

Now we discuss a simple heuristic method to improve balancing. The starting point is a perfect (N, k) DDOT collection. Now we apply the following steps in a loop.

1. Calculate a number of appearances of all k-qubit terms in the collection.
2. Find the set of non-overlapping k-bit marginals which appear in the collection the least number of times compared to the whole population. If k does not divide N, choose $\lfloor \frac{N}{k} \rfloor$ subsets and add random gates to the remaining bits.
3. Add circuit which implements those least-appearing marginals to the collection.

For example, say that in the last step of the above procedure for $k = 2$ and $N = 5$ we added circuit

$$01100\,. \tag{B.45}$$

This might mean that we found that the least appearing marginal is state 01 on qubits Q_0 and Q_1, the marginal 10 on qubits Q_2 and Q_3 appeared the same or second-least number of times in the whole collection, while the 0 on last qubit Q_4 was random. In this way, we are adding missing marginal terms "by hand" at each step of the loop. Clearly, this procedure is heuristic and it is not guaranteed to succeed since by adding certain circuits that implement desired marginals we also implement other marginals (in the example above, we add, e.g., marginal 00 on qubits Q_0 and Q_3). If it happens that those other marginals are in the opposite "tail" of the whole distribution (i.e., they are the most abundant ones), then the collection will not become more balanced (it can actually become less balanced). However, on average it is more likely that by doing so we add marginals that are closer to the "average marginals" (i.e., those which appear a number of times close to the mean appearance number).

Clearly, a lot of practical refinements of the described method are possible – for example, in the second step of the method, instead of adding the least appearing marginals one can focus on maximizing the number of low-appearance marginals. In practice, we found that the described method without modifications usually reduced $\sigma_{N,k}$ with a growing number of circuits added in the loop.

To conclude, let us note that the special case of the above technique can be also used to add circuits to the non-perfect collection of DDOT circuits in order to make it closer to being perfect. Namely, if the collection is not perfect, this means that it does not implement some computational-basis states on some k-qubit subsets. Hence in that case the above procedure would report "the least appearing marginals" to be those which are missing in the collection and it would

keep adding them in a loop until there are no missing terms – a strategy which might turn out to better then adding random circuits if the number of missing terms is not too big.

B.2.6 Overestimation of correlations

In the main text, we explained that when considering the implementation of DDOT circuits, some two-qubit correlations might be overestimated due to the fact the collection of circuits is not balanced. To understand this effect, let us now consider the following illustrative (but rather unrealistic) example.

Explanation of the effect and post-processing strategy

Say that we want marginal probability distribution on qubits Q_0 and Q_2 when input state was $|00\rangle$. Then if noise is *uncorrelated* it does not matter whether global input state on three qubits was $|000\rangle$ or $|010\rangle$ ($|001\rangle$ or $|011\rangle$). However, when there are correlations, it might happen that those two distributions *will be different* depending on the input state of Q_1. When we marginalize over Q_1 we forgot about "where the data came from". Normally, we would just add the marginal probability from circuits implementing both global states and then normalized it. But if global state $|000\rangle$ was implemented a different number of times than $|010\rangle$, it can cause that effectively correlations which are caused by Q_1 are wrongly identified as correlations between Q_0 and Q_2, because some global probability distributions contribute to the marginal with higher weights (i.e., are effectively counted more times when calculating marginal distributions). As mentioned in the main text, the natural way to reduce such effects is to create a collection that is balanced, hence it samples from all two-qubit states the same number of times. The other thing one can do is to perform post-processing of the experimental data in such a way, that all contributions to the given marginal are weighted by the inverse of a number of times they were implemented. Note that they come from different global distributions, so what is important is the number of times the given marginal state was implemented together with some specific state on all the other qubits. Importantly, this method is not perfect because for big systems the "state on all the other qubits" will likely be different each time anyway (this is due to the fact that collection of DDOT circuits will be random, hence it becomes quite unlikely to obtain two times the same bitstring if the number of qubits is high). Another thing one can do is to change the weights of the given contribution to the marginal distribution depending on the state of some particular subset of qubits in order to assess whether inferred correlations were correct. Specifically, one might perform a recursive procedure in which the structure of clusters and neighborhoods inferred from non-post-processed data is validated on particular subsets using this type of post-processing. There are certainly a lot of practical possibilities to improve the post-processing scheme and we intend to investigate them in future research.

Illustration of the effect

To have some idea how big a described effect can be, let us consider implementation of the following collection of circuits on three qubits $\{Q_0, Q_1, Q_2\}$

$$\{000, 001, 001, 010, 011, 100, 101, 110, 111\}\,,$$

where each 0 corresponds to implementation of identity gate and each 1 to the NOT gate. Note that considered collection is clearly abundant for 3-qubit characterization since it contains more than $2^3 = 8$ circuits. One can think of the above circuits as a part of a collection of DDOT circuits on a higher number of qubits, while we only look at a specific triple of qubits. Now let us assume that that the qubits Q_0 and Q_2 are completely *uncorrelated* in terms of readout noise. On the other hand, Q_1 and Q_2 are highly correlated in a following way

$$Q_1 = |0\rangle \implies \text{do nothing}\,, \tag{B.46}$$
$$Q_1 = |1\rangle \implies \text{apply bitflip to } Q_2\,. \tag{B.47}$$

