

Master's thesis

Master's Programme in Theoretical and Computational Methods

Practical Techniques for Accurate Measurements on Noisy Quantum Systems

Hetta Vappula

25th June 2025

Supervisors: Prof. Zoltán Zimborás
Dr. Daniel Cavalcanti
MSc Keijo Korhonen

Examiners: Prof. Zoltán Zimborás
Dr. Marco Cattaneo

University of Helsinki
Faculty of Science

P. O. Box 64 (Gustaf Hällströmin katu 2)
00014 University of Helsinki

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta — Fakultet — Faculty	Koulutusohjelma — Utbildningsprogram — Degree programme
Faculty of Science	Master's Programme in Theoretical and Computational Methods

Tekijä — Författare — Author

Hetta Vappula

Työn nimi — Arbetets titel — Title

Practical Techniques for Accurate Measurements on Noisy Quantum Systems

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidantal — Number of pages
Master's thesis	25th June 2025	54

Tiivistelmä — Referat — Abstract

High-precision measurements on near-term quantum devices are important for advancing quantum computing applications. The high readout errors of the current quantum computers complicate any effort to perform quantum simulations with high accuracy. This thesis implements practical techniques to reduce overhead and error, achieving accuracies suitable for quantum chemistry applications. To reduce circuit overhead, we use repeated settings and parallel Quantum Detector Tomography (QDT), and for shot overhead, we use a locally biased randomized measurement scheme. QDT is also used to mitigate readout errors and blending is used to mitigate the effects of drift. These techniques were validated by estimating the expectation values of the energy of the BODIPY molecule on a Hartree-Fock state. All experiments were performed on an IBM Eagle r3 quantum computer, where we were able to reduce the measurement errors by an order of magnitude, from 1-5% to 0.16%. These methods enable reliable quantum computations, essential for precise molecular energy calculations.

Avainsanat — Nyckelord — Keywords

quantum computing, NISQ, POVM, QDT, drift

Säilytyspaikka — Förvaringsställe — Where deposited

Muita tietoja — Övriga uppgifter — Additional information

Education is the most powerful weapon which you can use to change the world.
— Nelson Mandela

Acknowledgments

I would like to thank my supervisors and the entire team at Algorithmiq for everything they have taught me and for all the cool things I have experienced in the past two years. Ever since I stumbled into the quantum computation course by Prof. Sabrina Maniscalco and Matteo Rossi, I have been astounded by the topic and look forward to learning more.

I would also like to thank the Cleveland Clinic Quantum Innovation Catalyzer Program for giving us access to the IBM Cleveland Clinic System One quantum computer.

Thank you to all the amazing people who have taught me over the years. Thank you for believing in me and especially, thank you for inspiring me.

I would also like to thank all of my friends for supporting me throughout my studies and making sure I touch grass every once in a while. Thank you for being exactly who you are.

Special thanks to my family, and especially to my partner Aino, for listening to my ramblings on various topics related to theoretical physics and beyond. Thank you for sticking along for the journey for all these years.

Contents

1 Introduction **1**

2 Theoretical Background **5**

2.1 Quantum measurement 5

2.1.1 Estimating expectation values of observables 5

2.1.2 General measurement 6

2.1.3 Duals 9

2.2 Noise 11

2.2.1 State preparation and measurement (SPAM) errors 11

2.2.2 Classifications of noise 12

2.2.3 Superconducting qubits 14

2.3 Why is noise a problem? 15

2.4 Quantum Detector Tomography 16

3 Methods **19**

3.1 Ways of dealing with the noise 19

3.1.1 Dealing with noisy detectors 19

3.1.2 Dealing with temporal noise 20

3.1.3 Details of the blending procedure 22

3.1.4 Other methods 24

3.2 Hardware implementation 24

3.2.1 Implementation of the blending procedure on IBM 25

3.2.2 Repeated measurement settings for reducing circuit overhead . . 25

3.3 Case study: energy estimation of the BODIPY-4 molecule 29

4 Experimental Results **31**

4.1 Reduction of estimation bias using QDT 31

4.2 Measurement overhead of the measurement schemes 33

4.3 Characterization of temporal noise with blending 35

5 Conclusions **39**

Bibliography **40**

Appendix A Error suppression techniques **49**

A.1 Dynamical decoupling 49

A.2 Pauli twirling 49

Appendix B Additional results from the experiments **51**

B.1 Comparison of measurement schemes with and without QDT 51

B.2 Results for excited state Hamiltonians 52

B.3 Noise model for a fluctuating detector 54

1. Introduction

The study of physics has helped us to understand and predict the nature around us. This knowledge has allowed us to build, fly, feed, and heal. While progress is rarely linear, technological improvements often give us hope for a better future. As the theories of quantum mechanics are now applied to build the first quantum computers, we might be on the brink of a new technological leap.

Quantum computing has the potential to significantly impact fields such as climate and environmental sciences as well as the field of life science. By solving problems intractable for classical computers, they could enable us to tackle more and more complex problems. We are still in the very early days of quantum computing, and there are still many problems to solve before their full potential can be reached.

One of these most significant problems is the level of noise in current quantum computers, which affects both the precision and the accuracy of the results. Noisy gates, readout errors, and other types of noise cause the collected statistics to be unreliable. Adding further to the problem, measuring the inherently probabilistic quantum state is a source of statistical noise, sometimes referred to as shot noise. The limited scale and performance of current quantum computers, combined with the resource intensive nature of any real-world application, complicates any attempts of reaching useful quantum advantage. While these problems need to be also addressed on the hardware level, we investigate ways of enhancing the precision, accuracy, and reliability in the presence of noise using algorithms and post-processing techniques. Before reaching full fault-tolerance, these techniques aim to squeeze every last drop of performance out of a quantum computer.

In this thesis, we investigate how noise affects experimental results on a modern quantum computer and propose ways of mitigating the effects of noise as well as suppressing noise where possible. Our focus is mainly on measurement noise, but some of the strategies can also be useful when combating noise from other sources. We validate the proposed methods by using them to estimate the energy of a given physical system. For this purpose, we chose the Boron-dipyrromethene (BODIPY) molecule, an organic fluorescent dye used widely in various fields of chemistry.

In previous experiments, quantum computers have been used to estimate the

ground-state energies of certain molecules [1, 2, 3, 4, 5, 6, 7]. In order to accomplish this, one needs to estimate the expectation value of the state and the chemical Hamiltonian to a high precision. The upper bound of said precision has commonly been set to *chemical precision* at 1.6×10^{-3} Hartree [8]. We use the term *chemical precision* instead of the more common *chemical accuracy* in order to distinguish between the precision of estimating the energy of an ansatz state and the absolute error of the energy of the ansatz to the target energy of a molecule. The rationale for selecting this specific value comes from the sensitivity of reaction rates to variations in energy. Even in the fault-tolerant era, the efficient estimation of expectation values to high precision will be a relevant challenge due to the complexity of the observables representing the chemical molecules.

The focus of this work is on decreasing shot overhead (*i.e.* the number of times the state prepared on the quantum computer is measured), the circuit overhead (*i.e.* the number of unique circuits measured on the quantum computer), as well as mitigating the effects of measurement noise at different time scales. Each of the mentioned issues poses significant obstacles to the precision, accuracy, and efficiency of quantum computations. The techniques presented in this work are targeted solutions to one or more of the issues, and by combining them together, we are able to produce an efficient pipeline for high-precision measurements. We show that the combination allows us to reach estimation errors close to chemical precision when estimating complex observables, even though the measurement errors on current quantum computers are on the order of 10^{-2}.

A key component of our high-precision measurements toolbox is the use of informationally complete (IC) measurements. While the estimation of chemical observables is also possible via non-IC measurements such as Pauli grouping [2, 9, 10, 11, 12, 13], where the observable can be measured in groups of Pauli strings and not one Pauli string at a time, the IC measurements offer many benefits over the other methods. These include advantages such as allowing the estimation of multiple observables from the same measurement data [14, 15, 16, 17, 18], proven to be useful for measurement-intensive algorithms such as ADAPT-VQE [19], qEOM [20] and SC-NEVPT2 [21]. IC measurements also enable the use of efficient error mitigation methods [22, 23] and, more importantly, can be used with quantum detector tomography (QDT) [24, 25]. As a result, we can construct an unbiased estimator for the expectation value of the energy even with noisy data. They also allow us to use efficient post-processing methods to reach lower variances of the estimate and decrease the shot overhead [26, 27, 28, 29].

An additional method for tackling the shot overhead is the use of locally biased random measurements [30], introduced in Sec. 2.1.2. They allow us to allocate more of the shots to estimating the parts of the observable with a higher impact to the expec-

tation value of the energy. This allows us to reduce the total number of shots required, while still maintaining the informational completeness of the data. In Secs. 3.1.1 and 3.2.2 we introduce methods to reduce circuit overhead by using parallel QDT and repeated settings, optimizing the use of quantum resources. Finally, to mitigate the effects of drift observed on the device, in Sec. 3.1.2, we introduce the method of blending that takes any drift within a set experiments and makes it quasi-static.

The strategies presented in this thesis are a collection of practical methods for performing high-precision measurements. We validate the methods on a quantum computer, but most of the techniques can be useful in other contexts, such as in the study of quantum thermometry and quantum metrology [31, 32] as well. They significantly increase the reliability, accuracy, and precision of quantum computation, allowing for more impactful use cases for quantum technology.

In this thesis, we will explain each method used in detail and provide some basic background to the topics discussed as needed. The implementation of each method for IBM quantum computers is discussed as it seems necessary. The effectiveness of the techniques in improving the reliability, accuracy, and precision is validated by running experiments on an Eagle r3 quantum computer. This thesis is based on the research detailed in Ref. [33], where the author of the thesis is one of the contributors. The paper has been accepted for publication in *npj Quantum Information*.

2. Theoretical Background

2.1 Quantum measurement

To gain insight into a quantum state, it needs to be measured. Unlike its classical counter part, quantum measurement is inherently probabilistic. This is in the core of any quantum system, where the wave function of the system collapses into one of many possible outcomes at the point of measurement. This introduces a probabilistic aspect to the measurement that in the classical case is purely deterministic. In this section, we discuss both where quantum measurements are useful and how we describe them mathematically.

2.1.1 Estimating expectation values of observables

We are currently in the Noisy Intermediate-Scale Quantum (NISQ) era [34, 35], where all available quantum computers are inherently noisy and typically range from just a couple qubits to a few hundred qubits. While prominent quantum algorithms such as Shor's factoring require fault-tolerant quantum hardware [36], the current NISQ era is an important one for developing all of the tools required later. Furthermore, the noisy quantum computers could prove to be useful in some real-world applications even in the near future.

An interesting application of NISQ hardware is the estimation of expectation values of observables such as the ground state energy of a given molecule. The observable, possibly a chemical Hamiltonian describing some molecule, is transformed into qubit space using a fermion-to-qubit mapping. This transforms an observable described with creation and annihilation operators to be described using Pauli operators. The end result is a sum of Pauli strings with corresponding coefficients, where the individual Paulis (Z, X, Y and I) in the string describe the corresponding qubits.

There are many ways to estimate the expectation value of the described observable, but in the simplest case each Pauli string is estimated separately to a desired precision. The so-called Pauli weight of these strings is seldom very large, *i.e.* there are many identity terms in the Pauli strings. We can exploit the resulting commutativity

of the Pauli strings to reduce the measurement overhead by computing the expectation values of multiple terms at the same time. These so-called Pauli grouping methods have been studied extensively, see Refs. [2, 9, 10, 11, 12, 13].

The number of Pauli strings in chemical Hamiltonian increases polynomially as the system size increases, and thus the quantum resources required with naive methods prohibit us from estimating observables of big systems. This polynomial growth of the number of Pauli strings arises from the fact that the Hamiltonian may include interaction terms between distant particles in the molecule. For a chemical Hamiltonian the number of Pauli strings scales as n^4, with n being the number of qubits [12]. While a Pauli grouping method reduces the number of measurements needed, finding the optimal groupings is an NP-hard problem [37]. There are machine learning methods built to address the measurement problem [38] but in this thesis we focus on the classical shadows formalism first introduced in Ref. [16].

To get past the hurdles introduced by the scaling of the Hamiltonian, we use a randomized measurement scheme. With it, we construct a minimal classical sketch of the measured state, the so-called *classical shadow* [16]. This shadow is then used to estimate properties such as the ground energy of a molecule. This thesis builds on previous work on shadow estimation both by implementing a version of it in practice and addressing some of the practical problems with it. Understanding these topics is important in overcoming the problems posed by the polynomial scaling of the chemical Hamiltonians. Throughout this thesis, we will discuss a range of methods important in realizing accurate and precise measurements, especially important in the NISQ era but also in the early fault-tolerant one.

2.1.2 General measurement

To model a measurement on quantum hardware, we utilize a form of generalized measurements called the positive operator-valued measure (POVM). Using the POVM formalism, we reconstruct the shadow of a quantum state and use that to estimate different expectation values.

Suppose, we describe the measurement with the operator M_k acting on the quantum state $|\psi\rangle$. The probability of the outcome k is then

$$\mathbb{P}(k) = \langle\psi|M_k^\dagger M_k|\psi\rangle. \tag{2.1}$$

We now define $\Pi_k \equiv M_k^\dagger M_k$, where Π_k is a positive semi-definite Hermitian matrix with $\sum_k \Pi_k = \mathbb{I}$ and $\mathbb{I}$ is the identity matrix. The defined operators Π_k can be used to characterize the quantum state we want to measure. The set of *measurement effects* $\{\Pi_k\}$ define the POVM [39]. Each effect is associated with a possible experimental

result (or outcome) k so that the probability of obtaining the k^{th} result when measuring the state ρ is given by the Born rule $p_k = \text{Tr}[\rho\Pi_k]$ [39]. By the definition of POVMs, it is guaranteed that $p_k \geq 0$ and $\sum_k p_k = 1$.

In this work, we focus on informationally complete (IC) POVMs, for which the measurement effects form a basis in the Hilbert space of the system, denoted by $\mathcal{B}(\mathcal{H})$. Consequently, any operator $O \in \mathcal{B}(\mathcal{H})$ can be expressed as $O = \sum_i \omega_i \Pi_i$, where Π_i are POVM elements and ω_i are real coefficients. Thus, the expectation value of O can be written as

$$\langle O \rangle = \text{Tr}[\rho O] = \sum_i \omega_i \,\text{Tr}[\rho\Pi_i] = \sum_i \omega_i p_i, \tag{2.2}$$

for any quantum state ρ. This means that the expectation value of the operator O is the average of ω_i over the probabilities p_i. To implement this in action, we use a Monte Carlo estimator, where we sample S times from the underlying probability distribution with local POVMs, resulting in the following estimator:

$$\overline{O} = \frac{1}{S}\sum_{s=1}^{S} \omega_{i_s}. \tag{2.3}$$

In ref. [40] the estimator is explained in more detail, but it is important to mention that the estimator provides accurate estimates even when the number of total shots is significantly lower than all of the possible outcomes of the POVM *i.e.* when $S \ll 6^{N*}$. The choice of the IC-POVMs can be optimized based on the observable in question to minimize the variance of $\langle O \rangle$ [40, 25, 41]. This would allow the experimenter to use a lower number of shots to obtain similar results when compared to an unoptimized IC-POVM.

