Heinrich-Heine-University Düsseldorf


# Estimating Pauli Noise in Quantum Error Correction

Inaugural dissertation

presented to the Faculty of Mathematics and Natural Sciences
of Heinrich-Heine-University Düsseldorf
for the degree of
Doctor of Natural Sciences (Dr. rer. nat.)
by

**Thomas Wagner**
from Düsseldorf, Germany

07.12.2022

aus dem Institut für Theoretische Physik III
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. Dagmar Bruß
2. Prof. Dr. Martin Kliesch

Tag der mündlichen Prüfung: 27.02.2023

# Declaration

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.

*Düsseldorf, 07.12.2022*

Thomas Wagner

# Abstract

A detailed characterization of the noise affecting quantum devices is fundamental for their design, particularly for quantum error correction (QEC). Unfortunately, many characterization protocols are experimentally costly and require many measurements. Thus, it is an important task to obtain as much information as possible from easily available data, for example by tailoring protocols to specific applications.

Specifically in the context of QEC, a natural idea is to estimate the noise affecting a device from syndrome measurements that are performed anyway during standard error correction schemes. The main goal is to reduce the required effort for characterization by extracting additional information from already available data. Furthermore, this approach promises several other advantages. All components are characterized holistically in the context of the target application, which improves the detection of cross-talk. Furthermore, the resulting error models can be directly used with common decoders. Finally, since syndrome measurements preserve the encoded state, they allow for the online-characterization of a device during operation.

Unfortunately, efficient schemes for such estimation are currently only known for very specific error correction codes and noise models. Furthermore, there is a fundamental concern about *identifiability*. Since the class of syndrome measurements is necessarily quite limited, it is not clear under which conditions they are sufficient to uniquely estimate the parameters of a channel.

In this thesis, we address these problems and develop a general framework for the estimation of Pauli channels from syndrome data. Using this framework, we give comprehensive conditions under which a full characterization of the noise is possible. Furthermore, we consider the estimation of noise up to *logical equivalence*, i.e. focusing only on the information actually necessary for QEC. Here, we prove that the situation is as good as one could hope: estimation is possible as long as error correction itself is possible. We complement our fundamental results with efficient estimation protocols, which apply to arbitrary stabilizer codes. In contrast to previous proposals, these schemes neither require the computation of intractable likelihood functions, nor do they make heuristic assumptions about vanishing error rates. We quantify the performance of these estimators, both by providing a rigorous sample complexity bound and using simulations. The results suggest that noise estimation from syndrome data is a simple way to boost the performance of QEC schemes.

# Zusammenfassung

Eine detaillierte Charakterisierung des Rauschens auf Quantengeräten ist von grundlegender Bedeutung für ihre Entwicklung, insbesondere für Quantenfehlerkorrektur. Viele Charakterisierungsprotokolle sind jedoch experimentell aufwändig und erfordern eine große Anzahl Messungen. Es ist daher eine wichtige Aufgabe, so viele Informationen wie möglich aus leicht verfügbaren Daten zu gewinnen, zum Beispiel durch die Anpassung von Charakterisierungsprotokollen an spezifische Anwendungen.

Speziell im Kontext von Quantenfehlerkorrektur ist es eine naheliegende Idee, das Rauschen auf einem Gerät anhand von Syndrommessungen zu schätzen, die ohnehin im Verlauf von Standard-Fehlerkorrekturverfahren durchgeführt werden. Das Hauptziel besteht darin, den für die Charakterisierung erforderlichen Aufwand zu verringern, indem zusätzliche Informationen aus bereits verfügbaren Daten extrahiert werden. Darüber hinaus verspricht dieser Ansatz einige weitere Vorteile. Alle Komponenten werden ganzheitlich im Kontext der Zielandwendung charakterisiert, was die Erkennung von cross-talk verbessert. Weiterhin sind die resultierenden Fehlermodellen direkt mit gängigen Decodern kompatibel. Da Syndrommessungen den kodierten Zustand erhalten, ermöglichen sie zudem die Online-Charakterisierung eines Geräts während des Betriebs.

Leider sind effiziente Verfahren für eine solche Schätzung derzeit nur für sehr spezifische Fehlerkorrekturcodes und Rauschmodelle bekannt. Darüber hinaus besteht ein grundsätzliche Frage der *Identifizierbarkeit*. Da die Klasse der Syndrommessungen notwendigerweise sehr begrenzt ist, ist es nicht klar, unter welchen Bedingungen sie ausreichend für eine eindeutige Schätzung der Parameter eines Kanals sind.

In dieser Dissertation behandeln wir diese Probleme und entwickeln ein allgemeines Framework für die Schätzung von Pauli-Kanälen aus Syndromdaten. Mit Hilfe dieses Frameworks geben wir umfassende Bedingungen an, unter denen eine vollständige Charakterisierung des Rauschens möglich ist. Darüber hinaus betrachten wir die Schätzung des Rauschens bis auf *logische Äquivalenz*, d.h. wir konzentrieren uns nur auf die Informationen, die für Quantenfehlerkorrektur tatsächlich erforderlich sind. Hier beweisen wir, dass die Situation so gut ist, wie man nur hoffen kann: eine Schätzung ist möglich, solange die Fehlerkorrektur selbst möglich ist. Wir ergänzen unsere grundlegenden Ergebnisse mit praktischen und effizienten Schätzprotokollen, die für beliebige Stabilizer-Codes anwendbar sind. Im Gegensatz

zu früheren Vorschlägen erfordern diese Verfahren weder die aufwändige Berechnung von Likelihood-Funktionen, noch machen sie heuristische Annahmen über verschwindende Fehlerraten. Wir quantifizieren die Leistung dieser Schätzer, sowohl mit einer rigorosen Sample-Complexity Schranke als auch durch Simulationen. Die Ergebnisse legen nahe, dass die Schätzung des Rauschens aus Syndromdaten eine einfache Möglichkeit ist, die Effektivität von Quantenfehlerkorrekturverfahren zu steigern.