Now if one takes outcomes from circuits from the collection, add them up, normalize and consider resulting estimators of probabilities of obtaining different outcomes after implementation of collection from Eq. (B.46) *marginalized* over qubit 1, it follows from direct computation that the noise matrix on Q_0 and Q_2 is of the form

$$\mathrm{T}^{Q_0Q_2} = \begin{pmatrix} p(00|00) & p(00|01) & p(00|10) & p(00|11) \\ p(01|00) & p(01|01) & p(01|10) & p(01|11) \\ p(10|00) & p(10|01) & p(10|10) & p(10|11) \\ p(11|00) & p(11|01) & p(11|10) & p(11|11) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{3} & \frac{1}{2} \end{pmatrix}, \tag{B.48}$$

from which we obtain that the noise matrices on Q_0 depending on state of Q_2 are

$$\mathrm{T}^{Y_2='0'} = \begin{pmatrix} p(00|00) + p(01|00) & p(00|10) + p(01|10) \\ p(10|00) + p(11|00) & p(10|10) + p(11|10) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{B.49}$$

$$\mathrm{T}^{Y_2='1'} = \begin{pmatrix} p(00|01) + p(01|01) & p(00|11) + p(01|11) \\ p(10|01) + p(11|01) & p(10|11) + p(11|11) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{B.50}$$

(B.51)

and on Q_2 depending on state of Q_0

$$\mathrm{T}^{Y_0='0'} = \begin{pmatrix} p(00|00) + p(10|00) & p(00|01) + p(10|01) \\ p(01|00) + p(11|00) & p(01|01) + p(11|01) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \tag{B.52}$$

$$\mathrm{T}^{Y_0='1'} = \begin{pmatrix} p(00|10) + p(10|10) & p(00|11) + p(10|11) \\ p(01|10) + p(11|10) & p(01|11) + p(11|11) \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{1}{2} \\ \frac{2}{3} & \frac{1}{2} \end{pmatrix}. \tag{B.53}$$

The above matrices give correlation factors (Eq. (4.6)

$$c_{2\to 0} = 0 \tag{B.54}$$

$$c_{0\to 2} = \frac{1}{6}\,. \tag{B.55}$$

Hence the Algorithm 3 would report that the noise on Q_2 significantly depends on the state of Q_0 even though physically there is no dependence. The correlations between Q_1 and Q_2 give rise to false correlations between Q_0 and Q_2 after marginalizing over Q_1. This is solely due to the fact that in Eq. (B.46) different three-qubit states are sampled different numbers of times, which gives some contributions to the marginals have higher weight when constructing effective noise matrix on Q_0 and Q_2.

B.3 Noise characterization scheme overview

Here we provide a step-by-step description of our noise characterization procedure. We note that stages 0 and 1 were only briefly mentioned in the main text, and they correspond to the verification of the undertaken assumptions: stage 0 verifies the quality of the single-qubit gates, and stage 1 the assumption about classical nature of the noise in the measurement device. Stages 2 and 3 describe the proper characterization scheme of Diagonal Detector Overlapping Tomography which was discussed in detail in the main part of the work and in Appendix B.2.

B.3.1 Stage 0 – single-qubit gate fidelities

To begin, let us note that of our characterization procedure relies on the assumption of perfect state preparation. However, in practice, this assumption might be significantly violated. Therefore we propose not to use qubits with single-qubit gate infidelities above some threshold – in our experiments we arbitrarily chose this threshold to be 0.01.

In experiments on IBM's *Melbourne* backend, the single-qubit gate fidelities were good enough (fidelities above 99%) to use all of the qubits. In the case of Rigetti's *Aspen-8*, we discarded 8 qubits which had fidelities below 98%, while still using qubits 5 qubits which had fidelity in range $[98\%, 99\%]$.

B.3.2 Stage 1 – assessing classical form of the noise

In order to perform simultaneous estimation of single-qubit detectors with an overcomplete operator basis, one needs to implement 6 different circuits – each implementing eigenstate of a different Pauli matrix on every qubit at the same time. After this implementation, one needs to post-process data to obtain marginal single-qubit distributions and use standard detector tomography algorithms, for example, those described in [159] (and implemented in Python in online repository [197]).

Having reconstructed POVMs describing each single-qubit detector in a device, one can assess the classicality of the noise using methods described in Ref. [42] – for the sake of completeness, we will recall the main notions of that procedure here. First, we will need a notion of distance between quantum measurements. Such distances are usually related to the probability distributions that those measurements generate via Born's rule. Recall from Eq. (4.7) that the Total Variation Distance between probability distributions $\mathbf{p}$ and $\mathbf{q}$ as L1 norm of difference of those vectors

$$\mathrm{D}_{\mathrm{TV}}(\mathbf{p},\mathbf{q}) = \frac{1}{2}||\mathbf{p}-\mathbf{q}||_1 = \frac{1}{2}\sum_i |p_i - q_i| \,. \tag{B.56}$$

Now, the distance between quantum measurements related to TVD is the operational distance D_{op} defined for two POVMs $\mathbf{M}$ and $\mathbf{P}$ as [118]

$$\mathrm{D}_{\mathrm{op}}(\mathbf{M},\mathbf{P}) = \max \mathrm{D}_{\mathrm{TV}}(\mathbf{p_M},\mathbf{q_P}) \,, \tag{B.57}$$

where $\mathbf{p_M}$ (or $\mathbf{q_P}$) is a probability distribution generated by measurement $\mathbf{M}$ (or $\mathbf{P}$) via Born's rule, and the maximization goes over all quantum states. Hence, the operational distance between two quantum measurements is the worst-case distance between probability distributions they can generate. Now, to quantify readout noise one can simply calculate[2] operational distance for reconstructed POVM $\mathbf{M}$ for each qubit and the ideal measurement $\mathbf{P}$. Then, to quantify coherent part of the noise, one can make the following decomposition [42]

$$\mathbf{M} = \underbrace{\mathrm{T}\mathbf{P}}_{\text{classical part}} + \underbrace{\Delta}_{\text{coherent part}} \,. \tag{B.58}$$