A POVM is informationally complete when it has at least d^2 effects where d is the dimension of the Hilbert space of the operators. In our case, this means that there needs to be at least 4 effects to satisfy the informational completeness. An upside of using IC data is the available post-processing methods that benefit from it, such as the ones described in Refs. [22, 23] and the method of QDT, described later in Sec. 2.4.

A POVM can be defined with any set of operators that satisfies the rules above[†]. POVMs can be implemented in numerous ways, but the architecture of the quantum hardware used in this work and its limitations, such as low number of qubits and its connectivity, push us toward POVM implementations, where auxiliary qubits are not needed. In these PM-simulable cases, the measurement scheme consists of multiple projective measurements (PM), which in this case are Pauli measurements. We measure the qubit in different measurement bases and then combine these results in post-processing. This does not introduce any qubit overhead and in theory it is a

*True when the POVM used has 6 local effects, *e.g.* the Pauli measurements used in this thesis

[†]1. Positive semi-definite Hermitian matrix 2. Summing to identity

very simple operation. In practice, however, running these circuits on real hardware introduces other overheads that are addressed later.

The measurement schemes employed in this work belong to the space of PM-simulable IC-POVMs and are commonly referred to as classical shadows [16] and locally biased classical shadows [30]. In the classical shadows scheme, the measurement basis is randomly chosen from three non-trivial Pauli operators $\{Z, X, Y\}$, with each operator having an equal probability of selection. For each Pauli operator *i.e.* measurement basis, there exist two possible outcomes, corresponding to the eigenstates of the operator. Thus, for classical shadows, the POVM effect on qubit q is defined as:

$$\Pi_{k,i}^{(q)} = \frac{1}{3} E_{k,i}^{(q)}, \tag{2.4}$$

where $E_{k,i}^{(q)}$ is one of the two eigenstates of the Pauli operator $P_k \in \{Z, X, Y\}$ that is measured on qubit q, determined by the measurement outcome $i \in \{0, 1\}$. An N-qubit classical shadows POVM is a tensor product of N single qubit POVMs:

$$\Pi_{k,i} = \frac{1}{3^N} \bigotimes_{q=1}^{N} E_{k,i}^{(q)}. \tag{2.5}$$

When $N = 1$ we can write down all six of the effects $\Pi_{k,i}$ in the following manner:

$$\begin{aligned} \Pi_{0,0} = \frac{1}{3}|0\rangle\langle 0|, \quad \Pi_{0,1} = \frac{1}{3}|1\rangle\langle 1|, \quad \Pi_{1,0} = \frac{1}{3}|+\rangle\langle +|, \\ \Pi_{1,1} = \frac{1}{3}|-\rangle\langle -|, \quad \Pi_{2,0} = \frac{1}{3}|+i\rangle\langle +i|, \quad \Pi_{2,1} = \frac{1}{3}|-i\rangle\langle -i|, \end{aligned} \tag{2.6}$$

where $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$ and $|\pm i\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm i|1\rangle)$*. For a single qubit, we have $d = 2$ and since $d^2 = 4 > 6$, this POVM is informationally over-complete. The over-completeness allows us to utilize a post-processing method of dual optimization not available to the so called minimally complete POVMs that only have 4 effects. The method has been explained in Sec. 2.1.3

Locally biased classical shadows extend the classical shadows framework with non-uniform sampling probabilities [30]. The Pauli operators $\{Z, X, Y\}$ are sampled locally using a local probability $\tau_k^{(q)}$. Thus, each measurement outcome $E_{k,i}^{(q)}$, corresponding to a Pauli operator $P_k^{(q)}$ has the same $\tau_k^{(q)}$. As a result, the measurement effects can be represented as the following:

$$\Pi_{k,i} = \bigotimes_{q=1}^{N} \tau_k^{(q)} E_{k,i}^{(q)}. \tag{2.7}$$

The local probability $\tau_k^{(q)}$ is obtained by computing the total weights of the qubit-marginalized Pauli operators in the input observable $O = \sum_j c_j P_j$. This indicates each

*$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $|1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

Pauli operator's local weight in the observable. The sampling probability for the Pauli operator $P_k^{(q)}$ on qubit q is:

$$\tau_k^{(q)} = \frac{\sum_j |c_j| \mathbf{1}_{P_k^{(q)} = P_j^{(q)}}}{\sum_j |c_j| \mathbf{1}_{P_j^{(q)} \neq I}}, \tag{2.8}$$

where $\mathbf{1}_{P_k^{(q)} = P_i^{(q)}}$ denotes the indicator function, giving 1 if the Pauli operator $P_k^{(q)}$ equals the Pauli operator at qubit q of the jth Pauli string in O and 0 if not. The denominator is the absolute sum of all the coefficients corresponding to the Pauli strings in the observable where the identity operator is not found (or equivalently, any one of $\{Z, X, Y\}$ is found) at qubit q.

To ensure informational completeness, we introduce a heuristically determined minimum threshold value $\tau_{\min} = 0.2$ for the values $\tau_k^{(q)}$. This limit on the probability guarantees that each Pauli operator is sampled locally with a non-zero probability. The locally biased classical shadow scheme has the advantage that it is observable aware. Determining the sampling probabilities based on the observable allows us to focus efforts more effectively compared to the equal probability of the classical shadows method. Measuring local Pauli operators with a large contribution to the estimated expectation value more often, also lowers the contribution these terms have to the variance of the estimation. Thus the locally biased measurement scheme lowers the total variance, while preserving the informational completeness.

The local Pauli operators can be measured by applying different single-qubit gates before the canonical measurement in the Z-basis, which is natively implemented on IBM hardware. To measure the Z operator, one can directly perform the native measurement. To measure the X operator, one applies a Hadamard gate before the canonical measurement. Lastly, to measure the Y operator, one applies an $S^\dagger$ gate followed by a Hadamard gate and the canonical measurement.

2.1.3 Duals

When using POVMs to measure a state ρ, the results need to be interpreted. For any informationally complete POVM there exist a set of dual operators with which we can estimate the measured state. These *dual effects* D_i are defined by operators satisfying

$$O = \sum_i \mathrm{Tr}[O\Pi_i] D_i, \tag{2.9}$$

for any operator O [26].

In the case of minimally IC-POVMs, there exists only a single valid set of duals, sometimes referred to as the *canonical duals*. This is not the case with OC-POVMs where there exists infinitely many valid sets of duals in addition to the canonical ones.

To estimate the state ρ we construct an unbiased estimator by setting O in Eq. (2.9) to be ρ and get the following:

$$\bar{\rho} = \sum_i \mathrm{Tr}[\rho \Pi_i] D_i \tag{2.10}$$

$$= \sum_i f_i D_i \tag{2.11}$$

where f_i is the observed frequency of outcome i and D_i is the dual operator corresponding to the effect i [26].

The infinitely many valid dual frames allow us to optimize the duals used and with that reach lower variances, while using the same dataset. Optimal dual frames have been mentioned in Ref. [42] but the after the publication of Ref. [43] on *shadow estimation* there have been works proposing to optimize the dual operators using measurement data. In general these dual optimization methods are free in terms of the quantum resources required, and thus they are an important tool for increasing the precision of the estimations.

2.2 Noise

In this section, we discuss some of the characteristics of noise in current quantum hardware. We look at the time scales at which noise affects the experimental data gathered and finally explain the theoretical background for QDT.

2.2.1 State preparation and measurement (SPAM) errors

The errors related to the state preparation and measurement are often bundled together as it is hard to distinguish between the two. In general, the state preparation of a qubit in modern superconducting hardware is fairly good and most of the focus has thus been directed toward the measurement part. Sources of the readout error vary across different platforms but here we will focus on superconducting qubits. Typical sources of the readout error contain:

- As the physical measurement takes a significant amount of time, there is a non-zero probability that the measured state decays from the excited state $|1\rangle$ to the ground state $|0\rangle$, resulting in an error. Longer T_1 times help with this error but as the relaxation time might not be constant across the time window of the experiment [44], this type of error can also vary in time. This rises from a bath of parasitic two-level quantum systems, called TLS.
- The probability distributions of the measured physical quantities may have some overlap, resulting in assignment errors [45]. Furthermore, the distributions related to the computational states may not be symmetric, which makes it more probable to assign $|1\rangle$ as $|0\rangle$ than vice versa.
- There are many other sources that could effect the fidelity of the readout such as correlations between the qubits (cross-talk), thermal noise, problems in the measurement apparatus and many others.

The above-mentioned TLS noise and its sources have been a topic of extensive research [45, 46, 47, 48, 49, 44]. The parasitic TLS defect may absorb energy from the qubit if its resonance frequency matches the frequency of the qubit. While the qubits can be tuned to use frequencies where this is not a problem, the TLS frequency may vary in time. This variation may affect both the resonance frequency and the relaxation time of the qubit for some period of time. When the TLS frequency wanders away of the range of the qubit, the T_1 time returns to the levels it was previously [44] and the readout performance returns to its baseline level. The qubit-TLS interactions may introduce temporal fluctuations and could be the reason for the behavior shown in Fig.

4.4. This interaction may also be a source of memory effects [50], inherently linked to non-Markovian dynamics. These memory effects could arise from a scenario in which a qubit-TLS interaction at one time makes the energy absorption less likely at a later time. In another interesting case, the qubit absorbs energy from the parasitic TLS defect, introducing a source of information backflow. It is unclear to what extent these memory effects play a role in our experimental results, but the topic is definitely an interesting one.

Although completely removing SPAM noise is not feasible, there are definitely ways to engineer better measurement hardware. Meanwhile, there are ways to mitigate the effects of the errors and that is something we will focus on in this thesis.

The sources of the error is indeed an interesting topic, but it is not as useful when developing methods to counteract noise on software level. The errors can be modeled as random bit-flips of the qubit and even the time-dependent readout errors can be dealt with. In the single-qubit case, when the initialization errors are assumed to be insignificant, we can describe the errors classically and write them as:

$$\vec{p}_{noisy} = A\vec{p}_{ideal}, \tag{2.12}$$

where $\vec{p}_{noisy}$ corresponds to the vector of noisy outcome probabilities and $\vec{p}_{ideal}$ is its noiseless counterpart [51]. Since we are only looking at a single qubit, the confusion matrix can be written as:

$$A = \begin{bmatrix} P(0|0) & P(0|1) \\ P(1|0) & P(1|1) \end{bmatrix}. \tag{2.13}$$

In the case of more than one qubit the matrix and its estimation get more complicated, unless one assumes the errors to be independent, in which case it suffices to compute the matrix for each qubit separately. The matrix describing the full system A_{full} would in that case be a tensor product of the single-qubit confusion matrices A_q.

$$A_{full} = \bigotimes_{q=1}^{N} A_q. \tag{2.14}$$

2.2.2 Classifications of noise

In Ref. [46] quantum noise is divided into three categories, *process error*, *control error*, and the combination of these two, the *superprocess error*. *Process error* is an always on interaction that happens without any interventions by the experimenter. In it, the system interacts with an inaccessible environment. During this interaction, the system evolves according to some Hamiltonian other than the ideal one. These process errors are stochastic noise that cannot be turned off and thus will always be present. An example of this is the TLS noise, that appears independent of the gates being applied.

Unlike the always-on nature of the process errors, the *control errors* are tied to the actions of the experimenter. It is gate-dependent noise for which all the present dynamics originate in some way from the control equipment. These errors arise from the fact that the controls are not perfect, but they are tied to the actions performed on the system. Most of these types of errors are from gates applied to the qubits.

Finally, we have the overlap of the two aforementioned errors, *superprocess error*. Its effects on the system can be turned on an off, and thus it belongs to the category of control error, but it can interact with the environment and thus indirectly modify the process errors.

Process errors and control errors can be further subdivided into three overlapping categories, non-Markovian, time-dependent Markovian, and time-independent Markovian. A model constructed to describe non-Markovian noise can also describe Markovian noise, but the reverse is not true. Experiments on real quantum computers take a non-trivial amount of time and thus it is beneficial to look at the effects of any time-dependent noise at different timescales:

- The shortest of these is the circuit-level window, where the noise is looked at on timescales at which the different gates are performed. These dynamics repeat across shots of the same circuit. The noise map remains intact across repetitions of the circuit, *i.e.* the noise is static and thus it can be characterized easily.

- Quasi-static noise can affect the system through a single shot, for which a single stochastic variable is drawn from some stationary distribution. The stochastic variable can be different for each of the shots, resulting in dynamics that even with Markovian noise, can look non-Markovian after the results have been marginalized over the time window. While noise is never preferred, this type of noise is easy to characterize, unlike the one described next.

- The longest time scale is drift, which takes into account any of the dynamics for which the time scale is much larger than the repetition rate of the quantum device. This stays constant across a single set of shots, but across a longer experiment with hours of runtime these dynamics show up. Thus, in longer experiments, drift can be a significant source of error. As with quasi-static noise, even if the source of the drift is Markovian, the results may look non-Markovian, when marginalized.

The definition of non-Markovianity is not as clear as one would expect. When we said that the results may look non-Markovian, even if the noise is only time-dependent Markovian, we meant the following. If we do not consider the wall clock when gathering the data, we cannot determine if the source of some noise is time-dependent Markovian

or non-Markovian. While some consider the former to be included in the latter, we consider them to be separate, as the two have distinct differences. Let us consider a scenario where the first 10 minutes of an experiment, the device performs well, but due to faulty electronics, during the next 10 minutes, the results are much worse. In another scenario, a similar experiment is performed, but this time, due to interactions with the environment, the results seem similar to those of the first example. In the latter case, the interaction is somewhat static, whereas in the former case, there is a clear change in the dynamics. Only by looking at the marginalized results, we cannot differentiate between the two scenarios, but if we take into account the non-trivial time it takes to gather the data, we can easily make the distinction. Using this logic, Markovian noise, if time-dependent, can look as if it was non-Markovian.

The main difference between the quasi-static noise and the drift is that the quasi-static noise affects the results similarly across multi-shot experiments and seems to in that sense stay static. Compared to drift, which can make it hard to compare experimental results over time. This becomes apparent when competing versions of a similar method are tested. If the runs are performed one after another, the drift can affect some results more than others or in different ways. In the case of noise characterization this drift makes old noise models useless and forces us to learn the noise as close to the running of the experimental circuits as possible. As discussed in Sec. 3.1.2 the ideal case is to interleave the circuits to minimize the effects of drift by changing the dynamics from the regime of drift to the regime of the quasi-static noise. This comes with the added benefit that as the experiment time window is averaged over, two experiments performed at different times may be more comparable, as they both get closer to an "average" experiment.

2.2.3 Superconducting qubits

Many of the current quantum hardware providers have opted to build quantum computers utilizing Josephson junction based superconducting qubits that are connected with couplers. The layout of the qubits changes between the providers, but the used Eagle r3 device developed by IBM has a layout called the heavy-hex, shown in Fig. 2.1. The sparsity of the layout helps to keep any unwanted inter-qubit interactions at bay, but as a result the connectivity of the qubits is not ideal. With deeper circuits, limited connectivity may force an experimenter to increase the number of two-qubit gates drastically as interactions between far away qubits require the use of SWAP gates. This is especially impactful, as the noise related to a two-qubit gate is around an order of magnitude larger compared to that of a single-qubit gate.