# Acknowledgement

I would like to express my sincere thanks to my supervisor, Dagmar Bruß, for introducing me both to the field of quantum information and research more generally, and for all the discussions and support. I am very grateful to my co-supervisor, Martin Kliesch, who was always available for in-depth and thoroughly enjoyable discussions and provided much insight and advice. His unwavering enthusiasm, even when I myself felt doubtful or stuck, was always a source of inspiration. I would also like to thank Herman Kampermann for his comments, advice and help.

Further thanks goes to all my present and past colleagues at HHU Düsseldorf, and especially Lucas Tendick for being a great office mate. Special thanks also goes to Julia Kunzelmann, Nikolai Wyderka and Christopher Cedzich for offering to proofread parts of this thesis, and to Jens Bremer for help with everything related to our computing cluster.

My deepest gratitude goes to my family for their continuous and unconditional support. First, I want to thank my parents, Cornelia and Ralf, who shaped the person I am today. I would also like to thank my brother, Martin, for quite a bit of tech and life advice. Finally, I want to express my deepest thanks to Wibke for her love, unwavering support, and so much more.

# Contents

# 1 Introduction

Quantum computation promises to solve many problems more efficiently than any classical computer. After Feynman's idea of a "quantum simulator" [Fey82], Shor's famous factoring algorithm [Sho94] demonstrated the potential of quantum computers to solve problems which are believed to be classically intractable. Since then, many other quantum algorithms have been developed that outperform our best known classical algorithms, such as Grover's search algorithm [Gro96], the Harrow-Hassidim-Lloyd algorithm for linear systems [HHL09], and the quantum counting algorithm [BHT98]. Consequently, quantum computing has been of great interest, both scientific and commercial, in areas such as quantum chemistry [Cao+19].

Unfortunately, quantum algorithms rely on fragile resources, such as entanglement, in the underlying quantum systems. Consequently, they are very susceptible to noise caused by unwanted interactions with the environment. This problem is much more pronounced for quantum computers than for modern classical computers, whose hardware is very reliable. Thus, there has historically been much scepticism whether large quantum computation could possibly be realized. Fortunately, the development of quantum error correction (QEC) and fault-tolerant quantum computation provides a potential path to implementing quantum algorithms even on noisy hardware. The famous threshold theorem [AB99a] states that, if the physical noise can be reduced below a certain threshold, then arbitrarily large quantum computation can be realized accurately and with only poly-logarithmic overhead. However, in practice, this required overhead is still too much for current quantum hardware. To alleviate this problem, the QEC community has been developing more and more efficient codes, decoders and fault-tolerant gates [BE21; CTV17].

This progress in QEC is complemented by progress in the field of quantum characterization and benchmarking [Eis+20; KR21]. Characterization and benchmarking is used to verify the success of existing noise reduction techniques, and to compare different protocols in a fair way. Characterization can also play a significant role for error correction itself. At least on current hardware, noise levels can vary strongly both in time [Etx+21] and between different qubits [TQ19]. A detailed understanding of the noise affecting a device is very beneficial for both hardware and software calibration. In particular, it is possible to tailor both QEC codes and the corresponding correction algorithms to the noise at hand. Thus, many characterization protocols have been developed, with gate set tomography [Blu+17]

and randomized benchmarking [Kni+08; Hel+22] being among the most popular ones. Unfortunately, these protocols are often resource-intensive, requiring many experimental runs of a device. Characterizing noise on a quantum device based on easily available information is thus an important topic. One approach is to move beyond general-purpose methods and instead consider protocols adapted to specific applications.

## 1.1 Motivation and Previous Work

In this thesis, we develop quantum characterization protocols tailored to the context of QEC. Standard error correction already relies on many so-called syndrome measurements, which are used to detect and correct errors affecting a device. Thus, there is rich set of data measured during QEC. In standard error correction, this data is used to compute corrections based on an already established noise model. Usually, these noise models are obtained by characterizing the device before operation.

Going beyond both standard QEC and traditional characterization, it seems like a natural idea to use the syndrome measurements themselves to obtain additional information about the underlying noise processes. This suggest an approach which is complementary to the traditional characterization before operation. The main motivation is to extract more information from easily available data that is measured anyway during QEC. Furthermore, such an approach has the additional advantage of characterizing all components holistically in the context of the target application, which can make it easier to detect crosstalk. It also provides a way to obtain error rates information that is directly applicable to current decoding schemes. Finally, one could also imagine the online adaptation of QEC to time-varying noise, since the noise information can be updated during operation. Such a scheme is illustrated in Figure 1.1. In fact, the syndrome measurements are the only measurements that can be used in online-estimation, since they are the only measurements that can be performed without destroying the encoded state.

For general codes and noise however, it is not a priori clear that an estimation of error rates just from syndrome data should be possible at all. Since the syndrome measurements preserve the encoded logical information, they necessarily contain limited information. This manifests in the fact that there are generally exponentially many errors with the same syndrome, which we therefore cannot distinguish. Furthermore, even if the estimation should be possible in principle, there are practical concerns. The probability of a syndrome is an exponentially large sum of different error rates, and solving such a system for the error rates appears difficult at best.