For the ideal measurement $\mathbf{P}$ being the standard measurement in the computational basis, this decomposition is straightforward – the classical part of the noise is contained in the diagonal part of the measurement operators, while off-diagonal terms are a coherent part. The magnitude of the coherent part of the noise can be quantified as $\mathrm{D}_{\mathrm{op}}(\mathbf{M}, \mathrm{T}\mathbf{P})$. The assumption of fully classical noise leads to discarding the coherent part Δ in the Eq. (B.58) and performing error-mitigation *as if* POVM $\mathrm{T}\mathbf{P}$ was exact description of the detector. This leads to the propagation of coherent errors under error-mitigation, and it can be quantified via $||\mathrm{T}^{-1}||_{1\to 1}\mathrm{D}_{\mathrm{op}}(\mathbf{M}, \mathrm{T}\mathbf{P})$. Following guidelines in Ref. [42], we propose to *discard* every qubit which fulfills the following inequality

$$||\mathrm{T}^{-1}||_{1\to 1}\mathrm{D}_{\mathrm{op}}(\mathbf{M}, \mathrm{T}\mathbf{P}) \geq \mathrm{D}_{\mathrm{op}}(\mathbf{M},\mathbf{P}) \,. \tag{B.59}$$

Fortunately, in experiments on both IBM's and Rigetti's machines, this step did not lead us to discard any qubits – the noise in those devices remains highly classical, as indicated in previous experiments [42; 41].

Before proceeding further, let us note that while assessing classicality of the noise via single-qubit QDTs we make the following implicit assumption – the

[2] See Refs. [118; 42] on practical calculation of RHS of Eq. (B.57).

coherent part of the noise does not scale significantly with growing system size (by this we mean that it is at most additive in the number of qubits). Furthermore, in using the rule Eq. (B.59) we disregard effects of statistical deviations.

B.3.3 Stage 2 – Diagonal Detector Overlapping Tomography

The main idea of DDOT was described in the main text. The practical generation of relevant circuits was described in Appendix B.2.1 and summarized in Algorithm 1 (using random hash functions) and in Algorithm 2 (using random circuits).

B.3.4 Stage 3 – inferring correlations structure

The procedure of inferring correlations (i.e., structure of clusters and neighbourhoods) was described in detail in Appendix B.2.2 and summarized in Algorithm 3.

B.4 Sample complexity of energy estimation

In this section, we give some derivations related to the estimation of expectation values of local Hamiltonian terms on various quantum states. We start by discussing correlations in random states in Appendix B.4.1. Then in Appendix B.4.2, we prove Proposition 2 from the main text, which concerns states appearing in the QAOA algorithms. Finally, in Appendix B.4.3 we analyze what happens with the covariances of local Hamiltonian terms if the uncorrelated measurement noise and its mitigation are present.

B.4.1 Local correlations in random states

Proposition 4. *Let $|\psi\rangle$ be a pure state on $(\mathbb{C}^2)^{\otimes N}$, for a subset $\gamma \subset [N]$ of qubits we denote by ρ_γ and id_γ the marginal of $|\psi\rangle\langle\psi|$ corresponding to γ and the maximally mixed state on γ, respectively. Let H_α, H_β be local hamiltonians term that act on disjoint substets of qubits. We then have*

$$\mathrm{Cov}(H_A, H_B) \leq 3\|H_A\|\|H_B\|\|\rho_{\alpha\cup\beta} - \mathrm{id}_{\alpha\cup\beta}\|_1. \tag{B.60}$$

Proof. A simple algebra gives

$$\begin{aligned}\mathrm{Cov}(H_\alpha, H_\beta) &= \mathrm{tr}\left(H_\alpha \otimes H_\beta(\rho_{\alpha\cup\beta} - \rho_\alpha \otimes \rho_\beta)\right)\\ &= \mathrm{tr}\left(H_\alpha \otimes H_\beta(\Delta_{\alpha\cup\beta} - \Delta_\alpha \otimes \rho_\beta + \mathrm{id}_\alpha \otimes \Delta_\beta)\right)\ ,\end{aligned} \tag{B.61}$$

where $\Delta_\gamma = \rho_\gamma - \mathrm{id}_\gamma$. We now apply the well-known inequality

$$\mathrm{tr}(X) \leq \|A\|\|X\|_1\ , \tag{B.62}$$

for $A = H_\alpha \otimes H_\beta$ and $X = \Delta_{\alpha\cup\beta} - \Delta_\alpha \otimes \rho_\beta - \mathrm{id}_\alpha \otimes \Delta_\beta$. The 1-norm can be upper bounded as follows:

$$\|\Delta_{\alpha\cup\beta} - \Delta_\alpha \otimes \rho_\beta + \mathrm{id}_\alpha \otimes \Delta_\beta\|_1 \leq \|\Delta_{\alpha\cup\beta}\|_1 + \|\Delta_\alpha\|_1 + \|\Delta_\beta\|_1 \leq 3\|\Delta_{\alpha\cup\beta}\|_1\ , \tag{B.63}$$

where we used the following properties of 1-norm: triangle inequality, multiplicativity: $\|A\otimes B\|_1 = \|A\|_1\|B\|_1$, and data-processing inequality [198]. We conclude the proof by inserting (B.63) into (B.62) and using $\|H_\alpha \otimes H_\beta\| = \|H_\alpha\|\|H_\beta\|$. □

B.4.2 Proof of Proposition 2

Consider Hamiltonian with connectivity given by Erdös-Rényi random graph with N nodes and K edges with average degree equal to $q = \frac{K}{N}$. We will show that if the number of levels satisfies $p = c\log(N) + 1$ with $c \leq \frac{1}{2\log(2q/\ln 2)}$, then the variance of the energy $\mathrm{Var}(\mathcal{H})$ will scale as $O(N^{2-x})$, with $x > 0$ depending on c.