The initial states are produced by repeatedly measuring the qubits and applying

a bit-flip controlled on the result. The end goal is to have all the qubits in the $|0\rangle$ state when starting an experiment. Typically, the readout on a superconducting quantum computer is performed by coupling a qubit with a resonator and subsequently measuring the change of the state of the resonator using microwave pulses. During a circuit execution, the gates are applied using microwave pulses on the qubits. All of these processes are inherent sources of error, at least in the near future.

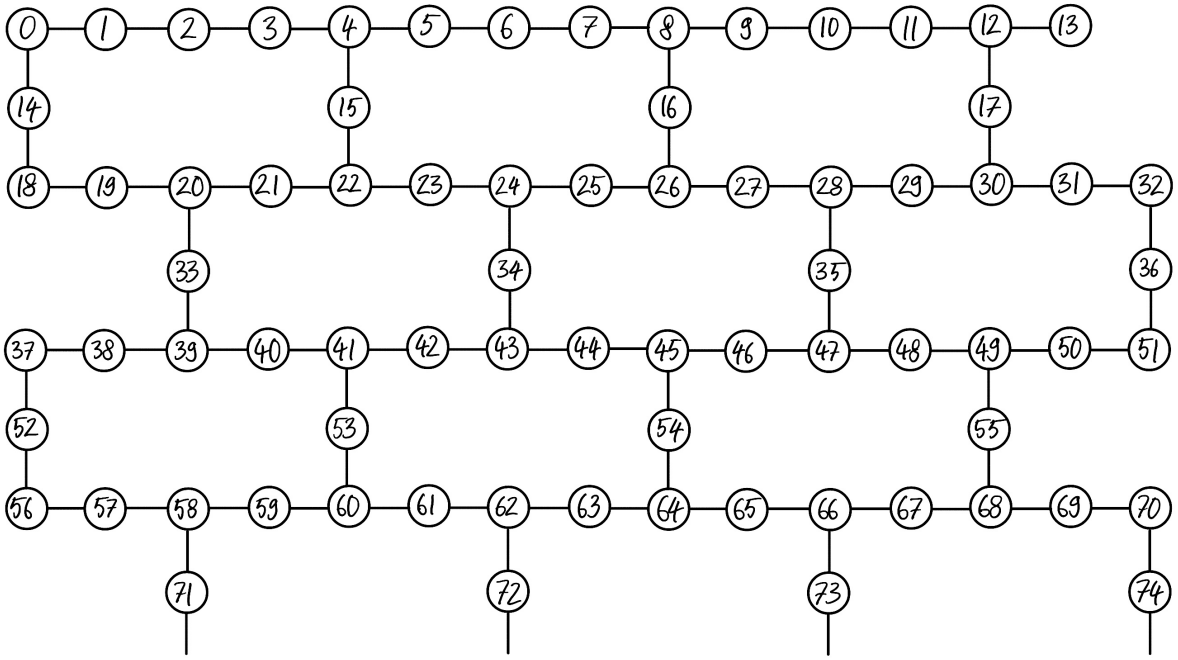


Figure 2.1: Partial layout of an IBM Eagle r3 chip, illustrating the heavy-hex structure.

2.3 Why is noise a problem?

This might seem like a silly question, but it is crucial for optimizing the methods used to mitigate noise. To maximize the performance of a quantum device, it is beneficial to employ a variety of techniques, each contributing uniquely. However, many of these methods have drawbacks, such as prolonging the total quantum runtime, adding complexity to the pipeline, and potentially introducing more components that could break the process.

In chemical applications, as described in Section 2.1.1, both accuracy and precision must be sufficiently high for experimental results to be useful. It is insufficient to ensure that the results indicate the correct energy if the error bars are excessively large. Similarly, it is not ideal for the variance to be very low if the energy approximation is biased. Some methods, such as Quantum Detector Tomography (QDT), primarily address accuracy, while others, such as the optimization of POVMs, help achieve more

precise approximations with the same number of shots.

When moving to experiments with higher qubit counts and deeper circuits, in addition to the measurement noise, we start to face other types of noise and limitations. The limited coherence time of a qubit combined with the gate errors in the best case introduces a bias to the estimated energy. At worst, they make the prepared state noisy at levels where the data obtained is no better than a coin flip. In order not to use quantum computers as very expensive random number generators, the circuits need to stay within reasonable limits for the given level of noise. As mentioned in Sec. 3.1.4, there are methods to mitigate some of this noise, but we have still not reached the point of practical quantum advantage [35]. Removing noise altogether would be fantastic, but not feasible in the near term, as the noise levels required for large-scale error correction are well below the current levels. In the meantime, there have been interesting proposals to reduce the noise in the experimental data, such as detecting errors by using a dual-rail qubit [52].

2.4 Quantum Detector Tomography

Measurements on current quantum hardware are noisy. To mitigate the effect of this noise, we can employ a plethora of different methods. For POVMs, Quantum Detector Tomography (QDT) is used to find the noisy effects of the POVM that model the measurement noise of the device. The tomography returns a POVM, which is modeling the one used to get the experimental data. To correctly characterize the detector, QDT has to get the data from the same detector used when estimating the energies. Thus, the performance of QDT relies on the noise being quasi-static, *i.e.* absent of drift. As long as this is satisfied, the effects of the readout noise on the energy estimation can be effectively mitigated. This improves the accuracy of the estimation, but comes with the inherent downside of also increasing the variance. This increase depends on the level of noise in the device.

To perform QDT, one needs a tomographically complete set of input states measured with the noisy detector. In other words, we need to have at least d^2 linearly independent input states. These states are assumed to be noiseless and while the assumption is incorrect, it is a good simplification as the SPAM errors are usually dominated by the measurement errors*. After obtaining the probabilities $p_{j,k} = \mathrm{Tr}[\rho_j \Pi_k]$ we can compute the noisy effect Π_k by solving the inverse problem [53, 54]. Semidefinite programming (SDP) is used to ensure that the obtained effects are physical, *i.e.* they are all positive. This ensures good performance, as there exist very efficient numerical

*This is the case with the Eagle r3 device we used.

methods for solving these types of problems. The optimization problem to be solved, is the following:

$$\min_{\Pi_k} \|\boldsymbol{f} - \boldsymbol{q}\| \tag{2.15}$$

$$s.t. \quad \Pi_k \geq 0 \quad \forall k \tag{2.16}$$

$$\sum_{k=1}^{m} \Pi_k = \mathbf{I}, \tag{2.17}$$

where $\boldsymbol{f}$ corresponds to the experimental frequencies and $\boldsymbol{q}$ to the probabilities given by $p_{jk} = \mathrm{Tr}[\rho_j \Pi_k]$. By modifying the POVM effects Π_k we can minimize the norm of the difference of the two vectors. The norm used can be of any degree, but in QDT implementation of this thesis, we have chosen to use the infinite norm. We call this method the single-delta SDP. It gives the maximal distance between the corresponding terms in the vectors. For a more in-depth analysis, Ref. [55] goes through the method in more detail.

3. Methods

Although this thesis focuses on methods to improve measurement results in quantum computers, the same methods are applicable in other fields as well. For instance, POVMs are employed in the study of quantum thermometry and quantum metrology [31, 32]. Consequently, some of the techniques and approaches introduced in this work may also be beneficial in that context, further demonstrating the usefulness of these methods.

3.1 Ways of dealing with the noise

In this section, we will introduce methods for countering errors related to the measurement. QDT is used to perform error mitigation and the so-called *blending* procedure is used to deal with the temporal behavior of the aforementioned measurement noise.

3.1.1 Dealing with noisy detectors

As already mentioned, the readout on a modern quantum computer is noisy. Thus, the POVM effects used by the hardware are not given by Eq. (2.6) directly, but a version of it, which also incorporates the noise in the associated with the measurement. As long as the noise level is appropriately low, these noisy effects are still a valid IC-POVM. In this case, the data obtained with the noisy measurement can still be used to get unbiased estimators of any observable. To obtain the noisy effects, we used QDT as described in Sec. 2.4.

The IC-POVMs used in the experiments are independent for each of the qubits. Thus, also the QDT performed on the qubits to characterize the noisy effects can also be independent, and as a result, the QDT can be implemented in parallel to avoid a potentially exponential number of unique measurements that would have to be sampled in a non-parallel QDT strategy. While parallel QDT decreases the circuit overhead, it does not take into account any possible correlated measurement noise, also known as cross-talk. This type of noise has been shown to occur [54], so we should be vigilant for any signs of it affecting the performance of the parallel QDT scheme and possibly

biasing the estimations performed.

We can prepare the tomographical input states on each of the qubits at the same time and then measure the qubits in one of the measurement bases. For classical shadows, the measurement bases used are the Pauli operators $\{Z, X, Y\}$, which means that for each of the input states, only three unique circuits needs to be measured. The input states used in this work are $\{|0\rangle, |1\rangle, |+\rangle, |+i\rangle\}$. This results in the total number of unique QDT circuits being 12, which stays constant regardless of the number of qubits N. An additional upside of the parallel QDT is that the resulting noisy effects can be utilized in any local classical shadow-style POVM. This includes the locally biased classical shadows explained in Sec. 2.1.2 or any other local measurement scheme utilizing Pauli measurements.

3.1.2 Dealing with temporal noise

Although measurement errors have decreased over time, on most available quantum computers they are on the scale of 10^{-2}. IBM quantum computers fall into the same range, and the qubits in the Eagle r3 device used in this work typically have readout errors of 1-3 %. As mentioned in Sec. 2.4, quasi-static noise is simple to characterize. Thus, correcting for these systematic errors can be done efficiently using QDT, with only the downside of the variance of the estimation increasing as a result.

While the quasi-static noise is easy to characterize, the same is not true when the quantum computer has drift. The current superconducting qubit hardware has a well-known problem with temporal fluctuations that happen on long enough timescales to be considered as drift. These (usually unpredictable) fluctuations affect, for example, the readout fidelity of the qubits [56, 57]. This drift, further described in 2.2.2 complicates efforts to characterize the noise. It makes the measurement effects of the detector fluctuate in time. The noise often takes a form where it can be modeled as the detector randomly switching between two regimes of quasi-static noise. For instance, the so-called "good value" regime can have measurement errors of around 1-3%, while the "bad regime" has errors of around 5-10%.

The drift poses a great threat to the performance of the IC-POVMs and QDT, introduced in Sections 2.1.2 and 2.4. As QDT is intended to characterize the measurement effects of the IC-POVM, it is important that the detector characterized is the same as used by the POVMs. In the presence of drift, performing the QDT at a different time as the estimation of the observable may lead to QDT using a detector different from the POVMs. This would mean that the estimation of the observable mean value as given by Eq. (2.2) could be significantly biased, *i.e.* the accuracy of the estimation would suffer.

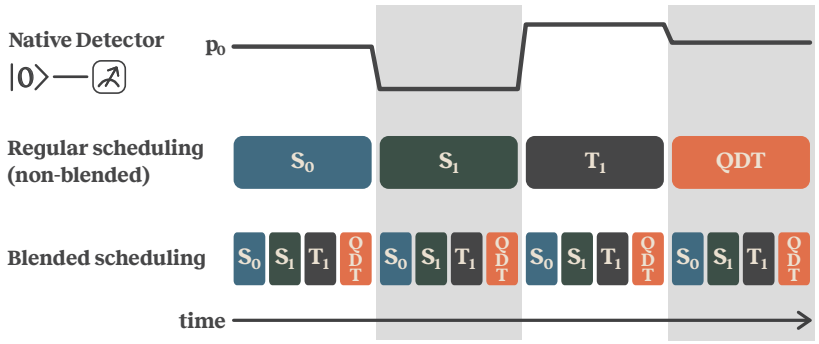


Figure 3.1: The diagram depicts a fluctuating detector, where the drift is clearly visible as a function of time. It is depicted as the probability of obtaining the outcome 0 when measuring $|0\rangle$ across the time window of all experiments. There are two way of scheduling the same experiments, the naive non-blended method performs each experiment fully before moving to the next, whereas in the blended method, the circuits are interleaved across the whole experiment. We measure the three states S_0, S_1 and T_1 as well as the required circuits needed to perform QDT. Due to drift in the device, in the naive case, each of the different circuits may use a detector different to the once used for other circuits. This mismatch may result in biased estimations of the observable and makes the results for each of the different states not comparable. When the circuits are scheduled in a blended way, we use the same average detector for each of the experiments. As a result the estimated energies of the states can be compared better and QDT characterizes the same average detector used in all the other experiments. Thus, we avoid any bias in the experiment and effectively mitigate the measurement errors in the device.

The effect of drift can be mitigated by removing any results obtained during the "bad" regimes. The method of "filtering" uses different statistical methods to remove the unwanted data to increase the accuracy of the estimate [56, 57]. While filtering is a common method used with near-term quantum devices, it has some downsides. Filtering out results from a time period where all qubits of a quantum computer are under-performing would be simple, but this is rarely possible. Our experimental observations have shown that different qubits fluctuate at different times. As the number of qubits used increases, the likelihood of one of the qubits being in the "bad" regime at any given time increases. As a result, we would be forced to throw out an increasing portion of the data, and thus the method is not viable when scaling into bigger system sizes.

We propose a new method for dealing with drift that relies on a time average that gives rise to an *unbiased* estimator of the expectation value of any observable. We call this procedure *blending*. With blending, samples for all experiments were taken evenly across the time window of the experiment. Effectively, the blending takes the noise from the regime of drift into quasi-static noise. The QDT characterization circuits are run alongside all the other experiments to ensure that fluctuation in the detector is accounted for. In Fig. 3.1, there is a diagram depicting a toy model of the drift and

how the naive scheduling, where each experiment is performed fully before moving to the next, differs from the blended scheduling. Blending marginalizes over the time window of the experiment, making the noise on the detector quasi-static. While the total noise level increases from using blending instead of filtering, it does not require discarding any of the data obtained, and thus it can still be used when moving to large numbers of qubits. Regardless of drift, it produces an unbiased estimator, where the level of noise determines how much the variance increases compared to not blending.

An analytical demonstration of blending is shown in Sec. 3.1.3 and details on how it was implemented on IBM hardware are available in Sec. 3.2. In addition to these, in Appendix B.3, we present a model for the drift encountered during the experiments on IBM Eagle r3. The noise model uses unbalanced random telegraph noise, to produce the random switching between the "good" and "bad" regimes. While the model can be used to model the noise well, the underlying sources of the noise are more complex as described in Section 2.2. The topic of unbalanced random telegraph noise is widely studied and we direct the reader to the following Refs. [58, 59, 60, 61, 62] for more in depth analysis.