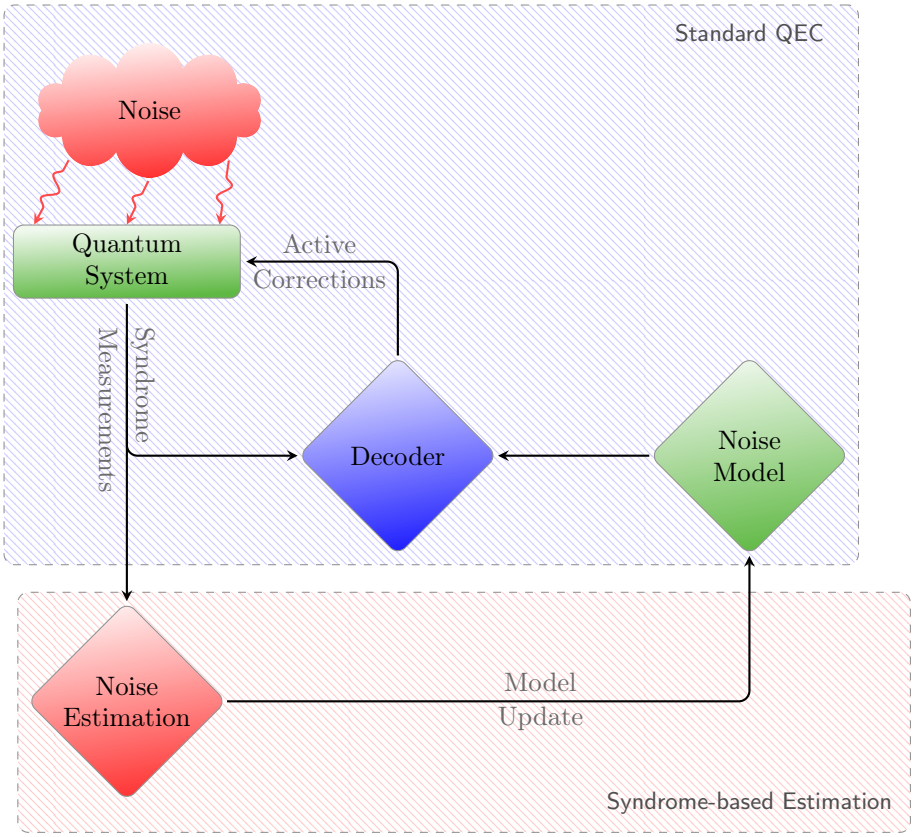



**Figure 1.1.:** Combined error correction and noise estimation using syndrome data.

Thus, there are two important questions about this novel estimation approach that need to be addressed. The first is the question of *identifiability*: Under which conditions is it possible (in principle) to uniquely estimate the parameters of a noise model from syndrome data? The second is how to design *practical algorithms* to actually perform the estimation.

The existing literature on this topic focuses mainly on the second question, treating the first as a bit of an afterthought, and offers approaches that fall into three broad categories:

The first category is *minimum-weight methods* [Fow+14; HL17; Woo20]. The key idea is that, if the error rates are small enough, only the smallest error matching each syndrome needs to be considered. All the other errors compatible with the syndrome can be ignored, since they are very unlikely. This results in a straightforward scheme to estimate error rates, simply by counting how often each different error occurred. However, this approach can only work in the limit of vanishing error rates, since otherwise high-weight errors can not be ignored. In Chapter 5, we demonstrate that "minimum-weight estimation" is clearly sub-optimal outside of this limit. Furthermore, if the error rates are small enough for the minimum-weight

approach to estimation to be successful, one might also expect simple minimum-weight decoding to perform very well, which requires no error rates information.

The second category is *likelihood-based methods* [FB16; Fuj14a; Com+14]. Here, the probability of each possible syndrome is explicitly expressed as a function of the parameters of our noise model, to obtain the likelihood of the observed data. Then, an estimate of the parameters can be computed via maximum-likelihood estimation. While this is a very general approach, there is a major drawback. As mentioned above, the probability of a syndrome is expressed as a sum over exponentially many compatible errors. Thus, the likelihood functions quickly become intractable, unless some simplifying assumptions are made. The most common assumption in this context is that the errors are independent between qubits, and the error rates of all qubits are equal. This greatly limits the applicability of such schemes.

Finally, for some specific codes, an *analytical expression* for the error rates in terms of the syndrome statistics has been developed by Spitz et al. [Spi+18]. This results in an efficient estimator, and also answers the question of identifiability. This scheme was used in experiments for decoder calibration in [Kri+22] and cross-talk analysis in [Che+21]. However, the applicability is limited to simple codes such as the surface and the repetition code. For example, the [4,1,2]-code considered experimentally in [Che+22] is not covered, and only approximate estimators were employed there. Furthermore, the scheme only applies to essentially classical noise models and independent errors, instead of more general models such as correlated Pauli noise.

## 1.2 Overview of Results

In this thesis, we develop a general framework for the estimation of error channels from the syndromes of a QEC code. This framework can be seen as a far-reaching generalization of the analytical expression developed by Spitz et al. [Spi+18], encompassing general codes and correlated noise. We use this framework both to answer the question of identifiability and to develop efficient estimation algorithms for arbitrary stabilizer, subsystem and data-syndrome codes. The estimator we propose has several desirable features. In particular, it

1. can be efficiently implemented,
2. applies to very general classes of codes,
3. does not rely on the assumption of vanishing error rates,
4. applies to Pauli noise, instead of just bit-flips,
5. and applies even to correlated errors.

As a specific application, we provide a scheme for decoder calibration from syndrome data for arbitrary stabilizer codes (Appendix A.1), generalizing the method of Spitz et al. [Spi+18] for the surface code. This provides an analytical solution for example for the calibration task considered in [Che+22], for which so far only approximate solutions were known. Furthermore, we show that even if the full channel is not identifiable, it is often still possible to extract all information that is relevant for QEC. More precisely, there are many errors that are *logically equivalent*, which means that they act in the same way on the encoded state. Such errors need not be distinguished for QEC. We consider the estimation of channels up to logical equivalence, and show that it is possible in many settings where the full channel can not be estimated.