Let $G = (V, E)$ be a graph (with vertex set V and edge set E), we denote by $B(i, r)$ the set of vertices that are in graph distance r or less from vertex i. Similarly, for an edge $\alpha = (i, j) \in E$, we define $C(\alpha, r)$ the set of vertices that are in graph distance r or less away from α. For $\alpha = (i, j)$ and $\beta = (v, w)$, we have that if $v \notin B(i, r+2)$, then $\beta \notin C(\alpha, r)$.

The random Hamiltonian corresponding to a graph can be written as sum of 2-qubit terms (with also single qubit terms incorporated into these) corresponding to the edges E of the random graph $G(N, q)$, i.e., $H = \sum_{(i,j)\in E} H_{(i,j)}$. The variance is then

$$\mathrm{Var}(\mathcal{H}) = \sum_{\alpha,\beta\in E} \mathrm{Cov}\left(H_\alpha, H_\beta\right) . \tag{B.64}$$

To bound this quantity we will utilize the following two facts. First, we notice that than any non-zero term in Eq. (B.64) can be bounded by

$$\mathrm{Cov}\left(H_\alpha, H_\beta\right) \leq \mathrm{Var}\left(H_\alpha\right)\ \mathrm{Var}\left(H_\beta\right) \leq ||H_\alpha||\ ||H_\beta||, \tag{B.65}$$

where we used known covariance-variance inequality together with Popoviciu's inequality.

Second, we reiterate an important observation from Ref. [171] about these types of QAOAs: Consider two operators O_1 and O_2 acting non-trivially only on the sets of qubits $A_1 \subset V$ and $A_2 \subset V$, respectively. If U is a unitary corresponding to a p-level QAOA (with any parameter setting), the set of nodes A_1 and A_2 is in graph distance at least $2p$ distance from each other and $|\psi\rangle$ is product state, then

$$\langle\psi| U^\dagger O_1 O_2 U |\psi\rangle = \langle\psi| U^\dagger O_1 U |\psi\rangle \langle\psi| U^\dagger O_2 U |\psi\rangle . \tag{B.66}$$

With these two ingredients, we can bound the variance as

$$\begin{aligned}
\mathrm{Var}(\mathcal{H}) &= \sum_{\alpha,\beta\in E} \mathrm{Cov}\left(H_\alpha, H_\beta\right) \\
&= \sum_{\alpha\in E} \sum_{\beta\in C(\alpha,2p)} \mathrm{Cov}\left(H_\alpha, H_\beta\right) \\
&\leq \sum_{\alpha\in E} \sum_{\beta\in C(\alpha,2p)} ||H_\alpha||\ ||H_\beta|| .
\end{aligned} \tag{B.67}$$

An even more rough upper bound can be given by

$$\mathrm{Var}\,(\mathcal{H}) \leq f_{\mathcal{H}} \sum_{\alpha\in E} \max_{\alpha\in E} |\mathrm{C}(\alpha, 2p)| = f_{\mathcal{H}}\ K\ \max_{\alpha\in E} |\mathrm{C}(\alpha, 2p)|\,, \tag{B.68}$$

with setting

$$f_{\mathcal{H}} = \max_{\alpha,\beta} ||H_\alpha||\ ||H_\beta|| \tag{B.69}$$

Now we can turn to the concrete case of Hamiltonians corresponding to graph sampled from the Erdös-Rényi graphs with average degree q. For such graphs the *Neighborhood Size Theorem* states the following [171]:
For any

$$r < \frac{w\log(N)}{4\log(2q/\ln(2)}, \tag{B.70}$$

where $0 < w > 1$, there exists a constants $a > 0$ and $A < 1$ such that

$$\mathrm{Prob}\Big[\max_{i\in V} |\mathrm{B}(i, r)| \geq N^{A/2}\Big] \leq e^{-N^{a/2}}. \tag{B.71}$$

To be more specific, we can give the expression of a and A in terms of r and N:

$$A = w\,\frac{(2 + |\log_{2q}(\ln(2))|)}{(1 + |\log_{2q}(\ln(2))|)}, \tag{B.72}$$

$$a = \frac{w}{3(1 + |\log_{2q}(\ln(2))|}\,. \tag{B.73}$$

The proof of the above comes from the proof of Neighborhood Size Theorem given in [171], page 13. Specifically our inequalities correspond to setting $s = 1$.

Now note that since for any $\alpha = (i, j) \in E$ and $\beta = (v, w) \in E$, we have that if $\beta \in \mathrm{C}(\alpha, r)$ then $v \in \mathrm{B}(i, r + 2)$. This immediately implies that generally for $\alpha = (i, j) \in E$ we have $|\mathrm{C}(\alpha, r)| \leq |\mathrm{B}(i, r + 2)|^2$, and thus also

$$\mathrm{Prob}\Big[\max_{\alpha\in E} |\mathrm{C}(\alpha, r-2)| \geq N^A\Big] \leq \mathrm{Prob}\Big[\max_{\alpha\in E} |\mathrm{B}(\alpha, r)|^2 \geq N^A\Big] = \mathrm{Prob}\Big[\max_{\alpha\in E} |\mathrm{B}(\alpha, r)| \geq N^{A/2}\Big] \leq \tag{B.74}$$

If we now choose $2p = r - 2$, and thus

$$p < \frac{w\log(N)}{8\log(2q/\ln(2))} - 1, \tag{B.75}$$

then by combining Eq. (B.74) with bound in Eq. (B.68), we get that with probability at least $1 - e^{-N^{a/2}}$ variance of the Hamiltonian is bounded by

$$\mathrm{Var}\,(\mathcal{H}) \leq f_{\mathcal{H}}\ K\ \max_{\alpha\in E} |\mathrm{C}(\alpha, r)| \leq f_{\mathcal{H}}\ K\ N^A = f_{\mathcal{H}}\ q\ N^{A+1} \tag{B.76}$$

which is statement of Proposition 2.