In addition to time-averaging the noise present, blending also allows us to monitor the performance of the quantum computer over time. As all of the shots for the different experiments are spread out evenly throughout the full time window, we can monitor the data for any significant fluctuations. For a given experiment, the measurement of a single qubit in one of the measurement bases is expected to remain constant, and while this only gives information about local fluctuations, we can use it to *e.g.* track drift due to TLS noise. While the method cannot give us an in-depth look at the non-local fluctuations, it can still alert us on when there is a need to look further into, for example, the known issues of time-dependent noise of the two-qubit gates [56]. To determine the source of the drift, a comparison of QDT circuits and some experimental circuit with two-qubit gates can reveal whether the problem lies in the preparation of the measured state or in the measurement itself.

3.1.3 Details of the blending procedure

Let us go through the details of the blending method using an example. In order for us to perform QDT we need to measure each of the probe states, where for the sake of the example we end up needing data for M unique QDT circuits. We denote these circuits as $\{\mathcal{C}_{\mathrm{QDT}}^{(j)}\}_{j=1}^{M}$. In addition to the QDT circuits, we have a single circuit $\mathcal{C}_O$ used to estimate the expectation value of the observable O. For each of these circuits, we need to measure the circuit multiple times in order to collect enough statistics to reach a desired precision. For the sake of the example, we collect the same number of shots

S for each circuit. Thus, we need to get $(M+1)S$ shots in total. Collecting a singular shot takes a set amount of time, let us say τ seconds. In the naive scheduling method described in Sec. 3.1.2 we would first collect all S shots for the circuit $\mathcal{C}^{(1)}_{\mathrm{QDT}}$, before moving to collecting the shots for $\mathcal{C}^{(2)}_{\mathrm{QDT}}$ and so on until $\mathcal{C}^{(M)}_{\mathrm{QDT}}$. After the last QDT circuit we would collect all of the S shots for the $\mathcal{C}_{\mathrm{O}}$ circuit and finish the hardware run there. Now, if the temporal fluctuations of the qubits occur on a time scale of τS seconds, the detector characterized using the QDT circuits does not necessarily match the one used to obtain the statistics for the observable O at the end of the hardware run.

Ideally, in the implementation of the blending method, initially only a single shot is obtained for the circuit $\mathcal{C}^{(1)}_{\mathrm{QDT}}$, after which the circuit $\mathcal{C}^{(2)}_{\mathrm{QDT}}$ is sampled once. We continue with this until the circuit $\mathcal{C}^{(M)}_{\mathrm{QDT}}$ after which we sample a shot of the circuit $\mathcal{C}_{\mathrm{O}}$, before moving to the next round of samples. These rounds are all identical in the order of the circuits and they are repeated S times to obtain the desired statistics. We assume the detector to stay stable for the period of a single cycle of shots *i.e.* for $(M+1)\tau$ seconds but varying across the different cycles. For each of the S cycles we can estimate a different set of effects $\{\Pi_{i,\alpha}\}$, where i labels the outcome and $\alpha = 1, \ldots, S$ labels the cycle.

Assuming that the state being measured does not depend on time, the probability of obtaining the i^{th} outcome of the POVM depends on α according to:

$$p_{i,\alpha} = \mathrm{Tr}[\rho \Pi_{i,\alpha}]. \tag{3.1}$$

As we perform QDT using *all* of the data obtained across all of the cycles, assuming the input states to be constant, we construct a time-averaged POVM with the effects on qubit q being:

$$\Pi^{(q)}_{i,\mathrm{avg}} = \frac{1}{S} \sum_{\alpha} \Pi^{(q)}_{i,\alpha}. \tag{3.2}$$

This time-averaged POVM can be used to estimate the expectation value of the observable O in an unbiased manner. We can write the observable O as:

$$O_{\mathrm{avg}} = \sum_{i} \omega_{i,\mathrm{avg}} \bigotimes_{q=1}^{N} \Pi^{(q)}_{i,\mathrm{avg}} \tag{3.3}$$

using the results of QDT. Then using the outcomes of the circuit $\mathcal{C}_O$ across *all* of the S cycles, we can estimate the expectation value of O as:

$$\langle O_{\mathrm{avg}} \rangle = \mathrm{Tr}[\rho O_{\mathrm{avg}}] \tag{3.4}$$

$$= \sum_{i} \omega_{i,\mathrm{avg}} p_{i,\mathrm{avg}}. \tag{3.5}$$

Probabilities $p_{i,\mathrm{avg}}$ are experimentally estimated from the joint measurement of local IC-POVMs $\{\Pi^{(q)}_{i,\mathrm{avg}}\}_{q=1}^{N}$. With blending the expectation value of O can be estimated in

an unbiased way, even in the presence of drift. For the purposes of the experiment, the method transforms the drift into quasi-static noise with a higher variance compared to the shorter time-scale quasi-static noise.

3.1.4 Other methods

Whilst experimenting with shallow circuits that consist solely of single-qubit gates, as discussed in this thesis, the aforementioned methods are sufficient. In other terms, the errors related to the measurement of the state of the system can mostly be dealt with. When moving to circuits with two-qubit gates and more depth, it might be beneficial to also use other error suppression and error mitigation methods. These should be used in addition and not instead of the techniques already used with the Hartree-Fock circuits.

Most chemistry circuits are quite sparse and thus the qubits may be idle for extended periods of time. During these periods the qubits decohere and thus the method of dynamical decoupling is needed to minimize these idle periods. The method has been further discussed in Appendix A.1.

With longer circuits, the physical time and the errors related to the applied gates, especially the two-qubit gates, take the prepared state away from the desired one. While dynamical decoupling can help suppressing the errors, we need methods like ZNE, PEC or TEM [22, 23, 63, 64] in order to remove the effects of the noise from the data. These methods usually come with their downsides, such as the computational overhead related to learning the noise. There are many ways to characterize the noise of the device but often the noise is modified to take on a Pauli form, making it easier to learn and mitigate. The technique used to manipulate the noise is called Pauli twirling and it is explained in further detail in Appendix A.2.

3.2 Hardware implementation

Next, we will focus on implementing the previously described methods on IBM's quantum hardware. There are differences between hardware providers in the ways the circuits are run, but most of the information in this section is applicable to other providers as well. The methods used here have been shown to be useful when working with hardware providers such as IQM, Oxford Quantum Circuits and Quantum Circuits Inc. Although each device has their own "personality", they do benefit from similar approaches to measurement error. This is due to the fact that most of the characteristics and limitations arise from the underlying physics of the quantum hardware. Additionally, we introduce the method of *repeated settings* and construct a formula

to estimate the repeated settings variance. The method is used to reduce the circuit overhead, when estimating expectation values of observables with IC-measurements.

3.2.1 Implementation of the blending procedure on IBM

The experiments shown in this work were run on the IBM Cleveland Clinic quantum computer `ibm_cleveland` with an Eagle r3 processor. The experiments were submitted as smaller jobs that in this case were limited to 300 circuits per job and 10^5 shots per circuit. The execution of a job is inherently blended, where the quantum computer first samples a single shot from each of the circuits within a job before moving to sampling the second shot. This also means that each of the circuits within a job need to have the same number of shots.

To implement the blending method with the limitations of the submission pipeline, we split all of the different experiments* equally between each of the jobs. Splitting the experiments proportionally between the jobs results in each of the jobs consisting of a microscopic version of the full experiment set. As each of the circuits in a job has to be run the same number of times, to increase the number of shots sampled for QDT, we repeated the circuits for a given number of times to reach a desired amount of data. As mentioned in Sec. 2.4 we only had 12 unique circuits for QDT, and thus most of the circuits executed were still energy estimation circuits. In order to give some concrete numbers, with the experiments shown in Sec. 4.2 each of the jobs contained 240 energy estimation circuits, where S_0, S_1 and T_1 states each had 80 circuits in a job. The remaining 60 circuits were allocated for QDT, which means that we repeated each of the unique QDT circuits 5 times.

As mentioned in Sec. 3.1.2, the results of the QDT circuits can also be used for monitoring purposes. In the absence of drift, the results of a given QDT circuit for a given qubit stay constant. When temporal fluctuations occur, we can track them by comparing the results across the different jobs. Each job is a snapshot of the average performance of the device during the given period. This principle has been used to construct the graph in Fig. 4.4.

3.2.2 Repeated measurement settings for reducing circuit overhead

The hardware implementation of a dilation-free measurement scheme involves both the sampling of the measurement setting and the consequent execution of a corresponding measurement circuit. For each shot, the measurement bases are sampled for each of the

*The circuits needed to estimate the expectation value of a given state and the QDT circuits

qubits, to produce a measurement setting. In this work, we use Pauli measurements and thus the sampled measurement bases are from the set $\{Z, X, Y\}$.

In the case of an N-qubit POVM with r effects per qubit, there are r^N possible measurement settings, from which we sample randomly. As a result, when we move to large system sizes, the number of settings sampled will also be the number of circuits that need to be executed on the hardware. Contrary to the case of small system sizes, where we end up sampling the same setting multiple times and thus can collect multiple shots at the same time for a given circuit.

Many of the quantum hardware providers, including IBM have some overhead related to the initialization of a circuit when starting to sample the shots. This limitation, partly due to the electronics involved, introduce a problem when executing a large number of circuits with a low number of shots. For a quantum circuit, there is a constant time τ it takes to sample a single shot of the given circuit. In addition to τ, when starting to collect statistics for a new circuit, there is another constant time θ, where $\tau \ll \theta$, related to the initialization. This means that to collect 100 shots for a single circuit, the required time is $\theta + 100\tau$. With a high number of shots the effect of θ is not prominent, but in the one-shot-per-setting scheme it is approximately an order of magnitude larger than τ. In return, this slows down experiments on quantum hardware and makes the procedure inefficient.

To combat this inefficiency, we propose a simple change to the dilation-free POVM measurement scheme. Using *repeated settings* we reduce the total number of measurement settings needed by sampling each setting multiple times, while still producing unbiased estimates of mean and variance. This reduces the number of circuits that need to be executed, minimizing the impact of θ on the total runtime of the experiment. As a result, an experiment reaching a set precision can be performed within a shorter time period. To prove the above statements, we set the number of measurement settings to be S and the number of repeats of each of the settings to be T. For a given measurement setting i, we label the outcome j as $m_{i,j}$. The corresponding Monte Carlo estimator $\overline{O}$ for dilation-free POVMs is then given by:

$$\overline{O} = \sum_{i=1}^{S} \sum_{j=1}^{T} \frac{\omega_{m_{i,j}}}{ST}, \tag{3.6}$$

where the estimator represent the average of all ω_m across the measurement settings and their outcomes. When setting $T = 1$ we still recover the standard one-shot-per-setting estimator, whereas with $T > 1$ we start collecting multiple shots for each of the measurement settings. By allocating the shot budget using repeated setting, we avoid overheads related to large values of S and use the quantum computer more efficiently.

To show that the repeated setting estimator in Eq. (3.6) is truly unbiased, we

start from the expectation value of $\overline{O}$:

$$\mathbb{E}[\overline{O}] = \mathbb{E}\left[\sum_{i=1}^{S}\sum_{j=1}^{T}\frac{\omega_{m_{i,j}}}{ST}\right] \tag{3.7}$$

$$= \frac{1}{ST}\mathbb{E}\left[\sum_{i=1}^{S}\sum_{j=1}^{T}\omega_{m_{i,j}}\right] \tag{3.8}$$

$$= \frac{1}{ST}\sum_{i=1}^{S}\sum_{j=1}^{T}\mathbb{E}\left[\omega_{m_{i,j}}\right] \tag{3.9}$$

$$= \frac{\mathbb{E}\left[\omega\right]}{ST}\sum_{i=1}^{S}\sum_{j=1}^{T}1 \tag{3.10}$$

$$= \mathbb{E}\left[\omega\right], \tag{3.11}$$

where $\mathbb{E}\left[\omega\right]$ is the expectation value of ω_m, the coefficient in the decomposition $O = \sum_m \omega_m \Pi_m$. Furthermore, $\mathbb{E}\left[\omega_{m_{i,j}}\right] = \mathbb{E}\left[\omega\right]$ for any i and j. Next we write the expectation value of ω_m as:

$$\mathbb{E}\left[\omega\right] = \sum_m p_m \omega_m \tag{3.12}$$

$$= \sum_m \mathrm{Tr}\left[\rho\Pi_m\right]\omega_m \tag{3.13}$$

$$= \sum_m \mathrm{Tr}\left[\rho\Pi_m\omega_m\right] \tag{3.14}$$

$$= \mathrm{Tr}\left[\rho\sum_m \Pi_m\omega_m\right] \tag{3.15}$$

$$= \mathrm{Tr}\left[\rho O\right] = \langle O\rangle. \tag{3.16}$$

From the equation (3.16) it is evident that the first moment of $\overline{O}$ is equal to the expectation value of O and thus $\overline{O}$ is an unbiased estimator of $\langle O\rangle$.

To find the variance of the estimator, we start by calculating the second moment as follows:

$$\mathbb{E}\left[\left(\overline{O}\right)^2\right] = \mathbb{E}\left[\sum_{i=1}^{S}\sum_{j=1}^{T}\sum_{i'=1}^{S}\sum_{j'=1}^{T}\frac{\omega_{m_{i,j}}\omega_{m_{i',j'}}}{S^2T^2}\right] \tag{3.17}$$

$$= \frac{1}{S^2T^2}\sum_{i=1}^{S}\sum_{j=1}^{T}\sum_{i'=1}^{S}\sum_{j'=1}^{T}\mathbb{E}\left[\omega_{m_{i,j}}\omega_{m_{i',j'}}\right]. \tag{3.18}$$

From here, we proceed by considering all the possible combinations of $\omega_{m_{i,j}}$ and $\omega_{m_{i',j'}}$. There are a total of S cases where $i = i'$ and $j = j'$. For these terms we write the expectation values as $\mathbb{E}\left[\omega_{m_{i,j}}\omega_{m_{i,j}}\right] = \mathbb{E}\left[\omega^2\right]$ and thus these terms contribute to the total second moment by $\langle\omega^2\rangle/ST$.

Moving on to the cases where $i = i'$ and $j \neq j'$, of which there are $ST(T-1)$. This means that the measurement setting picked is the same between the two, but the

outcome differs. Plugging in $i = i'$ and $j \neq j'$ we get:

$$\mathbb{E}\left[\omega_{m_{i,j}}\omega_{m_{i'=i,j'}}\right] = \mathbb{E}\left[\omega_{m_{i,j}}\omega_{m_{i,j'}}\right] = \mathbb{E}\left[\mathbb{E}\left[\omega\right]_i^2\right], \tag{3.19}$$

where $\mathbb{E}\left[\omega\right]_i$ is the expectation value of all outcomes obtained with the measurement setting i. $\mathbb{E}\left[\mathbb{E}\left[\omega\right]_i^2\right]$ is denoted as $\langle\langle\omega\rangle_i^2\rangle$, we call it the *conditional second moment*. The contribution from these terms to the total second moment is $\langle\langle\omega\rangle_i^2\rangle(T-1)/ST$.