The focus of our work is on the estimation of *Pauli channels*. One motivation for this is mathematical convenience. In contrast to general quantum channels, Pauli channels are much more analytically tractable, and we are able to prove rigorous results. This can be viewed as a first step to the study of more complex noise models. Another reason is that Pauli channels are commonly used in QEC. Most decoding schemes are designed for Pauli noise, and can incorporate Pauli error rates to improve their decoding. Finally, it is possible to map arbitrary quantum noise onto a Pauli channel via *randomized compiling* [WE16], implying that Pauli noise is more than a mere toy model. This has also been demonstrated experimentally [War+21]. For reasons similar to ours, the estimation of Pauli channels has recently received considerable attention, using a variety of approaches [FW20; FO21; HFW20; HYF21].

In addition to our main results about noise estimation from syndromes, we also contributed to a work about entanglement in grid states [Ghi+22]. This work is listed in Appendix E.

## 1.3 Structure of the Thesis

The content of this thesis is organized as follows. We start with three preliminary chapters, introducing the main concepts that we use in this work.

- In Chapter 2 we give some useful mathematical background, mainly on group characters and the related Fourier transform. This provides a useful language for the concepts of QEC, and our main results rely on these tools. We also give an overview over our notation here.
- We give a brief introduction to QEC in Chapter 3. We describe the most important classes of quantum codes, and also show how they can be viewed in a unified framework. For this purpose, we stress the connection to group characters, which is so far mostly implicit in the literature on QEC. We

explain the decoding of quantum codes, with focus on the importance of noise information for this task. Finally, we discuss common noise models that are used in QEC and their limitations, which also informs what kind of model we can ultimately estimate from syndrome data.

- In Chapter 4, we give some background on probabilistic graphical models (PGMs). PGMs are a useful tool from machine learning, and many of our results can be viewed through the perspective of PGMs. Particularly important are the concepts of factor graphs and Bayesian networks.

Then, we give an overview over the main results of our doctoral research.

- In Chapter 5, we discuss the estimation of Pauli channels from syndrome data for concatenated quantum codes. We introduce an algorithm based on tools from the learning of PGMs, and show simulations verifying its effectiveness. This algorithm has the advantage that it can be integrated with optimal decoding algorithms for concatenated codes. This is based on our published work [Wag+21].

- In Chapter 6, we present our general framework for the estimation of Pauli channels from syndrome data. We give comprehensive conditions under which the channels are identifiable, and develop an efficient estimation algorithm. We furthermore consider the estimation of channels up to logical equivalence, and show that it is possible under minimal conditions. This is based on our published works [Wag+22a] and [Wag+22b].

# Mathematical Preliminaries

# 2

> *The greatest challenge to any thinker is stating the problem in a way that will allow a solution*
>
> **— Bertrand Russell**

In this section, we collect some mathematical preliminaries, mainly about group characters and the Fourier transform on finite Abelian groups. We will see in Chapter 3 that the language of group characters is very convenient to express the concepts of QEC. In particular, it allows for a unified description of many classes of quantum codes, and avoids the somewhat cumbersome Pauli-to-binary isomorphism that is often used in the description of QEC. Many fundamental facts in QEC directly correspond to facts about group characters. Fourier analysis is the basis of our framework developed in Chapter 6. We introduce the basic definitions and the most important results for our purposes in an abstract setting, mostly following Mao and Kschischang [MK05], with some notation from Kalachev and Sadov [KS22]. Later, we are mostly concerned with the group $\mathbb{F}_2$ and the effective Pauli group P. Further information is available in the books by Fulton and Harris [FH13], and Terras [Ter99]. All proofs omitted in this section can also be found in the above sources.

## 2.1 Notation

We denote as $[n] \coloneqq \{1, \dots, n\}$ the set of the first $n$ positive integers. The field with two elements is denoted $\mathbb{F}_2$. For a statement $Q$, we denote with $[Q]$ the *Iverson bracket* of $Q$, which takes the value $1$ if $Q$ is true and $0$ if $Q$ is false. The *powerset* of a set $A$ is the set of all subsets of $A$, including the empty set, and it is denoted as $2^A$. We denote the four *Pauli matrices* as $I = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, $X = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, $Y = \left(\begin{smallmatrix} 0 & -i \\ i & 0 \end{smallmatrix}\right)$ and $Z = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. We also use $I$ for the generic identity matrix, or a generic identity element of a group. For an element $a = (a_1, \dots, a_n)$ of a group $A = \prod_{i=1}^n A_i$, we denote its *weight* with $|a| = |\{i \in [n] \,:\, a_i \neq I\}|$. The Pauli group $\mathcal{P}^n$ on $n$ qubits its the group of Pauli strings generates by the Pauli operators $\{I, X, Y, Z\}$ with phases,

$$\mathcal{P}^n = \{\epsilon \bigotimes_{i=1}^{n} e_i \,:\, \epsilon \in \{\pm 1, \pm i\},\, e_i \in \{I, X, Y, Z\},\, i \in [n]\}\,. \tag{2.1}$$

Often however, it is possible to ignore the phases, and work in the *effective Pauli group* $\mathsf{P}^n = \mathcal{P}/\{\pm 1, \pm i\}$. A *Pauli channel* $\Lambda_P$ is quantum channel that acts on the density matrix $\rho$ of a quantum state as

$$\Lambda_P(\rho) = \sum_{e \in \mathsf{P}^n} P(e) e\rho e \,. \tag{2.2}$$

Here, $P : \mathsf{P}^n \to [0, 1]$ is a probability distribution over Pauli errors. Since the channel is fully described by this distribution, we sometimes use "Pauli channel" also as a shorthand when referring to the distribution $P$. We do however reserve the symbol $\Lambda$ for the actual channels, and denote the distributions as $P$. Furthermore, the following notation is used throughout:

**Notation 1.** *For a group $A$ and a subgroup $B \subseteq A$, the* indicator function *of $B$ is given by*

$$\Phi_B(a) = [a \in B] \,, \tag{2.3}$$

*and the* scaled indicator function*, or* uniform probability distribution*, $U_B$ over $B$ is given by*

$$U_B = \frac{1}{|B|} \Phi_B \,. \tag{2.4}$$

## 2.2 Characters of Finite Abelian Groups

A *group character* of a finite Abelian group $A$ is a group homomorphism

$$\chi : A \to S^1 \,, \tag{2.5}$$

where $S^1 \coloneqq \{c \in \mathbb{C} : \, |c| = 1\}$ is the unit circle. The group characters of $A$ form a group $\widehat{A}$ under point-wise multiplication, which is called the *dual group* of $A$. In other words, $\widehat{A} = \mathrm{Hom}(A, S^1)$. This is similar to the notion of a dual vector space. So called Pontryagin duality guarantees that for any finite Abelian group $A$, its double dual $\widehat{\widehat{A}}$ is canonically isomorphic to $A$. Furthermore for finite Abelian groups, $\widehat{A}$ is isomorphic to $A$, but not canonically so. This is again similar to the situation for vector spaces.