B.4.3 Covariances of local terms in presence of uncorrelated readout noise

Presence of uncorrelated readout noise

As a starting point, let us assume that we have a state with correlations bounded in trace norm

$$||\rho_{\alpha\beta} - \rho_\alpha \otimes \rho_\beta||_1 \leq \epsilon_{\alpha\beta}\,, \tag{B.77}$$

where $\rho_{\alpha\beta}$ is marginal quantum state on subsystems α and β which is close to a product state of those subsystems. The special instance $\epsilon = 0$ corresponds to the product state case. We are interested in what happens with the covariances of local Hamiltonian terms if the measurement is affected by uncorrelated classical noise of the form $\mathsf{T} = \bigotimes_i \mathsf{T}_{Q_i}$, where T_{Q_i} is noise matrix acting on qubit Q_i (see Eq. (4.15)). Similarly, as in the main text, the Hamiltonian we consider is classical (i.e., diagonal), and local terms can be decomposed into sums of products of Pauli σ_z terms. Let us denote by $\mathbf{p}_\alpha^{\text{noisy}} = \mathsf{T}_\alpha \mathbf{p}_\alpha^{\text{ideal}}$ a noisy marginal distribution on the qubit subset α corresponding to local Hamiltonian term H_α. Note that in the uncorrelated noise model $\mathsf{T}_\alpha = \bigotimes_{Q_i \in \alpha} \mathsf{T}_{Q_i}$. Similarly to Eq. (B.20), we can write the expectation value of the noisy local Hamiltonian term as

$$\left\langle \tilde{H}_\alpha \right\rangle_{\rho_\alpha} \coloneqq \langle \mathsf{T}_\alpha | \mathbf{p}_\alpha^{\text{noisy}} \rangle = \langle \mathsf{T}_\alpha | \mathsf{T}_\alpha \mathbf{p}_\alpha^{\text{ideal}} \rangle\,, \tag{B.78}$$

where T^α is a vectorized spectrum of term H_α, the subscript ρ_α indicates the marginal state on which the expectation value is calculated and we used a convenient braket notation to denote scalar product.

We are now interested in bounding the covariance $\mathrm{Cov}\left(H_\alpha^{\text{noisy}}, H_\beta^{\text{noisy}}\right)$ between two local Hamiltonian terms when measured in a quantum state from Eq. (B.77). Let us decompose a marginal quantum state as

$$\begin{aligned}\rho_{\alpha\beta} &= \rho_\alpha \otimes \rho_\beta + \Delta_{\alpha\beta} \\ \Delta_{\alpha\beta} &\coloneqq \rho_{\alpha\beta} - \rho_\alpha \otimes \rho_\beta\,.\end{aligned}$$

Now can write

$$\begin{aligned}\mathrm{Cov}\left(\tilde{H}_\alpha, \tilde{H}_\beta\right) &= \left\langle \tilde{H}_{\alpha\beta} \right\rangle_{\rho_{\alpha\beta}} - \left\langle \tilde{H}_\alpha \right\rangle_{\rho_\alpha} \left\langle \tilde{H}_\beta \right\rangle_{\rho_\beta} = \\ &= \left\langle \tilde{H}_{\alpha\beta} \right\rangle_{\Delta\alpha\beta} + \left\langle \tilde{H}_{\alpha\beta} \right\rangle_{\rho_\alpha \otimes \rho_\beta} - \left\langle \tilde{H}_\alpha \right\rangle_{\rho_\alpha} \left\langle \tilde{H}_\beta \right\rangle_{\rho_\beta} = \\ &= \underbrace{\langle \mathsf{T}_{\alpha\beta} | \mathsf{T}_{\alpha\beta} \mathbf{p}_{\Delta_{\alpha\beta}}^{\text{ideal}} \rangle}_{\text{correlated part}} + \underbrace{\langle \mathsf{T}_{\alpha\beta} | \mathsf{T}_{\alpha\otimes\beta} \mathbf{p}_{\alpha\beta}^{\text{ideal}} \rangle - \langle \mathsf{T}_\alpha | \mathsf{T}_\alpha \mathbf{p}_\alpha^{\text{ideal}} \rangle \langle \mathsf{T}_\beta | \mathsf{T}_\alpha \mathbf{p}_\beta^{\text{ideal}} \rangle}_{\text{uncorrelated part}}\,.\end{aligned}$$