Finally, we consider the cases where $i \neq i'$, of which there are $S(S-1)T^2$ terms. They correspond to cases where the measurement setting used is not the same. In these cases, the expectation value is written as:

$$\mathbb{E}\left[\omega_{m_{i,j}}\omega_{m_{i',j'}}\right] = \mathbb{E}\left[\omega_{m_{i,j}}\right]\mathbb{E}\left[\omega_{m_{i',j'}}\right] = \mathbb{E}\left[\omega\right]^2. \tag{3.20}$$

The contribution of these term to the second moment is $\langle\omega\rangle^2(S-1)/S$.

We use the above to obtain the variance by summing all of the separate contributions together and subtracting the first moment squared:

$$\mathrm{Var}[\overline{O}] = \frac{\langle\omega^2\rangle}{ST} + \frac{\langle\langle\omega\rangle_i^2\rangle(T-1)}{ST} + \frac{\langle\omega\rangle^2(S-1)}{S} - \langle\omega\rangle^2 \tag{3.21}$$

$$= \frac{\langle\langle\omega\rangle_i^2\rangle - \langle\omega\rangle^2}{S} + \frac{\langle\omega^2\rangle - \langle\langle\omega\rangle_i^2\rangle}{ST}. \tag{3.22}$$

Using the above formula we can efficiently estimate the variance for the repeated settings scheme in an unbiased way.

The conditional second moment $\langle\langle\omega\rangle_i^2\rangle$ is bounded from above by the second moment $\langle\omega^2\rangle$ and from below by the first moment squared $\langle\omega\rangle^2$. This results in the following inequality:

$$\langle\omega\rangle^2 \leq \langle\langle\omega\rangle_i^2\rangle \leq \langle\omega^2\rangle. \tag{3.23}$$

In the case where the conditional second moment reaches the lower bound, *i.e.* $\langle\langle\omega\rangle_i^2\rangle = \langle\omega\rangle^2$, the variance can be reduced by allocating more shots for each measurement setting. On the other hand, at the upper bound, the variance cannot be reduced by increasing the shots per setting. In that case, the variance can only be decreased by sampling more measurement settings.

The repeated settings method can, at best, only be as good as the one-shot-per-setting scheme, and thus for a given number of shots it is always safer to use the latter if hardware limitations are not a problem. In some cases, there are multiple ways to allocate shots between S and T to reach the same variance, but there is a limit to the precision that can be reached by only increasing the shots per measurement setting T.

To further investigate this limit, we compare the variances of both the one-shot-per-setting scheme and the repeated settings scheme. The former is denoted as V_A

with $R \coloneqq S$ and $T = 1$, whereas the latter is denoted as V_B at the limit of $T \to \infty$. Both can be derived from Eq. (3.22):

$$V_A = \frac{\langle \omega^2 \rangle - \langle \omega \rangle^2}{R}, \tag{3.24}$$

$$V_B = \frac{\langle \langle \omega \rangle_i^2 \rangle - \langle \omega \rangle^2}{S}. \tag{3.25}$$

Next, we take the ratio of the two variances and use the fact that $\langle \omega^2 \rangle \geq \langle \langle \omega \rangle_i^2 \rangle$ to obtain the following inequality.

$$\frac{V_B}{V_A} = \frac{\langle \langle \omega \rangle_i^2 \rangle - \langle \omega \rangle^2}{\langle \omega^2 \rangle - \langle \omega \rangle^2} \frac{R}{S} \leq 1. \tag{3.26}$$

This inequality holds when $V_A > 0$, *i.e.* $\langle \omega^2 \rangle - \langle \omega \rangle^2 > 0$. Moving terms around we obtain:

$$\left(\frac{\langle \langle \omega \rangle_i^2 \rangle - \langle \omega \rangle^2}{\langle \omega^2 \rangle - \langle \omega \rangle^2} \right) R \leq S, \tag{3.27}$$

$$\left(1 - \frac{\langle \omega^2 \rangle - \langle \langle \omega \rangle_i^2 \rangle}{\langle \omega^2 \rangle - \langle \omega \rangle^2} \right) R \leq S. \tag{3.28}$$

Finally, we identify the term we call the *saving factor*:

$$S \geq \left(1 - \frac{\langle \omega^2 \rangle - \langle \langle \omega \rangle_i^2 \rangle}{\langle \omega^2 \rangle - \langle \omega \rangle^2} \right) R = F_{\text{saving}} R, \tag{3.29}$$

The quantity $F_{\text{saving}} R$ is used to find the number of measurement settings S, given $T \to \infty$, needed to reach the same precision as the one-shot-per-setting scheme with R shots. In practice, to reproduce the behavior at the limit $T \to \infty$ the shot per measurement setting T can be finite and often fairly low, depending on the value of the conditional second moment $\langle \langle \omega \rangle_i^2 \rangle$.

The saving factor is state, observable, and POVM dependent and is bounded by $0 \leq F_{\text{saving}} \leq 1$. Due to hardware limitations, it would be ideal that the value of F_{saving} would be small such that more shots could be allocated to T, while still reaching the same precision as the one-shot-per-setting scheme. This would decrease S and, in return, make the experiment run faster on the hardware.

3.3 Case study: energy estimation of the BODIPY-4 molecule

In this work we use the boron-dipyrromethene (BODIPY) molecule as a testbed of the methods introduced. BODIPY and its derivatives are a class of organic fluorescent dyes that possess favorable chemical and physical characteristics. These include, but are not

limited to, their high modularity, strong absorption of visible light, and significant fluorescence quantum yields. In addition, they are exceptionally stable both thermally and chemically. As a result, BODIPY compounds have wide-spread applications in medical imaging, biolabelling, optoelectronics, photoelectrochemistry, photocatalysis, artificial photosynthesis, and photo-dynamic therapy. This work is part of a larger project that focuses on simulating the aforementioned BODIPY compounds on a quantum computer. For further details, we refer the reader to Ref. [65].

To study the effectiveness of the methods introduced in this work, we used an initialization state of a ΔADAPT-VQE [66] procedure for an in-solvent BODIPY-4 molecule. In order to obtain results for different system sizes, we studied the molecule in the active spaces of 4e4o (8 qubits), 6e6o (12 qubits), 8e8o (16 qubits), 10e10o (20 qubits), 12e12o (24 qubits), and 14e14o (28 qubits). We estimated the energy of the ground state (S_0), the first excited single state (S_1) and the first excited triplet state (T_1) for each of the active spaces. In order to obtain the excited state energies, we used the techniques in Ref. [66] to generate Hamiltonians where the original excited states are ground states and then used the Hartree-Fock states of the said Hamiltonians.

For a given active space, all three Hamiltonians have the same number of Pauli strings as shown in Table 3.1.

Qubits	8	12	16	20	24	28
Num. Pauli strings	361	1819	5785	14243	29693	55323

Table 3.1: Number of Pauli strings in each of the Hamiltonians for a given system size. As mentioned in Sec. 2.1.1 the number of Pauli strings increase as a function of the system size as $\mathcal{O}(N^4)$ [8].

As our focus is on measurement fidelity, we have opted for circuits that do not use any two-qubit gates. The initialization state represented by the Hartree-Fock state is separable and thus suits the purpose well. Even though the measured states are simple, the Hamiltonians themselves have a complex structure and contain a large number of terms, as is evident from Table 3.1. Thus, reaching the chemical precision of 1.6×10^{-3} Hartree is not a trivial task even when measuring the Hartree-Fock state.

The goal of ΔADAPT-VQE is to estimate the gaps between the energies of the three states. To reach desired accuracies for these gaps, we utilize blending to avoid drift affecting the estimates. While there will still be quasi-static noise affecting the results, this should make the statistics for the different states as comparable as possible.

While the Hartree-Fock states were the same for all three states, we ran different circuits for each of them to test the Hamiltonian-inspired locally biased classical shadows measurement scheme. Although we used the same Pauli measurements for each state, the probability of sampling a basis for a given qubit differed between the states.

4. Experimental Results

To analyze the performance of the methods introduced in this work, we utilized them in experiments on a real quantum computer. Next, we will go through all the experiments conducted and their results. While analyzing the results, we used two types of errors to determine the accuracy and the precision of the estimates. For accuracy we used the absolute error *i.e.* the distance from the estimated energy E_{est} to the reference energy E_{ref} for the given problem. In general we do not have access to this reference energy, but due to the size of the problem we can use the reference value to determine the absolute error as $|E_{\text{est}} - E_{\text{ref}}|$.

To estimate the precision of E_{est}, we calculated the standard error of the estimated energy E_{est}. It was calculated as the square root of the estimator variance, as shown in Eq.(3.22).

The standard error can be interpreted as a likely distance from the sample mean to the true population mean. It estimates the amount of random statistical noise in the estimated mean. It is used to calculate confidence intervals, where there is a probability of 68% that the true mean lies within 1σ of the sample mean. Similarly 2σ refers to a probability of 95% and 3σ to a probability of 99.7%. The standard error can be used to determine whether an absolute error in the estimation of an energy is due to noise or if it is likely that some other source of error is effecting the results. For example, a noisy detector could bias the results significantly enough, such that the absolute error is consistently larger than even three times the standard error.

4.1 Reduction of estimation bias using QDT

In Sec. 2.4 we anticipated that QDT would decrease the absolute errors of energy estimates, while increasing their variance. In order to demonstrate the performance of QDT we consider the following 8-qubit experiment. We blend the QDT circuits with the energy estimation circuits for the S_0 Hartree-Fock state and measure the circuits on `ibm_cleveland`. The circuits were measured using the classical shadows measurement scheme. A single experiment consisted of $S = 7 \times 10^4$ measurement settings, where the number of shots per setting was $T = 100$. In total, the state S_0 was measured using

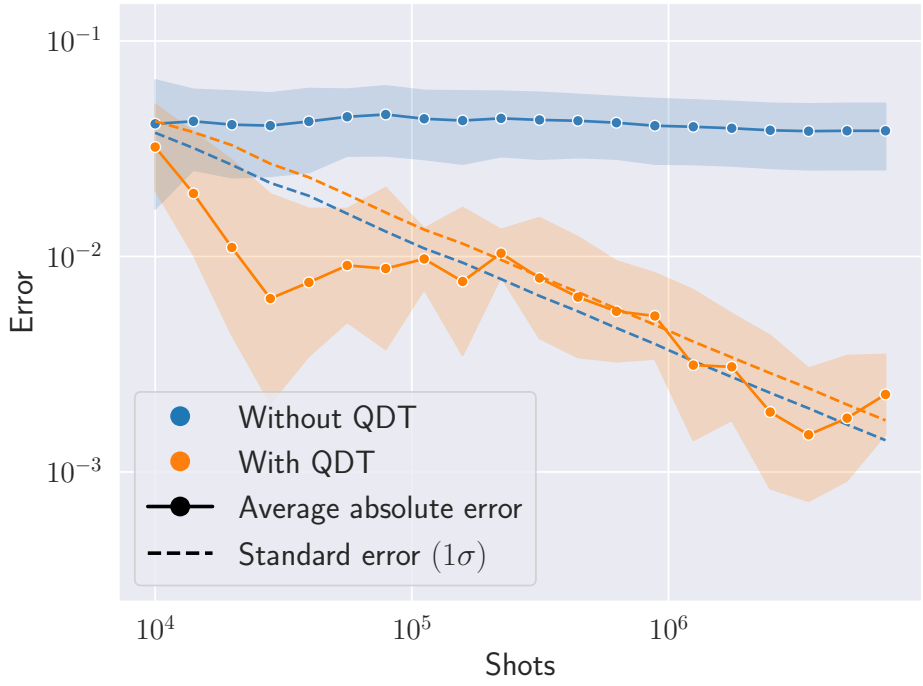


Figure 4.1: Errors corresponding to the estimation of the 8-qubit Hartree-Fock state of S_0 on `ibm_cleveland`. The measurements were performed using the classical shadows measurement scheme. The graph shows the errors of the estimated energies with and without QDT as a function of the number of shots over 10 repetitions. Each of the 10 repetitions consisted of $S = 7 \times 10^5$ measurement settings and $T = 100$ shots for each setting. The plot has been constructed by sampling varying sized data sets from the original and by calculating the means and standard errors of each. Thus, for a given color, the points all come from the same data set, where the final point takes into account all of the gathered data. It is a depiction of how increasing the number of shots affects the results. The shaded area represent the standard error of the estimated means at the 95% confidence interval. The absolute errors represent the difference between the noiseless energy of the state prepared, and the estimate obtained. The standard errors have been computed using the equation for repeated settings variance in Eq. (3.22).

$S = 7 \times 10^6$ shots. The circuits were divided into different jobs, where each QDT circuit was repeated 4 times. Thus, each job had 252 S_0 circuits and 48 QDT circuits, each measured 100 times. After splitting the circuits into individual jobs, there were a total of 278 jobs and 8.33×10^6 shots, including QDT. This means that each of the 12 QDT circuits was measured 1.11×10^5 times throughout the 45-50 minute experiment. To obtain statistics across multiple independent experiments, we performed the described experiment 10 times.

From the results, it is clear that noise in the measurement process affects the outcomes greatly. As shown in Fig. 4.1, the average sample means seem to converge on a value significantly different from the reference energy. When using the complete data set, the average absolute error is 3.83×10^{-2}, which makes the estimates significantly

biased. When we assume that the detector can be described using the ideal POVM effects, even increasing the sample size does not seem to decrease the average absolute error.

To obtain a correct description of the measurement effects, we characterize the detector using QDT and this gets our average absolute errors all the way down to 2.72×10^{-3}. For the 10 repetitions performed, the average standard error was 1.75×10^{-3}. Although the first results showed a clear bias, thanks to QDT, the estimator now obeys the $1/\sqrt{S}$ scaling of the standard error. This means that we reach errors nearing chemical precision (1.6×10^{-3}).

4.2 Measurement overhead of the measurement schemes

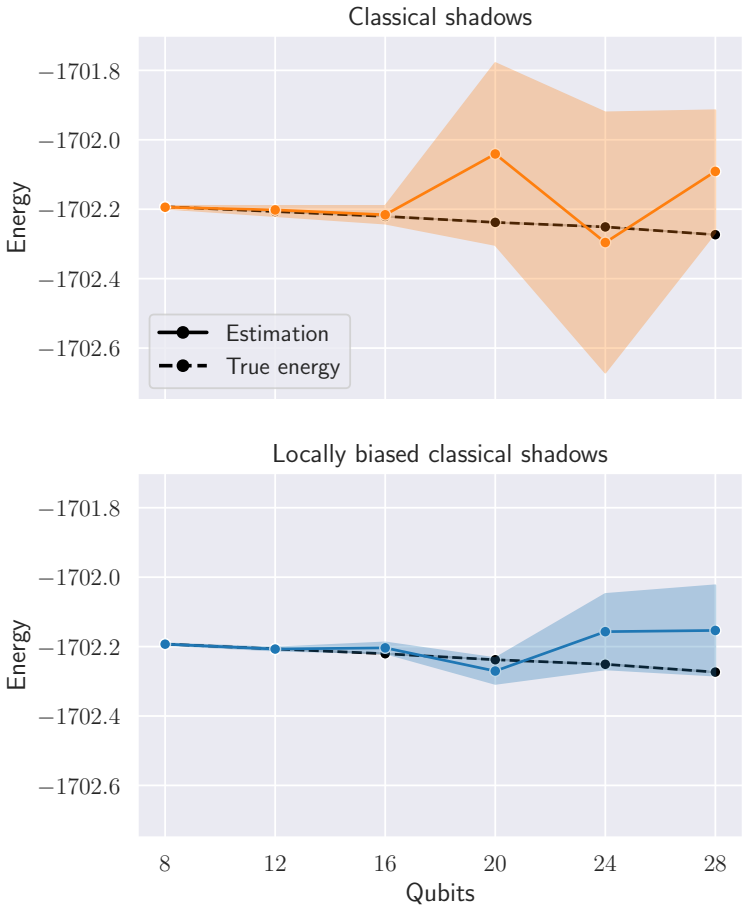


Figure 4.2: Comparison of the CS and LBCS measurement strategies when measuring the Hartree-Fock state of S_0 in different system sizes. The black dashed lines in the plots correspond to the reference energies for each system size. The shaded regions are the standard errors (1σ) for each of the estimated energies, determined using the equation for repeated settings variance in Eq. (3.22).