For $\chi \in \widehat{A}$ and $b \in A$, we also use the notation $\langle \chi, b \rangle \coloneqq \chi(b)$. This is similar to the well known bracket-notation of quantum mechanics. Furthermore, we often want to identify $A \cong \widehat{A}$, such that $\chi$ is replaced by an element of $A$. This leads to the notion of a *bicharacter* [KS22], which is a convenient way to express a fixed ismorphism $A \cong \widehat{A}$.

**Definition 2.** *A* bicharacter *of a finite Abelian group $A$ is a map*

$$\langle \cdot, \cdot \rangle : A \times A \to S^1 , \tag{2.6}$$

*such that the map $a \mapsto \langle a, \cdot \rangle$ is an isomorphism of $A$ and $\widehat{A}$.*

A bicharacter behaves very similar to a scalar product from linear algebra, where the value $0$ for a scalar product is replaced with $+1$ for a bicharacter, since we map into a multiplicative group. Therefore, we also extend the bicharacter notation to matrices, which is useful in the context of QEC:

**Notation 3.** *For a matrix $H = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} \in A^{n \times m}$ and an element $a \in A^m$ we write*

$$\langle H, a \rangle = \begin{pmatrix} \langle h_1, a \rangle \\ \vdots \\ \langle h_n, a \rangle \end{pmatrix} \tag{2.7}$$

Note that if $\langle \cdot, \cdot \rangle$ is an ordinary scalar product, then this is just ordinary matrix multiplication.

Furthermore, we can define a notion of "orthogonal complement":

**Definition 4.** *Let $A$ be a finite Abelian group and $B \subseteq A$ a subgroup. The* annihilator *$B^\perp$ of $B$ is*

$$B^\perp = \left\{ \chi \in \widehat{A} : \chi(b) = 1 \, \forall \, b \in B \right\} . \tag{2.8}$$

While properly the annihilator is defined as a subgroup of $\widehat{A}$, we usually make implicit use of the isomorphism $A \cong \widehat{A}$ and view $B^\perp$ as a subgroup of $A$:

$$B^\perp = \{ a \in A : \langle a, b \rangle = +1 \, \forall \, b \in B \} . \tag{2.9}$$

A key difference between a bicharacter and a scalar product is that we often have $\langle a, a \rangle = 1$, i.e. it is not positive-definite in the linear algebra sense. Thus, it is possible that $B \subseteq B^\perp$. However, we always have $(B^\perp)^\perp = B$. Furthermore, taking the annihilator reverses the order of inclusions. That is, for any two subgroups $B, C \subseteq A$, if $B \subseteq C$, then $C^\perp \subseteq B^\perp$. Finally, the following property of the annihilator is useful later in Section 3.2. A proof can also be found e.g. in [Ter99, Chapter 12, Lemma 1].

**Lemma 5.** *For any finite Abelian group $A$ and any subgroup $B \subseteq A$,*

$$B^\perp \cong A/B . \tag{2.10}$$

*Proof.* Let $q : A \to A/B$ be the quotient map. Then, we can check that the dual map

$$\begin{aligned} \widehat{q} : \widehat{A/B} &\to B^{\perp} \\ \chi &\mapsto \chi \circ q \end{aligned} \tag{2.11}$$

is an isomorphism. The fact that a finite Abelian group is isomorphic to its dual group completes the proof. □

### 2.2.1 The Fourier Transform

Using group characters, it is possible to define the Fourier transform for any finite Abelian group. In fact, this theory can be developed even more generally for locally compact Abelian groups [Rud90], but we are not concerned with this generalization.

**Definition 6.** *Let $A$ be a finite Abelian group. The* Fourier transform *of a complex valued function $f : A \to \mathbb{C}$ is the function $\mathscr{F}[f] : \widehat{A} \to \mathbb{C}$ given by*

$$\mathscr{F}[f](\chi) = \sum_{b \in A} \langle \chi, b \rangle f(b) \,. \tag{2.12}$$

Usually, we identify $A$ and $\widehat{A}$ and view $\mathscr{F}[f]$ as a function on $A$. The most important lemma connected to the Fourier transform is the following:

**Lemma 7.** *If $B$ is a subgroup of a finite Abelian group $A$ and $\chi$ is a character of $A$, then*

$$\sum_{b \in B} \langle \chi, b \rangle = |B| \Phi_{B^{\perp}}(\chi) \,. \tag{2.13}$$

A proof can be found e.g. in [MK05, Lemma 1]. In particular, setting $B = A$ and identifiying $\widehat{A}$ and $A$, one obtains for any $a \in A$,

$$\sum_{b \in A} \langle a, b \rangle = |A| [a = I] \,. \tag{2.14}$$

Using this lemma, one can directly verify that the Fourier transform is an invertible function, with inverse given by

$$\mathscr{F}^{-1}[g](a) = \frac{1}{|A|} \sum_{\chi \in \widehat{A}} \langle \chi, a^{-1} \rangle g(\chi) \,, \tag{2.15}$$

for $g : \widehat{A} \to \mathbb{C}$.