In the above we slightly abused the notation – in general, α and β can overlap, in which case one needs to insert proper identities and accordingly redefine product state $\rho_\alpha \otimes \rho_\beta$ (together with bound in Eq. (B.77) which will now correspond to the different, more refined division of qubits) and corresponding noise matrices. We also denoted by $\mathbf{p}_{\Delta_{\alpha\beta}}^{\text{ideal}}$ a formal vector given by diagonal elements

of $\Delta_{\alpha\beta}$ (this corresponds to measurement in computational basis). Now note that in the last line of Eq. (B.79) the part underlined as "uncorrelated part" contains terms without any correlations except those which can appear if α and β overlap. In particular, if $\alpha \cap \beta = \varnothing$, then since the noise is uncorrelated, we have $\left\langle \mathsf{T}_{\alpha\beta} \right| = \langle \mathsf{T}_\alpha | \otimes \left\langle \mathsf{T}_\beta \right|$ and $\mathsf{T}_{\alpha\otimes\beta}\mathbf{p}^{\text{ideal}}_{\alpha\otimes\beta} = \left(\mathsf{T}_\alpha \mathbf{p}^{\text{ideal}}_\alpha\right) \otimes \left(\mathsf{T}_\beta \mathbf{p}^{\text{ideal}}_\beta\right)$, therefore then this part is equal 0. Otherwise, it gives a non-zero contribution, which however would be present even without any measurement noise.

The only part which adds non-trivial correlations is therefore $\langle \mathsf{T}_{\alpha\beta} | \mathsf{T}_{\alpha\beta}\mathbf{p}^{\text{ideal}}_{\Delta_{\alpha\beta}} \rangle$ which using Eq. (B.77) and elementary transformations can be bounded as

$$|\langle \mathsf{T}_{\alpha\beta} | \mathsf{T}_{\alpha\beta}\mathbf{p}^{\text{ideal}}_{\Delta_{\alpha\beta}} \rangle| \leq ||H_\alpha H_\beta||\ ||\mathsf{T}_{\alpha\beta}||_{1\to 1}\ ||\Delta_{\alpha\beta}||_1 \leq ||H_\alpha H_\beta||\ \epsilon_{\alpha\beta}\ . \tag{B.79}$$

In the last inequality we made use of the following facts. If α and β do not overlap, we have $||\mathsf{T}_{\alpha\beta}||_{1\to 1} = ||\mathsf{T}_\alpha \otimes \mathsf{T}_\beta||_{1\to 1} = ||\mathsf{T}_\alpha||_{1\to 1}\ ||\mathsf{T}_\beta||_{1\to 1} = 1$. In case α and β overlap, the non-overlapping parts will simply give stochastic matrices (and they have $1 \to 1$ norm equal to 1), while overlapping parts will be squared. We therefore obtain $||\mathsf{T}_{\alpha\beta}||_{1\to 1} = \prod_{i\in\alpha\cap\beta} ||\mathsf{T}_i^2||_{1\to 1} \leq \prod_{i\in\alpha\cap\beta} ||\mathsf{T}_i||^2_{1\to 1} = 1$, where we used submultiplicity of norm. Of course, if the noise model is known, the $||\mathsf{T}_{\alpha\beta}||_{1\to 1}$ could be also calculated explicitly.

From the above, it follows that if $\epsilon_{\alpha\beta}$ is small, then under uncorrelated measurement noise the covariances between local terms will be, unsurprisingly, small as well.

Effects of error-mitigation

To include error-mitigation on the above considerations let us note that since error-mitigation is operation performed classically, it can be incorported into the spectrum of the Hamiltonian by defining spectrum $\left\langle \mathsf{T}_\alpha \mathsf{A}^\dagger_\alpha \right|$ of "error-mitigated Hamiltonian" term H^{corr}_α which energy is estimated on the noisy probability distribution $\left|\mathbf{p}^{\text{noisy}}\right\rangle$. Here we define a dual of correction matrix $\mathsf{A}^\dagger_\alpha$ acting on the spectrum of the local Hamiltonian. Note that if the correction matrix is exact (i.e., not approximate), it immediately follows that expected value of such defined Hamiltonian coincides with the true value of the energy

$$\left\langle H^{\text{corr}}_\alpha \right\rangle = \langle \mathsf{T}_\alpha \mathsf{A}^\dagger_\alpha | \mathbf{p}^{\text{noisy}} \rangle = \langle \mathsf{T}_\alpha \mathsf{A}^\dagger_\alpha | \mathsf{T}_\alpha \mathbf{p}^{\text{ideal}} \rangle = \langle \mathsf{T}_\alpha | \underbrace{\mathsf{A}_\alpha \mathsf{T}_\alpha}_{\mathbb{I}} \mathbf{p}^{\text{ideal}} \rangle = \langle \mathsf{T}_\alpha | \mathbf{p}^{\text{ideal}} \rangle = \langle H_\alpha \rangle\ . \tag{B.80}$$

Using this perspective we can derive the bounds on covariances between error-mitigated local Hamiltonian terms in a manner fully analogous to previous derivations. The "correlated part" this time is bounded as

$$|\langle \mathsf{T}_{\alpha\beta} \mathsf{A}^\dagger_{\alpha\beta} | \mathsf{T}_{\alpha\beta}\mathbf{p}^{\text{ideal}}_{\mathbf{\Delta}} \rangle| \leq ||H_\alpha H_\beta||\ ||\mathsf{A}^\dagger_{\alpha\beta}||_{1\to 1} ||\mathsf{T}_{\alpha\beta}||_{1\to 1}\ ||\Delta_{\alpha\beta}||_1 \leq ||H_\alpha H_\beta||\ ||\mathsf{A}^\dagger_{\alpha\beta}||_{1\to 1}\ \epsilon_{\alpha\beta}\ . \tag{B.81}$$

parameter	α	γ	A	a	c
starting value	0.602	0.101	200	0.06	0.12

Table B.1: *Starting values of hyperparameters used in optimization. The meaning of the parameters is in agreement with standard conventions (see for example Refs. [177; 178]).*

The rest of discussion is identical to the previous analysis of the effects of uncorrelated noise, the difference being that now we have additional terms coming from the duals of correction matrices. Note that those duals in special cases can have huge values – indeed, if the noise matrix is barely invertible in the first place, one can expect that error-mitigation will highly increase the uncertainty in the estimation of energy.