To validate the efficacy of the measurement techniques presented in this work, we conducted a set of experiments across all considered system sizes and Hamiltonians. In total, twelve distinct experiments were performed on `ibm_cleveland` quantum computer. In each of the experiments we blended the circuits corresponding to the three different initialization states and the QDT. This was done for all six different system sizes, while using both the classical shadows (CS) and locally biased classical shadows (LBCS) measurement schemes.

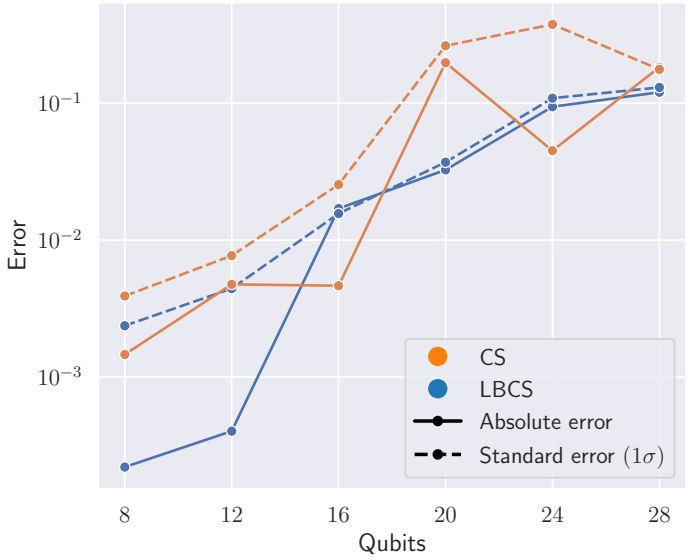


Figure 4.3: A direct comparison of the errors of LBCS and CS measurement schemes. The values correspond to the Hartree-Fock state of S_0 in various system sizes. The figure shows both the absolute error and the estimated standard error (1σ) for each case. Each data point consists of an experiment using $S = 3 \times 10^4$ measurement settings and $T = 100$ shots for each setting. QDT was used in order to mitigate the effects of measurement noise in the estimates. The absolute errors represent the difference between the noiseless energy of the state and the estimate obtained. The standard errors have been computed using the equation for repeated settings variance in Eq. (3.22).

In all twelve separate experiments, we used $S = 3 \times 10^4$ measurement settings with $T = 100$ for each of the three states. In each of the jobs consisting of 300 circuits, we repeated the twelve QDT circuits 3 times. Similarly to the other circuits, we sampled 100 shots for each of the QDT circuits. Across all 341 jobs, this totaled 1.02×10^5 shots per QDT circuit and 1.23×10^6 shots for the complete QDT procedure. Each of the twelve experiments took approximately 1 hour of quantum runtime.

To compare the CS and LBCS measurement schemes, we estimate the expectation values of the S_0 Hamiltonians as presented in Figures 4.2 and 4.3. All of the results utilize noisy POVM effects, characterized by QDT. The results for the S_1 and T_1 Hamiltonians are available in Appendix B.2. In Appendix B.1, we show that QDT reduces the absolute error of the estimated energies for all of these results and discuss

its effect on the variance of the estimate.

From the results obtained, we see that both measurement schemes yield energy estimates that lie within the error bars. While all of the estimates seem to be unbiased, the LBCS has consistently lower standard errors. The smaller variance means that we are getting more precise energy estimates, and thus the required sample size to reach a given level of confidence is smaller for LBCS.

Despite this advantage, we observe that the standard errors increase substantially with the number of qubits for both measurement strategies. This growing variance poses a significant challenge for high-precision experiments on near-term quantum hardware. It can be attributed to both the inherent limitations of the measurement schemes and the noise levels of the quantum hardware.

4.3 Characterization of temporal noise with blending

Finally, we look at an example of the drift we have observed during one of the experiments of Sec. 4.1. The experimental setup is further described in the aforementioned section. In addition to the original experiment, we also ran an additional job that included only QDT circuits, where each of the twelve circuits were sampled 10^5 times, totaling 1.2×10^6 shots. We used the results of this job to simulate a measurement scheme in which no blending was applied, and the energy estimation circuits were run independently of the QDT circuits. As described in Sec. 3.1.2, in the presence of drift, the detector characterized would not necessarily be the same used to measure the energy estimation circuits. The total quantum runtime of the experiment on `ibm_cleveland`, was approximately 50 minutes.

To monitor the drift, we look at the local measurement results for each of the qubits separately. In this experiment the qubit with index 1 is a model example of what the drift on modern quantum devices looks like. In Fig. 4.4 we show the results for the qubit with index 1, when it is measured in the Pauli Z basis. In both of the plotted cases the qubit has been prepared in the state $|0\rangle$ but one corresponds to the initialization state and has approximately 2×10^4 shots per job *i.e.* per data point, while the other is one of the QDT circuits with only 400 shots per data point.

The results are clearly noisy, but in the absence of drift, the plotted values are expected to remain constant. In the first twenty minutes of the experiment, this holds, as the fluctuations between jobs do not exceed the error bars, while still exhibiting quasi-static noise. After the 22 minute mark, especially the initialization state results with a larger set of data show a clear and statistically significant fluctuation. After the

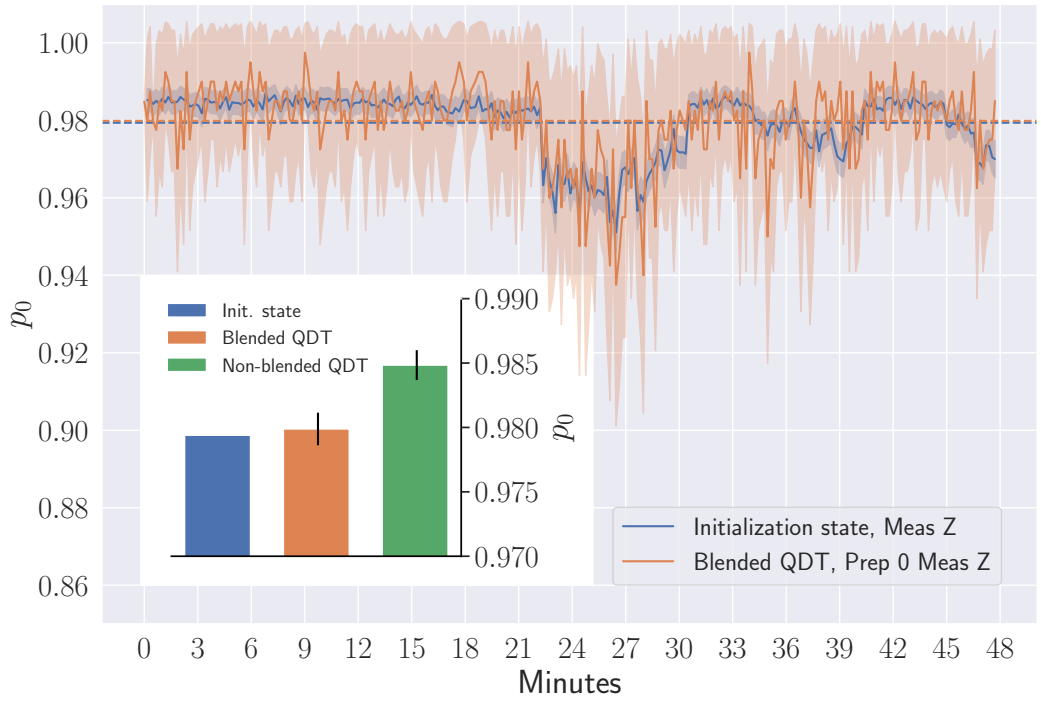


Figure 4.4: Comparison of the frequencies obtained for the initialization state in the Pauli Z basis and the QDT experiment with the measurement of $|0\rangle$ in the Pauli Z basis on the qubit with index 1 on `ibm_cleveland`. Each of the data points represent the total frequency of measuring the outcome zero in the given job. For the initialization state, each data point was computed from approximately 2×10^4 shots, while the QDT data points only consist of 400 shots. The line plot shows a significant fluctuation between the 22nd and the 30th minute, after which the fluctuations are not as drastic. The fluctuations are a clear example of drift described in Sec. 2.2.2, while the time before the drift represent the quasi-static noise described in the same section. In the inset bar plot we show the average frequencies for each of the three experiments, where the non-blended QDT does not appear in the line plot as it consist of data from a single job. All error bars represent 3σ, where the standard error for binomial random variables has been computed using $\sigma = \sqrt{p_0(1-p_0)/S}$.

clear fluctuation between the 22nd and 30th minutes, the detector fidelity returns to similar levels it was in the beginning but still remains more unstable up until the end of the experiment. With the QDT results, we do not have enough data to determine if the fluctuations are statistically significant but the same drift is still visible in the data.

For a clearer illustration, the inset in Fig. 4.4 shows how close the total frequencies of the two datasets are and how far off the non-blended frequency is. The frequencies from the two blended experiments match each other up to statistical noise even when the qubit experiences drift. As a result, we can determine that the blended QDT truly characterizes the same detector as is used for the energy estimation circuits. On the other hand, the non-blended QDT with much higher total frequency would not accurately describe the same detector. Thus, by using the resulting characterization,

we would not take all of the measurement noise in the experiment into account when estimating the expectation value of a given observable. In this case, the drift caused the energy estimation circuit frequency to differ from the non-blended frequency by 5×10^{-3}, while the blended method yielded an order of magnitude smaller discrepancy of 5×10^{-4}.

From these results, we can state that the drift on the quantum hardware occurs at the time scale of minutes, and thus the set of QDT circuit performed approximately every 100 milliseconds suffices to at least on average characterize the detector in the experiment. Although we do not show results using any other quantum computer, blending at these time scales is not always possible. In cases where blending only involves interleaving batches of energy estimation circuits with batches of QDT circuits, the performance, as expected, could suffer. For example, if the detector is characterized only every five minutes, there may have been a drop in performance and a subsequent return to the baseline within the time period. On the other hand, to minimize bias in the estimated energy, it is always better to blend the circuits in some way, even when the setup is suboptimal.

Due to the nature of drift, it can be difficult to detect from calibration data alone. Although we have observed qubits with higher readout fidelity to also exhibit fewer and weaker temporal fluctuations, there are no easy ways to avoid drift. Even in the cases where we can pick the very best qubits in a quantum computer, with larger circuits, we are forced to also use the "bad" qubits. Additionally, with more complex circuits it is important to factor in the two-qubit gate performance as it can often be more important than the measurement. It should be mentioned that, in the cases where the measurement noise is caused by TLS noise, the decrease in the qubit T_1 time obviously affects the quality of the results more generally.

Due to the aforementioned reasons, we believe that blending is a crucial part of a high-precision measurement. By averaging over all of the possible drift, we are able to take the drift into a regime of quasi-static noise.

5. Conclusions

In this work, we presented methods for performing high-precision measurements on current quantum hardware, where the average readout errors are high. To validate the methods, we performed experiments on a real quantum computer, where we estimated the expectation values of complex Hamiltonians. We discussed the noise and other limitations of current quantum hardware and introduced methods to overcome the resulting challenges.

We started by proposing a hardware-aware approach to the randomized measurement scheme, called *repeated settings*. By measuring each of the sampled measurement settings multiple times, we are able to reach the same precision for the estimated energy, while using less quantum resources. This reduction also decreases the resources required both in pre- and post-processing. To further reduce the measurement overhead, we used efficient observable-aware POVMs. With these locally biased classical shadow measurements we were able to consistently reach lower variances when estimating expectation values of complex BODIPY-4 Hamiltonians.

The noisy hardware introduces errors to the measurements performed on them, and as a result, the estimated energies suffer as well. To counteract the effects of any quasi-static noise, we applied parallel QDT. While mitigating the measurement errors, it also makes both the pre- and post-processing more efficient, compared to a non-parallel scheme. During the experiments, we have observed drift in the results, which could present a threat to the performance of the QDT method used. To take these temporal measurement errors from the regime of drift to quasi-static noise, we proposed the method of *blending*. By executing the QDT circuits in parallel with the other circuits used in the experiments, we characterize the same detector as used for all of the other circuits. With blending, we take into account noise across the different time scales of an experiment and thus are able to mitigate the effect of drift.

The methods mentioned were tested on IBM quantum computers, and while the devices between providers differ, the methods have already been used on other devices with small changes. In addition to the use cases on quantum computers, these methods may prove themselves to be useful wherever high-precision measurements are needed.

We acknowledge that the methods presented do not take into account some types

of noise, such as correlated measurement noise, referred to as cross-talk. The parallel QDT used treats each qubit as independent, and thus any possible cross-talk is not characterized. In practice cross-talk has been shown to occur [54] and thus its effects on IC-POVM measurement schemes should be studied, as well as possible methods to mitigate it. While in the experimental results of this work, we were able to achieve high precisions, without showing any significant bias, by taking cross-talk into account we might be able to further improve the performance of the pipeline.

From the experimental results it is clear that the errors grow very large when moving past system sizes of 16 qubits. While LBCS performed better than CS, the method is based on simple heuristics, and thus the performance could be improved significantly as in Refs. [40, 30]. Additionally, the variance of the estimate could be decreased by different measurement schemes and measurement noise suppression methods. Other ways of improving the results could be, for example, post-processing methods of dual optimization of Sec. 2.1.3 and TN-ICE [29]. Although in this work we have used QDT to perform measurement error mitigation, it is not the only method available. Other techniques include TREX [67], M3 [68] and those proposed for structured and sparse probability distributions in Refs. [69, 70]. The way these methods are affected by drift and how they could be applied to the IC-POVM formalism used should be studied thoroughly. The techniques introduced in this work should also be investigated in systems beyond separable Hartree-Fock states, in order to gain a deeper understanding of them.

Bibliography

[1] Peruzzo, A. *et al.* A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.* **5**, 1–7 (2014). URL http://dx.doi.org/10.1038/ncomms5213.

[2] Kandala, A. *et al.* Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **549**, 242–246 (2017). URL https://doi.org/10.1038/nature23879.

[3] O'Malley, P. *et al.* Scalable Quantum Simulation of Molecular Energies. *Physical Review X* **6**, 031007 (2016). URL https://doi.org/10.1103/PhysRevX.6.031007.