Another important property of the Fourier transform is related to convolutions.

**Definition 8.** *The* convolution *of two complex valued functions $f, g, : A \to \mathbb{C}$ of a finite Abelian group $A$ is given by*

$$(f * g)(a) = \sum_{b \in A} f(b) g(ab^{-1}) \tag{2.16}$$

Thus, the convolution evaluated at $a \in A$ is defined by summing over all possible ways to obtain the element $a$ as a product of two other elements in the group. In particular, for two random variables $b, c$ taking values in $A$ and distributed according to $P_b, P_c$, their product is distributed according to $P_a * P_b$. Analogously to the standard Fourier transform for functions on the real line, the Fourier transform maps convolutions to products.

**Lemma 9.** *For any two complex valued functions $f, g : A \to \mathbb{C}$ over a finite Abelian group $A$ we have,*

$$\mathscr{F}[f * g] = \mathscr{F}[f] \cdot \mathscr{F}[g] \,, \tag{2.17}$$

*and inversly,*

$$\mathscr{F}[f \cdot g] = \frac{1}{|A|} \mathscr{F}[f] * \mathscr{F}[g] \,. \tag{2.18}$$

This is a standard result, found e.g. in [Ter99, Chapter 10, Theorem 2].

### 2.2.2 Averaging/Subsampling Duality

We now explore the behavior of the Fourier transform for functions defined on subgroups $B$ of a finite Abelian group $A$, following [MK05]. In the next section, we then focus specifically on groups with a product structure.

In this context, the most important corollary to Lemma 7 is the behavior of indicator functions under Fourier transform. For a proof, see e.g. [MK05, Theorem 7].

**Lemma 10.** *Let $A$ be a finite Abelian group and $B \subseteq A$ a subgroup. Then,*

$$\mathscr{F}[U_B] = \Phi_{B^\perp} \,, \qquad\qquad \mathscr{F}[\Phi_B] = |A| U_{B^\perp} \,. \tag{2.19}$$

There is two natural classes of functions in relation to a subgroup $B \subseteq A$. A function $f : A \to \mathbb{C}$ is called *$B$-impulsive* if $f(a) = 0$ for all $a \notin B$. It is called *$B$-periodic* if $f(a) = f(a + b)$ for all $b \in B$. These notions are closely related to the operations of averaging and sampling. The *$B$-averaging* of a function $f : A \to \mathbb{C}$ is $f * U_B$, explicitly

$$(f * U_B)(a) = \frac{1}{|B|} \sum_{b \in B} f(ab) \,. \tag{2.20}$$

The averaging acts similar to a projection in that $f * U_B * U_B = f * U_B$, and the $B$-averaged function is always $B$-periodic. The $B$-*sampling* of a function is $f \cdot \Phi_B$, and this similarly behaves as a projection. The $B$-sampling of a function is always $B$-impulsive.

Furthermore, these two notions are dual to each other under Fourier transform, which directly follows from Lemma 10.

**Lemma 11** (Averaging/Subsampling Duality)**.** *For any complex valued function $f : A \to \mathbb{C}$ on a finite Abelian group $A$ and any subgroup $B \subseteq A$:*

$$\mathscr{F}[f \cdot \Phi_B] = \mathscr{F}[f] * U_{B^\perp} \,, \tag{2.21}$$

*and*

$$\mathscr{F}[f * U_H] = \mathscr{F}[f] \cdot \Phi_{B^\perp} \,. \tag{2.22}$$

### 2.2.3 Local Functions and their Duality Properties

Often, we are concerned with groups that have a direct product structure, i.e. $A = \prod_{i=1}^n A_i$. In this case, $\widehat{A} \cong \prod_{i=1}^n \widehat{A}_i$. Thus, a bicharacter on $A$ can be constructed from the bicharacters $\langle \cdot, \cdot \rangle_{A_i}$ of the groups $A_i$, as

$$\langle a, b \rangle = \prod_{i=1}^{n} \langle a_i, b_i \rangle_{A_i} \,. \tag{2.23}$$

We sometimes refer to this as *product bicharacter*.

Furthermore, we are often concerned with *locally defined functions* on such groups, i.e. functions defined only on some factors of $A = \prod_{i=1}^n A_i$. We now collect some useful notation and basic facts related to direct products, local functions and their duality properties. This material closely follows [MK05], with some sections taken from [Wag+22b].

The *support* of an element $a \in A$ is the set

$$\{i \in [n] \,:\, a_i \neq I\} \,. \tag{2.24}$$

We say that $a$ is supported on a region $R \subseteq [n]$ if $\mathrm{supp}(a) \subseteq R$. The subgroup corresponding to a region $R \subseteq [n]$ is denoted $A_R \coloneqq \prod_{i \in R} A_i$, which we view as embedded in $A$. Similarly, we denote as $a_R$ the projection of $a \in A$ onto that subgroup. For example, the projection of $a = (a_1, a_2, \ldots, a_n)$ onto $A_{\{1,2\}}$ is $a_{\{1,2\}} = (a_1, a_2, 0, \ldots, 0)$. The *complement* of a region $R \subseteq [n]$ is $R^c = [n] \setminus R$. Under the product bicharacter, and using the identification $A \cong \widehat{A}$, we can write $A_R^\perp = A_{R^c}$.

A local function $f_R : A_R \to \mathbb{C}$ can be extended to a function $f$ on $A$ in two different ways: The *impulsive extension* is given by $f(a) = f_R(a)$ if $a \in A_R$ and $f(a) = 0$ otherwise. The *periodic extension* is given by $f(a) = f_R(a_R)$. For example, the periodic extension $f$ of a function $f_{\{1,2\}} : A_{\{1,2\}} \to \mathbb{R}$ to $A$ fulfills $f(a_1, a_2, a_3, \ldots, a_n) \coloneqq f_{\{1,2\}}(a_1, a_2)$ for any $a_3, \ldots, a_n$. As the names suggest, the impulsive extension is $A_R$-impulsive, while the periodic extension is $A_{R^c}$-periodic. As a direct corollary to the averaging/subsampling duality (Lemma 11), these two options are dual to each other under Fourier transform. A proof can be found e.g. in [MK05, Theorem 9].