B.5 Details of numerical simulations and additional experimental data

In this short section, we provide some additional information on numerical simulations and experiments. This includes detailed discussion of methods used to simulate QAOA in Appendix B.5.1, and additional experimental results on characterization of correlations in Appendix B.5.2.

B.5.1 Simulation of the Quantum Approximate Optimization Algorithm

Here we provide some details of the performed numerical simulations. Simulation of QAOA was performed on the system of $N = 8$ qubits. In the main text, we described how the algorithm works, however, let us now repeat it for the sake of completeness. In standard implementation [144], one initializes quantum system to be in $|+\rangle^{\otimes N}$ state, where $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. Then p-layer QAOA is performed via implementation of unitaries of the form

$$U_p = \Pi_j^p \exp\left(-i\,\alpha_j \mathcal{H}_D\right) \exp\left(-i\,\beta_j \mathcal{H}_O\right), \tag{B.82}$$

where $\mathcal{H}_D$ is driver Hamiltonian, which we take to be

$$\mathcal{H}_D = \sum_k^N \sigma_x^k, \tag{B.83}$$

and $\mathcal{H}_O$ is objective Hamiltonian that one wishes to optimize (i.e., to find approximation for its ground state energy), and $\{\alpha_j\}, \{\beta_j\}$ are the angles to-be-optimized. The quantum state after p-th layer is

$$\left|\psi_p\right\rangle = U_p\,|+\rangle^{\otimes N}\,, \tag{B.84}$$

and the function which is passed to classical optimizer is the estimator of the expected value $\left\langle\psi_p\right|\mathcal{H}_O\left|\psi_p\right\rangle$ (note that this makes those estimators to effectively

be a function of parameters $\{\alpha_j\}, \{\beta_j\}$). The estimator is obtained by sampling from the distribution defined by the measurement of $|\psi_p\rangle$ in the computational basis, taking the relevant marginals, and calculating the expected value of $\mathcal{H}_O$ using values of those estimated marginals.

Theoretically one could optimize p-layer QAOA by simultaneous optimization of all $2p$ angles. However, this is hard in practice, since the number of optimized parameters increases the complexity of classical optimization. In our optimizations, we therefore modified the optimization to be divided into steps in the following way. In each step of optimization, we optimized over the set of 6 angles (i.e., 3 QAOA layers). Then the input to the next step was the optimized state obtained in the previous steps. For example, input to second optimization step was the quantum state $U_3^{(1)} |+\rangle^{\otimes N} = U_3 |+\rangle^{\otimes N}$, where we used superscript to denote optimization step. Then input to third optimization step was the state $U_3^{(2)} U_3^{(1)} |+\rangle^{\otimes N} = U_6 |+\rangle^{\otimes N}$, etc.

In summary, for p-layer QAOA, the optimization was effectively divided into $\frac{p}{3}$ steps, and in each step, we optimized over 6 angles, i.e., 3 QAOA layers. Each step used $2 * 800 = 1600$ function evaluations (the factor 2 comes from two evaluations needed for gradient evaluation which was done 800 times) plus a single final function evaluation. We further performed each such procedure additional time and chose the better run (out of 2). Each energy estimator was obtained using 10^4 energy measurements. Therefore the total number of function evaluations was $\approx \frac{p}{3} \times 3.2 \times 10^7$.

As a classical optimizer, we used Simultaneous perturbation stochastic approximation (SPSA) [177; 178; 156; 61]. Then with a growing number of optimization steps, we gradually changed the parameters to (heuristically) make the optimization more adaptive. Parameters α and γ were not changed, while other parameters were changed according to prescription

$$
\begin{aligned}
a_p &= a_0 \, 0.9^p \,, \\
c_p &= c_0 \, 0.9^p \,, \\
A_p &= A_0 \, 1.1^p \,,
\end{aligned}
$$

where 0 subscript denotes starting values of parameters given in Table B.1.

We note that the results presented in Fig. 4.6 exhibit rather poor convergence of the algorithm in a sense that adding more layers above $p = 9$ changed the resulting energies only slightly. This can be explained by the fact that we did not perform an exhaustive search over hyperparameters, but can also be a manifestation of the recently reported fact that QAOA might have problems with reaching global minima for relatively complicated Hamiltonians (like high-density MAX-2-SAT used in our work) [199]. Clearly, in the context of our work, only the comparison of noisy and noise-mitigated optimization to the noiseless run was of significance.

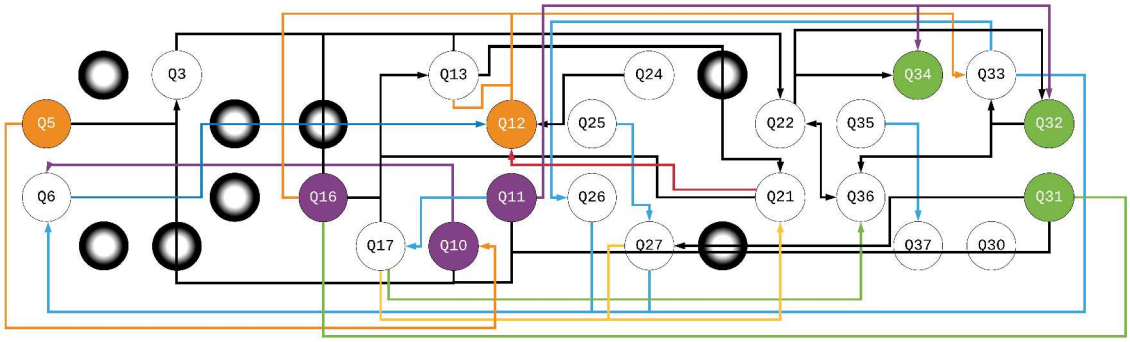


Figure B.1: *Depiction of correlations in Rigetti device. In Fig. 4.2 we presented the above image splited into two parts for clarity.*