[4] Colless, J. *et al.* Computation of Molecular Spectra on a Quantum Processor with an Error-Resilient Algorithm. *Physical Review X* **8**, 011021 (2018). URL https://doi.org/10.1103/PhysRevX.8.011021.

[5] Hempel, C. *et al.* Quantum chemistry calculations on a trapped-ion quantum simulator. *Physical Review X* **8**, 031022 (2018). URL https://doi.org/10.1103/PhysRevX.8.031022.

[6] Robledo-Moreno, J. *et al.* Chemistry beyond exact solutions on a quantum-centric supercomputer. *preprint arXiv:2405.05068* (2024). URL https://arxiv.org/abs/2405.05068.

[7] Nützel, L. *et al.* Solving an industrially relevant quantum chemistry problem on quantum hardware. *preprint arXiv:2408.10801* (2024). URL https://arxiv.org/abs/2408.10801.

[8] Cao, Y. *et al.* Quantum chemistry in the age of quantum computing. *Chem. Rev.* **119**, 10856–10915 (2019). URL https://doi.org/10.1021/acs.chemrev.8b00803.

[9] Izmaylov, A. F., Yen, T.-C., Lang, R. A. & Verteletskyi, V. Unitary partitioning approach to the measurement problem in the variational quantum eigensolver

method. *Journal of Chemical Theory and Computation* **16**, 190–195 (2020). URL https://doi.org/10.1021/acs.jctc.9b00791.

[10] Verteletskyi, V., Yen, T.-C. & Izmaylov, A. F. Measurement optimization in the variational quantum eigensolver using a minimum clique cover. *The Journal of Chemical Physics* **152**, 124114 (2020). URL https://doi.org/10.1063/1.5141458.

[11] Huggins, W. J. *et al.* Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers. *npj Quantum Information* **7**, 1–9 (2021). URL https://doi.org/10.1038/s41534-020-00341-7.

[12] Crawford, O. *et al.* Efficient quantum measurement of pauli operators in the presence of finite sampling error. *Quantum* **5**, 385 (2021). URL https://doi.org/10.22331/q-2021-01-20-385.

[13] Miller, D., Fischer, L. E., Sokolov, I. O., Barkoutsos, P. K. & Tavernelli, I. Hardware-tailored diagonalization circuits (2022). URL https://arxiv.org/abs/2203.03646.

[14] D Ariano, G. M., Perinotti, P. & Sacchi, M. F. Informationally complete measurements and group representation. *Journal of Optics B: Quantum and Semiclassical Optics* **6**, S487–S491 (2004). URL http://dx.doi.org/10.1088/1464-4266/6/6/005.

[15] D'Ariano, G. M. & Perinotti, P. Optimal data processing for quantum measurements. *Phys. Rev. Lett.* **98**, 020403 (2007). URL https://link.aps.org/doi/10.1103/PhysRevLett.98.020403.

[16] Huang, H.-Y., Kueng, R. & Preskill, J. Predicting many properties of a quantum system from very few measurements. *Nat. Phys.* **16**, 1050–1057 (2020). URL https://www.nature.com/articles/s41567-020-0932-7.

[17] Elben, A. *et al.* The randomized measurement toolbox. *Nature Reviews Physics* **5**, 9–24 (2022). URL http://dx.doi.org/10.1038/s42254-022-00535-2.

[18] Dutt, A. *et al.* Practical benchmarking of randomized measurement methods for quantum chemistry hamiltonians. *arXiv preprint arXiv:2312.07497* (2023). URL https://arxiv.org/abs/2312.07497.

[19] Nykänen, A., Rossi, M. A. C., Borrelli, E.-M., Maniscalco, S. & García-Pérez, G. Mitigating the measurement overhead of ADAPT-VQE with optimised informa-

tionally complete generalised measurements. *preprint arXiv:2212.09719* (2023). URL https://arxiv.org/abs/2212.09719.

[20] Morrone, D., Talarico, N. W., Cattaneo, M. & Rossi, M. A. C. Estimating molecular thermal averages with the quantum equation of motion and informationally complete measurements. *Entropy* **26**, 722 (2024). URL https://www.mdpi.com/1099-4300/26/9/722.

[21] Fitzpatrick, A., Talarico, N. W., Eikås, R. D. R. & Knecht, S. Quantum-centric strong and dynamical electron correlation: A resource-efficient second-order n-electron valence perturbation theory formulation for near-term quantum devices. *preprint arXiv:2405.15422* (2024). URL https://arxiv.org/abs/2405.15422.

[22] Filippov, S. *et al.* Matrix product channel: Variationally optimized quantum tensor network to mitigate noise and reduce errors for the variational quantum eigensolver. *arXiv preprint arXiv:2212.10225* (2022). URL https://arxiv.org/abs/2212.10225.

[23] Filippov, S., Leahy, M., Rossi, M. A. C. & García-Pérez, G. Scalable tensor-network error mitigation for near-term quantum computing. *arXiv preprint arXiv:2307.11740* (2023). URL https://arxiv.org/abs/2307.11740.

[24] Maciejewski, F. B., Zimborás, Z. & Oszmaniec, M. Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography. *Quantum* **4** (2020). URL https://quantum-journal.org/papers/q-2020-04-24-257/.

[25] Glos, A. *et al.* Adaptive POVM implementations and measurement error mitigation strategies for near-term quantum devices. *arXiv preprint arXiv:2307.11740* (2022). URL https://arxiv.org/abs/2208.07817.

[26] Malmi, J., Korhonen, K., Cavalcanti, D. & García-Pérez, G. Enhanced observable estimation through classical optimization of informationally overcomplete measurement data: Beyond classical shadows. *Phys. Rev. A* **109**, 062412 (2024). URL https://link.aps.org/doi/10.1103/PhysRevA.109.062412.

[27] Fischer, L. E., Dao, T., Tavernelli, I. & Tacchino, F. Dual-frame optimization for informationally complete quantum measurements. *Phys. Rev. A* **109**, 062415 (2024). URL https://link.aps.org/doi/10.1103/PhysRevA.109.062415.

[28] Caprotti, A., Morris, J. & Dakić, B. Optimising quantum tomography via shadow inversion. *arXiv preprint arXiv:2402.06727* (2024). URL https://arxiv.org/abs/2402.06727.

[29] Mangini, S. & Cavalcanti, D. Low-variance observable estimation with informationally-complete measurements and tensor networks. *arXiv preprint arXiv:2407.02923* (2024). URL https://arxiv.org/abs/2407.02923.

[30] Hadfield, C., Bravyi, S., Raymond, R. & Mezzacapo, A. Measurements of Quantum Hamiltonians with Locally-Biased Classical Shadows. *Commun. Math. Phys.* **391**, 951–967 (2022). URL https://link.springer.com/article/10.1007/s00220-022-04343-8.

[31] Khan, M. M., Mehboudi, M., Terças, H., Lewenstein, M. & Garcia-March, M. A. Subnanokelvin thermometry of an interacting d-dimensional homogeneous bose gas. *Phys. Rev. Res.* **4**, 023191 (2022). URL https://link.aps.org/doi/10.1103/PhysRevResearch.4.023191.

[32] Ngo, T. V., Tsarev, D. V., Lee, R.-K. & Alodjants, A. P. Bose–einstein condensate soliton qubit states for metrological applications. *Scientific Reports* **11**, 19363 (2021). URL https://doi.org/10.1038/s41598-021-97971-4.

[33] Korhonen, K. *et al.* Practical techniques for high precision measurements on near-term quantum hardware: a case study in molecular energy estimation (2024). URL https://arxiv.org/abs/2409.02575. 2409.02575.

[34] Preskill, J. Quantum Computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018). URL https://doi.org/10.22331/q-2018-08-06-79.

[35] Zimborás, Z. *et al.* Myths around quantum computation before full fault tolerance: What no-go theorems rule out and what they don't (2025). URL https://arxiv.org/abs/2501.05694. 2501.05694.

[36] Cai, J.-Y. Shor's algorithm does not factor large integers in the presence of noise. *Science China Information Sciences* **67**, 173501 (2024). URL https://doi.org/10.1007/s11432-023-3961-3.

[37] Jena, A., Genin, S. & Mosca, M. Pauli partitioning with respect to gate sets (2019). URL https://arxiv.org/abs/1907.07859. 1907.07859.

[38] Koutný, D., Motka, L., Hradil, Z., Řeháček, J. & Sánchez-Soto, L. L. Neural-network quantum state tomography. *Physical Review A* **106** (2022). URL http://dx.doi.org/10.1103/PhysRevA.106.012409.

[39] Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010).

[40] García-Pérez, G. *et al.* Learning to Measure: Adaptive Informationally Complete Generalized Measurements for Quantum Algorithms. *PRX Quantum* **2**, 040342 (2021). URL https://doi.org/10.1103/prxquantum.2.040342.

[41] Nguyen, H. C., Bönsel, J. L., Steinberg, J. & Gühne, O. Optimizing shadow tomography with generalized measurements. *Physical Review Letters* **129** (2022). URL http://dx.doi.org/10.1103/PhysRevLett.129.220502.

[42] Zhu, H. Quantum state estimation with informationally overcomplete measurements. *Phys. Rev. A* **90**, 012115 (2014). URL https://link.aps.org/doi/10.1103/PhysRevA.90.012115.

[43] Innocenti, L. *et al.* Shadow tomography on general measurement frames. *PRX Quantum* **4**, 040328 (2023). URL https://link.aps.org/doi/10.1103/PRXQuantum.4.040328.

[44] Lisenfeld, J., Bilmes, A. & Ustinov, A. V. Enhancing the coherence of superconducting quantum bits with electric fields. *npj Quantum Information* **9**, 8 (2023). URL https://doi.org/10.1038/s41534-023-00678-9.

[45] Nachman, B., Urbanek, M., de Jong, W. A. & Bauer, C. W. Unfolding quantum computer readout noise. *npj Quantum Information* **6**, 84 (2020). URL https://doi.org/10.1038/s41534-020-00309-7.

[46] White, G. A. L., Jurcevic, P., Hill, C. D. & Modi, K. Unifying non-markovian characterisation with an efficient and self-consistent framework (2023). URL https://arxiv.org/abs/2312.08454. 2312.08454.

[47] Agarwal, A., Lindoy, L. P., Lall, D., Jamet, F. & Rungger, I. Modelling non-markovian noise in driven superconducting qubits. *Quantum Science and Technology* **9**, 035017 (2024). URL https://dx.doi.org/10.1088/2058-9565/ad3d7e.

[48] Agarwal, A. *et al.* Fast-tracking and disentangling of qubit noise fluctuations using minimal-data averaging and hierarchical discrete fluctuation auto-segmentation (2025). URL https://arxiv.org/abs/2505.23622. 2505.23622.

[49] Müller, C., Cole, J. H. & Lisenfeld, J. Towards understanding two-level-systems in amorphous solids: insights from quantum circuits. *Reports on Progress in Physics* **82**, 124501 (2019). URL https://dx.doi.org/10.1088/1361-6633/ab3a7e.

[50] Wang, Q. *et al.* Evidence of memory effects in the dynamics of two-level system defect ensembles using broadband, cryogenic transient dielectric spectroscopy (2025). URL https://arxiv.org/abs/2505.18263. 2505.18263.

[51] Nation, P. D., Kang, H., Sundaresan, N. & Gambetta, J. M. Scalable mitigation of measurement errors on quantum computers. *PRX Quantum* **2**, 040326 (2021). URL https://link.aps.org/doi/10.1103/PRXQuantum.2.040326.

[52] Chou, K. S. *et al.* A superconducting dual-rail cavity qubit with erasure-detected logical measurements. *Nature Physics* **20**, 1454–1460 (2024). URL https://doi.org/10.1038/s41567-024-02539-4.

[53] Lundeen, J. S. *et al.* Tomography of quantum detectors. *Nature Physics* **5**, 27–30 (2009). URL http://www.nature.com/articles/nphys1133.

[54] Maciejewski, F. B., Baccari, F., Zimborás, Z. & Oszmaniec, M. Modeling and mitigation of cross-talk effects in readout noise with applications to the Quantum Approximate Optimization Algorithm. *Quantum* **5**, 464 (2021). URL https://quantum-journal.org/papers/q-2021-06-01-464/.

[55] Cattaneo, M. *et al.* Self-consistent quantum measurement tomography based on semidefinite programming. *Physical Review Research* **5**, 033154 (2023). URL https://link.aps.org/doi/10.1103/PhysRevResearch.5.033154.

[56] Kim, Y. *et al.* Evidence for the utility of quantum computing before fault tolerance. *Nature* **618**, 500–505 (2023). URL https://www.nature.com/articles/s41586-023-06096-3.

[57] Hirasaki, Y., Daimon, S., Itoko, T., Kanazawa, N. & Saitoh, E. Detection of temporal fluctuation in superconducting qubits for quantum error mitigation. *arXiv preprint arXiv:2307.04337* (2023). URL http://arxiv.org/abs/2307.04337.

[58] Galperin, Y. M., Altshuler, B. L., Bergli, J. & Shantsev, D. V. Non-gaussian low-frequency noise as a source of qubit decoherence. *Phys. Rev. Lett.* **96**, 097009 (2006). URL https://link.aps.org/doi/10.1103/PhysRevLett.96.097009.

[59] Wold, H. J., Brox, H., Galperin, Y. M. & Bergli, J. Decoherence of a qubit due to either a quantum fluctuator, or classical telegraph noise. *Phys. Rev. B* **86**, 205404 (2012). URL https://link.aps.org/doi/10.1103/PhysRevB.86.205404.

[60] Li, J. *et al.* Motional averaging in a superconducting qubit. *Nature Communications* **4** (2013). URL http://dx.doi.org/10.1038/ncomms2383.

[61] Lo Franco, R., D'Arrigo, A., Falci, G., Compagno, G. & Paladino, E. Spin-echo entanglement protection from random telegraph noise. *Physica Scripta* **T153**, 014043 (2013). URL http://dx.doi.org/10.1088/0031-8949/2013/T153/014043.

[62] Daniotti, S., Benedetti, C. & Paris, M. G. A. Qubit systems subject to unbalanced random telegraph noise: quantum correlations, non-markovianity and teleportation. *The European Physical Journal D* **72** (2018). URL http://dx.doi.org/10.1140/epjd/e2018-90450-x.

[63] Giurgica-Tiron, T., Hindy, Y., LaRose, R., Mari, A. & Zeng, W. J. Digital zero noise extrapolation for quantum error mitigation. In *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)*, 306–316 (2020).

[64] van den Berg, E., Minev, Z. K., Kandala, A. & Temme, K. Probabilistic error cancellation with sparse pauli–lindblad models on noisy quantum processors. *Nature Physics* **19**, 1116–1121 (2023). URL https://doi.org/10.1038/s41567-023-02042-2.

[65] Fitzpatrick, A. *et al.* Excitation Energies of a Thiophene-fused Boron Dipyrromethene (BODIPY) Fluorophore simulated on a Superconducting Quantum Computer. *In preparation* (2024).