**Lemma 12.** *Let $R \subseteq [n]$ and $A = \prod_{i=1}^n A_i$ a finite Abelian group, equipped with the product bicharacter. Let $f_R : A_R \to \mathbb{C}$ be a locally defined function with Fourier transform $g = \mathscr{F}[f] : A_R \to \mathbb{C}$ on $A_R$. Let $f_i$ and $f_p$ respectively be the impulsive and periodic extension of $f$ to $A$, and similarly $g_i$ and $g_p$ the impulsive and periodic extension of $g$ to $A$. Then,*

$$\mathscr{F}[f_i] = g_p \qquad\qquad \mathscr{F}[f_p] = |A_{R^c}| g_i \tag{2.25}$$

## 2.3 Important Groups

A very important group in the context of QEC is the *Pauli group*. The Pauli group $\mathcal{P}^n$ on $n$ qubits its the group of Pauli strings generates by the Pauli operators $\{I, X, Y, Z\}$ with phases,

$$\mathcal{P}^n = \{\epsilon \bigotimes_{i=1}^n e_i \,:\, \epsilon \in \{\pm 1, \pm i\},\, e_i \in \{I, X, Y, Z\},\, i \in [n]\}\,. \tag{2.26}$$

Often however, it is possible to ignore the phases, and work in the *effective Pauli group* $\mathsf{P}^n = \mathcal{P}/\{\pm 1, \pm i\}$. Another important group for QEC is the group of of bit-strings $\mathbb{F}_2^n$.

Since all elements of $\mathbb{F}_2^n$ and $\mathsf{P}^n$ have order two, the corresponding bicharacters only take values in $\{+1, -1\}$. For the group $\mathbb{F}_2^n$, we use the bicharacter that is related to the usual scalar product on $\mathbb{F}_2^n$,

$$\langle a, b\rangle = (-1)^{a\cdot b} = \prod_{i=1}^n (-1)^{a_i b_i}\,. \tag{2.27}$$

For the effective Pauli group $\mathsf{P}^n$, the bicharacter encodes commutation relations between the corresponding elements of $\mathcal{P}^n$,

$$\langle a, b\rangle = \begin{cases} +1, & a \text{ and } b \text{ commmute in } \mathcal{P}^n \\ -1, & a \text{ and } b \text{ anti-commmute in } \mathcal{P}^n \end{cases} . \tag{2.28}$$

Thus, although $\mathsf{P}^n$ is an abelian group, we still retain the relevant information about commutation relations in a natural way. This also implies that the annihilator of a subgroup $B \subseteq \mathsf{P}^n$ corresponds to the centralizer of $B$ in $\mathcal{P}^n$, i.e. to the subgroup of all elements of $\mathcal{P}^n$ that commute with all elements of $B$. The bicharacter (2.28) is also known as "scalar commutator" in the literature.

# 3 Quantum Error Correction

In this chapter, we describe the basics of quantum error correction (QEC), starting from the the fundamental idea of redundant encoding in a classical setting. We describe several important classes of quantum codes and summarize them in a unified framework, using the language of group characters. Basic ideas from character and representation theory also allow us to prove fundamental facts about quantum codes in ways which are arguably conceptually simpler than the standard proofs in the literature. Focusing on the role of noise information in QEC, we describe common decoding procedures. We also describe common error models that are used in QEC and their limitations. Namely, we will distinguish between simple *phenomenological noise models* and more fine-grained *circuit-noise models*. The content of this section is mostly inspired by the books by Nielsen and Chuang [NC11], and Lidar and Brun [LB13], and the tutorial by Gottesman [Got09]. An overview over our notation can be found in Section 2.1. Our discussion will be mostly abstract in terms of groups and characters. However, we do assume familiarity with the fundamentals of quantum information, such as state vectors and quantum channels, as found e.g. in [NC11].

## 3.1 Classical Linear Codes

Before we turn to the problems of QEC, let us give a brief description of some classical error correction codes. Later, we see how quantum codes can be constructed in a similar manner, but also what additional challenges arise in the quantum setting.

The simplest example of a classical code is the repetition code. To protect a logical bit from errors, we encode it into $d$ physical bits by simple repetition:

$$0 \mapsto 00\ldots0\,, \qquad\qquad 1 \mapsto 11\ldots1\,. \tag{3.1}$$

When we transmit the physical bits through some noisy channel, some of them could be flipped. Then, we can recover the logical information by a simple majority vote, which will be successful if at most $\lfloor \frac{d-1}{2} \rfloor$ physical bits are erroneous. The parameter $d$ is also called the *distance* of the code.

More generally, a classical linear code encoding $k$ logical bits into $n$ physical bits is a $k$-dimensional subspace $C \subset \mathbb{F}_2^n$. A code can be represented by a generator matrix

$G$, whose rows are $k$ basis vectors of $C$. Often however, a code is instead represented in terms of parity relations, which correspond to the *dual code*. The dual code $C^\perp$ is the annihilator of $C$ under the bicharacter $\langle \cdot, \cdot \rangle$ of $\mathbb{F}_2^n$. This means that $\langle a, c \rangle = +1$ for any $c \in C$ and $a \in C^\perp$, i.e. the dual code describes parity relations that are fulfilled by all codewords. The code $C$ is fully defined by these parity restrictions. For example, for the repetition code the parity of any two adjacent bits is equal, and a basis of the dual code is $(1, 1, 0, 0, \dots), (0, 1, 1, 0, \dots)$ and so on. Generally, one chooses a basis of $C^\perp$ and calls the elements of this basis the *parity-checks*. The parity-checks can be collated into the rows of a matrix $H$, called *parity-check matrix*.