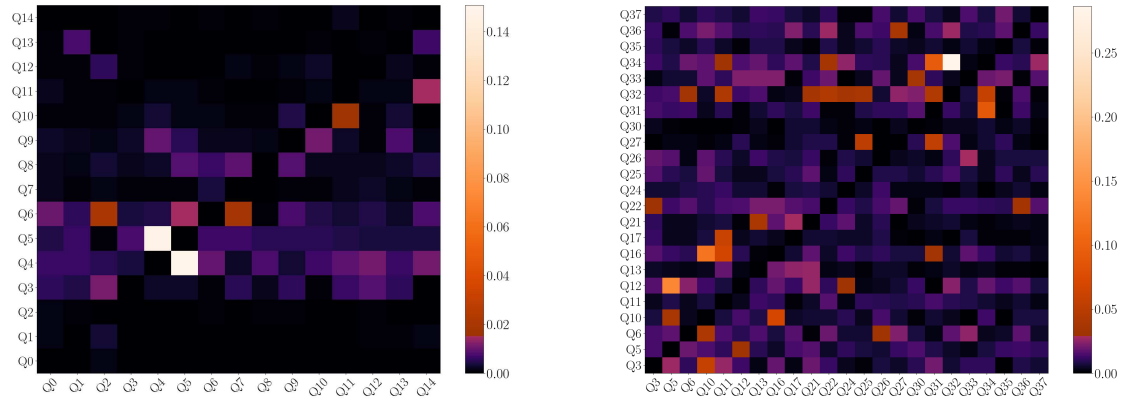


Figure B.2: *Heatmap of the correlations (Eq. (4.6)) in IBM's 15-qubit Melbourne device (left) and a 23-qubit subset of Rigetti's Aspen-8 device (right). The convention is that "row is affected by column", i.e., the measurement noise on the qubit with the label given by row index depends on the state of the qubit with the label given by column index, and the magnitude of the dependence is indicated by colors.*

B.5.2 Additional experimental data

Here we present additional data concerning correlations reconstructed in DDOT characterization of IBM's 15-qubit *Melbourne* device and a 23-qubit subset of Rigetti's *Aspen-8* device. The full depiction of correlations in Rigetti's device is presented in Fig. B.1. The heatmaps showing the reported correlations in both devices are presented in Fig. B.2. In Fig. B.3 we show how many parameters are needed to describe various noise models, and table in Fig. B.4 shows how much time data-processing took. The data was processed using laptop with 32GB DDR4 RAM (speed 2667MT/s) and Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz. We note that no multi-threading was implemented – in principle, one likely can reduce the run-time further be exploiting parallel calculations. We intend to optimize the code used for data-processing during development of our online repository [197].

	IBM's Melbourne - 15 qubits					
noise model	uncorrelated	only neighbors	**clusters & neighbors**	CTMP [15]	full classical	full quantum
parameters	30	72	76	450	$\approx 1.0 * 10^9$	$\approx 3.5 * 10^{13}$

	Rigetti's Aspen-8 - 25 qubits					
noise model	uncorrelated	only neighbors	**clusters & neighbors**	CTMP [15]	full classical	full quantum
parameters	50	342	756	1058	$\approx 7.0 * 10^{13}$	$\approx 5.9 * 10^{20}$

Figure B.3: *Total number of parameters needed to describe a noise model. The uncorrelated noise model is tensor product of single-qubit stochastic noise matrices. The "only neighbors" model corresponds to considering only single-qubit (trivial) clusters and their neighborhoods estimated in our experiments. The "clusters & neighbors" model corresponds to our full noise model containing both non-trivial clusters and their neighborhoods. CTMP is a number of parameters needed to describe 2-local classical noise model from Ref. [120] without any assumptions on the correlations structure. We note that combining our DDOT characterization with CTMP model could reduce the number of parameters in CTMP by pointing to negligible correlations that can be disregarded. For comparison, the "full classical" noise model refers to generic stochastic map, and "full quantum" to a generic d-outcome POVM.*

	DDOT circuits	**Shots per DDOT circuit**	**Pre-processing time**
IBM - 15q	749	8192	≈ 2.5 min
Rigetti - 23q	504	1000	≈ 4.5 min

	Error-mitigated experiments	**Shots per experiment**	**Total post-processing time**	**Error-mitigation time**
IBM - 15q	1200	40960	≈ 5.5 min	≈ 9 s
Rigetti - 23q	399	1000	≈ 2 min	≈ 6 s

Figure B.4: *Time of data processing. In first table, the "pre-processing time" includes calculation of marginal noise matrices on 2-qubit subsets, calculation of pairwise correlations, reconstruction of the noise model and calculation of inverse noise matrices needed for corrections of all possible two-qubit marginal probability distributions (note that this in general includes also higher-dimensional corrections, as explained in Fig. 4.3). In the second table, the "Total post-processing time" includes both calculation of marginal distributions needed to estimate energies of investigated 2-local Hamiltonians and performed error-mitigation on them. The "Error-mitigation time" shows only the latter.*