[66] Nykänen, A. *et al.* Toward Accurate Calculation of Excitation Energies on Quantum Computers with ΔADAPT-VQE: A Case Study of BODIPY Derivatives. *J. Phys. Chem. Lett.* **15**, 7111–7117 (2024). URL https://doi.org/10.1021/acs.jpclett.4c01301.

[67] van den Berg, E., Minev, Z. K. & Temme, K. Model-free readout-error mitigation for quantum expectation values. *Phys. Rev. A* **105**, 032620 (2022). URL https://link.aps.org/doi/10.1103/PhysRevA.105.032620.

[68] Nation, P. D., Kang, H., Sundaresan, N. & Gambetta, J. M. Scalable mitigation of measurement errors on quantum computers. *PRX Quantum* **2**, 040326 (2021). URL https://journals.aps.org/prxquantum/abstract/10.1103/PRXQuantum.2.040326.

[69] Yang, B., Raymond, R. & Uno, S. Efficient quantum readout-error mitigation for sparse measurement outcomes of near-term quantum devices. *Phys. Rev. A* **106**, 012423 (2022). URL https://link.aps.org/doi/10.1103/PhysRevA.106.012423.

[70] Bravyi, S., Sheldon, S., Kandala, A., Mckay, D. C. & Gambetta, J. M. Mitigating measurement errors in multiqubit experiments. *Phys. Rev. A* **103**, 042605 (2021). URL https://link.aps.org/doi/10.1103/PhysRevA.103.042605.

[71] Pokharel, B., Anand, N., Fortman, B. & Lidar, D. A. Demonstration of fidelity improvement using dynamical decoupling with superconducting qubits. *Phys. Rev. Lett.* **121**, 220502 (2018). URL https://link.aps.org/doi/10.1103/PhysRevLett.121.220502.

[72] Wallman, J. J. & Emerson, J. Noise tailoring for scalable quantum computation via randomized compiling. *Phys. Rev. A* **94**, 052325 (2016). URL https://link.aps.org/doi/10.1103/PhysRevA.94.052325.

Appendix A. Error suppression techniques

A.1 Dynamical decoupling

Dynamical decoupling is an error suppression method countering the errors rising from idle qubits and it is especially important for sparse circuits, where between gates, a qubit is in idle for extended periods of time. These long idle periods are not as common for dense circuits, and thus the need for dynamical decoupling should always be assessed in a case by case basis.

By adding gates that add up to the identity when a qubit is in idle, one can counter coherent errors related to any unwanted interactions with other qubits. There are numerous ways to implement this method, but even the simplest ones like applying two X-gates to any qubits in idle can significantly suppress the errors in the circuit [71]. This method, as the one we will move to next, assumes that the single-qubit gates have fairly low error rates.

A.2 Pauli twirling

Pauli twirling, also known as randomized compiling, is a method that attempts to bring the noise on the quantum device from an arbitrary form to the Pauli form [72]. This makes it easier to characterize the effects of noise and subsequently use the noise models to do circuit noise mitigation such as TEM [22, 23]. Even alone it can combat coherent errors as they usually accumulate quadratically, while the Pauli error only accumulates linearly [72].

As mentioned in Sec. 2.2.3, the main source of control noise emerges from the use of two-qubit gates. With most devices the noise associated with them is an order of magnitude higher, compared to the single-qubit gates. As a result, it is common that only this dominant noise is addressed, and thus, in the case of twirling, only the two-qubits gates are twirled. Nevertheless, the implementations are still similar for all types of gates.

For each gate in the circuit, one samples a set of Pauli gates, two for a single-qubit gate, and four for a two-qubit one. They are chosen such that the logic of the original gate remains intact. When the circuits are executed on a quantum computer, there needs to be multiple instances of twirling of the same circuit. This is important to ensure that any errors or biases introduced by the twirling can be mitigated.

Quite often, the gates added by Pauli twirling can be “compiled away” so that

the total number of gates does not increase. In cases where this is not possible, the added noise from the single-qubit gates is minor compared to the noise associated with the two-qubit gates. As mentioned previously, it is often faster to run a single circuit with many shots than to run many circuits with a low number of shots. Due to this, the twirling could increase the total computation time, but when combined with randomized measurements, twirling can be implemented without any additional overhead.

Appendix B. Additional results from the experiments

B.1 Comparison of measurement schemes with and without QDT

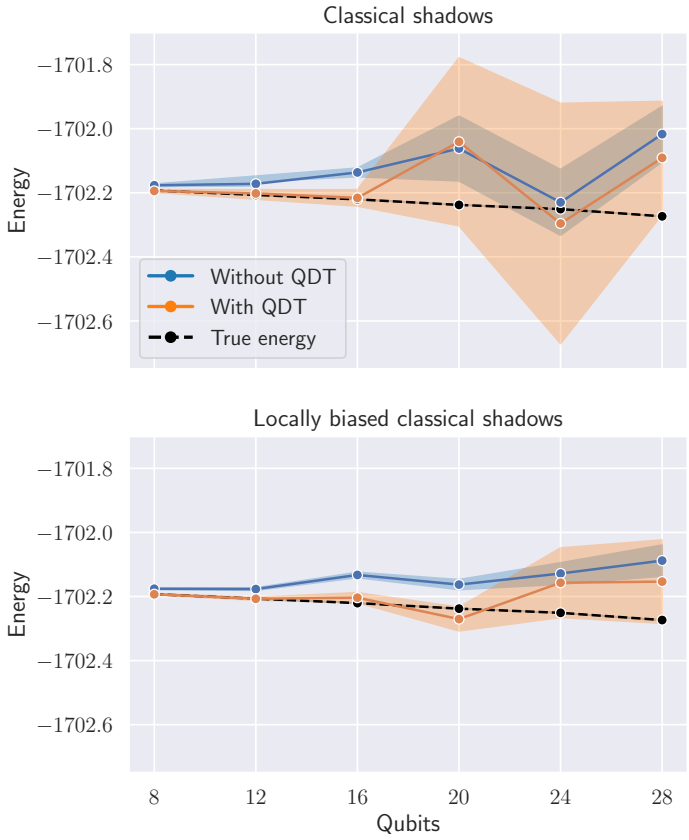


Figure B.1: A comparison between the S_0 energies estimated using CS and LBCS measurement schemes with and without QDT. The shaded regions are the standard errors (1σ) for each of the estimated energies, determined using the equation for repeated settings variance in Eq. (3.22).

In Sec. 4.1 we assessed the effectiveness of the QDT when estimating the expectation value of the 8-qubit S_0 Hamiltonian. Figs. B.1 and B.2 show how QDT affects the results across the different system sizes. From these results, we deduce that on average QDT increases the accuracy of the estimate.

The estimated energies computed without the use of QDT show a bias, where the reference value is not within the error bars. As this is a result of the errors in the quantum hardware, we can decrease this bias using QDT, as visible in the results. Although the standard error increases as a result of QDT, it is a worthy sacrifice as we obtain results with less bias, and the reference energies now fit within the error bars.

When the system size is increased, the absolute error and variance of the estimates also increase. This can be partially explained by larger system sizes requiring more shots in order to reach similar precision. Another partial reason for the behavior is that the increase in qubits increases the noise affecting the results. The qubits in a quantum computer are not uniformly noisy. When selecting the best possible set of qubits for a given experiments, when the system size is increased, it is likely that the additional qubits are noisier than the ones selected already. Nevertheless, from the results, it is clear that the LBCS method decreases the measurement overhead compared to CS, making the case for the method even stronger.

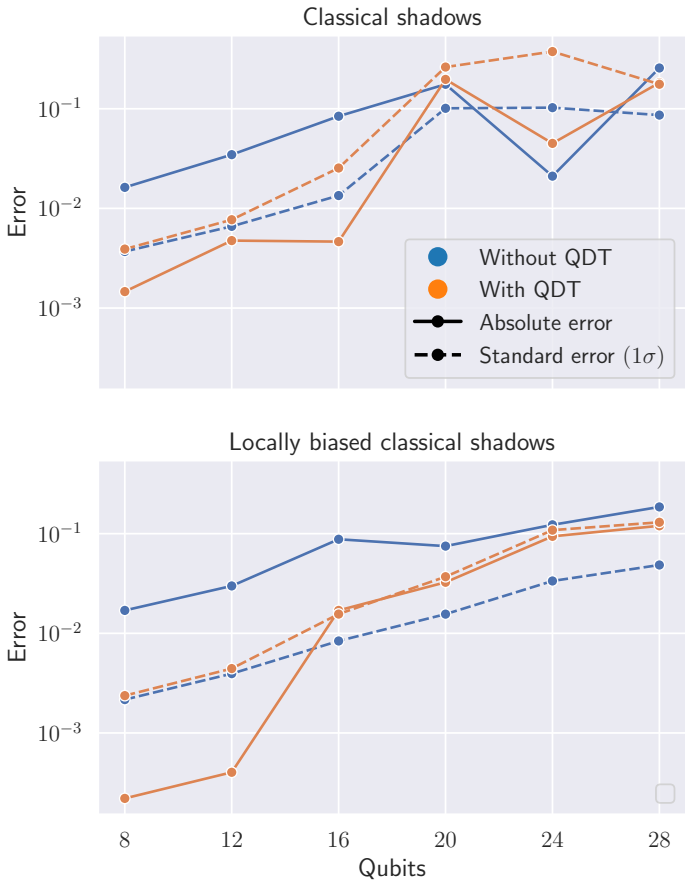


Figure B.2: The errors of the energies estimated using CS and LBCS measurement schemes with and without QDT. The error bars correspond to a standard error of 1σ, determined using the equation for repeated settings variance in Eq. (3.22).

B.2 Results for excited state Hamiltonians

When collecting the data presented in Sec. 4.2, for each of the 12 experiments, we blended together the circuits for the states S_0, S_1 and T_1 as well as the circuits for QDT. In the aforementioned section, only the results for S_0 are presented, and thus

in this section we present the results for S_1 and T_1. While the LBCS instance was unique for each of the three states, due to parallel QDT, the same set of QDT results can be used in all cases. While the three experiments did not necessarily use the same set of qubits, the QDT was performed independently for each qubit, and thus a QDT experiment involving the full set of overlapping qubits used sufficed. Since the four sets of circuits were blended, we can assume that any noise within the experiment can be approximated to be quasi-static. Of course, one cannot assume the quasi-static property across the twelve experiments as blending was not done on that scale. As a result, the experiments are not as comparable, even though blending allows each experiment to use a detector closer to the "average" one. The estimated energies of the excited-state Hamiltonians are shown in Fig. B.3. Across the data sets, the estimates are all unbiased, evident from the fact that in all cases except for one, the energy is within 1σ of the reference energy and when this is not the case, the true value is well within 2σ. Similarly to the case of S_0, it is evident that the LBCS performs much better than the CS, significantly reducing the variance of the estimated energy.

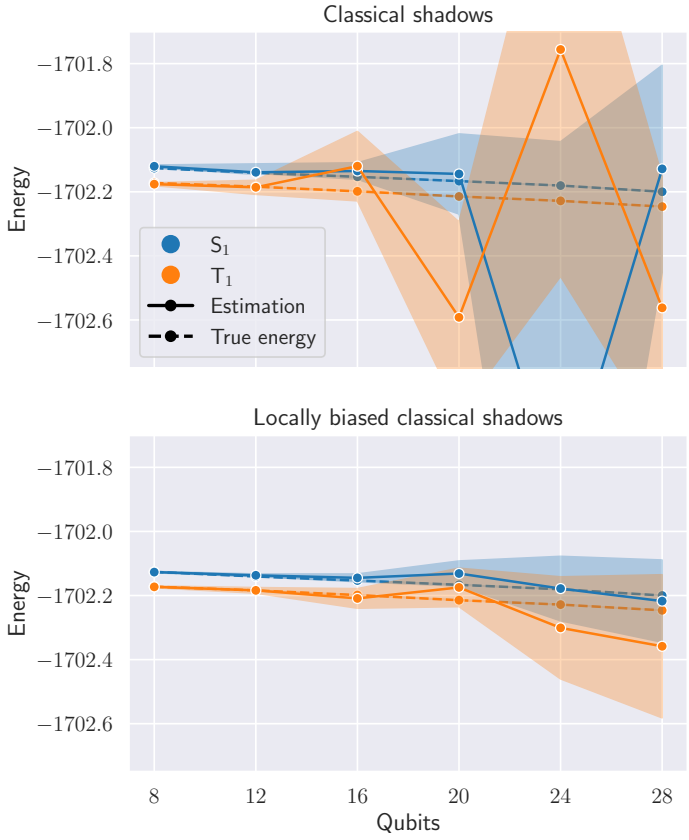


Figure B.3: Estimated energies for the S_1 and T_1 Hamiltonians in the various active spaces using the CS and LBCS measurement schemes and QDT. The error bars correspond to a standard error of 1σ, determined using the equation for repeated settings variance in Eq. (3.22).

B.3 Noise model for a fluctuating detector

In Sec. 3.1.2 we discussed drift in near-term quantum computers and present the method of blending to counter it. The method takes drift into the regime of quasi-static noise by scheduling the sampled circuits in an interleaved manner. In this section, we will present a noise model inspired by the noise we have encountered on real quantum computers. The noise model utilizes the observation that the detector seems to have "good" and "bad" regimes, between which it appears to switch according to a stochastic process. We model this as unbalanced random telegraph noise (RTN) with switching rates μ_{gb} and μ_{bg}, corresponding to the switching rates from a bad regime to a good regime or vice versa. The probability of switching from a good regime to a bad one within a time period t is given by:

$$p_{gb}(t) = e^{-\mu_{gb}t}\mu_{gb}t, \tag{B.1}$$

where a similar equation can also be written for the case where the detector switches from a bad regime to a good one. In order for the model to match the observations, we assume that $\mu_{gb} < \mu_{bg}$.

As discussed in Sec. 2.1.2 the measurement effect of an ideal detector for qubit q are $\{\Pi_i^{(q)}\}$. In the good regime, the detector is subject to quasi-static noise, where we model it as a small coherent error U_ϕ and a depolarizing error with probability p affecting the ideal effects. As the noise level is fairly small and quasi-static, the measurement strategy used is not influenced significantly. In the good regime, we can write down the measurement effects as:

$$\Pi_{i,\text{good}}^{(q)} = (1-p)\left[U_\phi \Pi_i^{(q)} U_\phi^\dagger\right] + p\frac{\mathbb{I}}{2}. \tag{B.2}$$

To model the noise in the bad regime, we add another term, a coherent error V_B. This term is larger than V_A and thus simulates the reduction in detector accuracy, observed when a detector experiences temporal fluctuations. This drift is harmful to the accuracy of our estimations, and thus we need to use blending to minimize its effects. We can describe the effects in the bad regime as:

$$\Pi_{i,\text{bad}}^{(q)} = (1-p)\left[U_\phi V_B \Pi_i^{(q)} V_B^\dagger U_\phi^\dagger\right] + q\frac{\mathbb{I}}{2}. \tag{B.3}$$

In Sec. 3.1.3 we use the index α to denote a specific time window, within which the noise is quasi-static. In the context of the noise model introduced, α selects whether the detector is in the good or bad regime.