This leads to a convenient description of the error correction process. Let us start with some codeword $c \in C$, and assume is corrupted by an error $e \in \mathbb{F}_2^n$, resulting in the received word $f = c + e$ (where addition is in $\mathbb{F}_2^n$). Then we can apply the parity-check matrix to obtain the *syndrome* $S$,

$$S = \langle H, f \rangle = \langle H, c \rangle \langle H, e \rangle = \langle H, e \rangle \,, \tag{3.2}$$

which just depends on the error since the codewords have by definition a trivial syndrome. Here, we used Notation 3. The task of a *decoder* is to find for each possible syndrome a choice of recovery $r$ that is most likely to match the actual error $e$. We defer discussion on how to construct such a decoder to Section 3.2, where we consider the decoding task in the context of quantum codes.

## 3.2 Stabilizer Codes

The most popular class of quantum codes are *stabilizer codes* [Got97; NC11], which are in many ways similar to classical linear codes. A stabilizer code with $n$ qubits is defined by an Abelian subgroup $\mathscr{S} \subseteq \mathcal{P}^n$ such that $-I \notin \mathscr{S}$, called the *stabilizer group*. The elements of this group are called *stabilizers* of the code. The stabilizers play a similar role for quantum codes as the parity-checks for classical codes. However, all stabilizers must commute, which is a strong restriction that has no analogue for classical linear codes. This makes the construction of quantum codes much more challenging [Bab+15].

Let us briefly discuss the description of a stabilizer code in terms of quantum states. The codespace of a stabilizer code is the simultaneous $+1$-eigenspace of all stabilizers,

$$C = \left\{ |\psi\rangle \in \mathbb{C}^{2^n} : s|\psi\rangle = +1|\psi\rangle \, \forall \, s \in \mathscr{S} \right\}. \tag{3.3}$$

The elements of the codespace are the *codewords* that we use to encode our information. If the codespace $C \subseteq \mathbb{C}^{2^n}$ has dimension $2^k$, we say that the code encodes

$k$ logical qubits into $n$ physical qubits. That is, each possible state $|\phi\rangle \in \mathbb{C}^{2^k}$ of $k$ qubits is encoded as an element of $C$. The error correction capabilities of a code arise from the fact that we only use a small part of the total Hilbert space to encode information, which adds a lot of redundancy. If errors occur, the state will usually be mapped to a state outside of $C$, which can be detected. We then try to recover by mapping back to the correct state in the codespace. The following theorem relates the number of stabilizers to the number of encoded qubits. The standard proof can be found e.g. in [NC11, Proposition 10.5].

**Theorem 13.** *If $r$ is the number of independent generators of $\mathscr{S}$, then $k = n - r$.*

*Proof.* This is essentially a simple fact from representation theory, e.g. found in [FH13, Chapter 2.2]: If $\mathscr{S}$ has $r$ generators, it has $2^r$ elements since all Pauli matrices square to the identity. We can check that the projector onto the codespace $C$ is given by $\Pi_C = \frac{1}{|\mathscr{S}|}\sum_{s\in\mathscr{S}} s$. Since all Pauli matrices except the identity are traceless, we have

$$2^k = \dim(C) = \mathrm{Tr}\left(\Pi_C\right) = \frac{1}{|\mathscr{S}|}\sum_{s\in\mathscr{S}} \mathrm{Tr}\left(s\right) = \frac{\mathrm{Tr}\left(I\right)}{|\mathscr{S}|} = \frac{2^n}{|\mathscr{S}|}\,, \tag{3.4}$$

and thus $2^r = |\mathscr{S}| = \frac{2^n}{2^k}$. □

In QEC, errors are often described as Pauli errors, i.e. elements of $\mathcal{P}^n$, and a common noise model is that of a *Pauli channel* [NC11]. Since the phases are not relevant for QEC, we often describe stabilizers and errors instead as elements of the effective Pauli group $\mathsf{P}^n$ (Section 2.3). A Pauli channel $\Lambda_P$ is quantum channel that acts on the density matrix $\rho$ of a quantum state as

$$\Lambda_P(\rho) = \sum_{e\in\mathsf{P}^n} P(e) e\rho e\,. \tag{3.5}$$

Here, $P : \mathsf{P}^n \to [0,1]$ is a probability distribution over Pauli errors. The action of a Pauli channel corresponds to randomly applying Pauli operators according to $P$. Since the channel is fully described by this distribution, we will sometimes use the term "Pauli channel" also as a shorthand when referring to the distribution $P$. We do however reserve the symbol $\Lambda$ for the actual channels, and denote the distributions as $P$. Pauli channels are a commonly used noise model, but certainly not the most general model possible. Further discussion of this can be found in Section 3.4.

Let us now describe the *error correction process* using stabilizer codes, focusing on the correction of Pauli noise (compare e.g. [NC11, Chapter 10.5.5]). Each round of standard error correction follows a 3-step scheme. First, a chosen set of generators of $\mathscr{S}$ is measured. Since all stabilizers commute, they can be measured simultaneously.

Each measurement results in a $\pm 1$ outcome, and the collection of these outcomes is called the *syndrome*. If no errors are present, the outcomes of all measurements is $+1$, since the codespace is the $+1$-eigenstate of all stabilizers. Now assume a Pauli error $e \in \mathsf{P}^n$ is present on the data qubits, i.e. the encoded state $|\psi\rangle \in C$ is mapped to $e|\psi\rangle$. Then, it can be shown that the outcomes of all stabilizer measurements that anti-commute with this error are flipped. Thus, the outcome of the measurement of $s \in \mathscr{S}$ is exactly given by the bicharacter defined in Eq. (2.28), as $\langle s, e\rangle = \pm 1$. Note that the measurement of the stabilizers does not destroy the logical information, since the codespace is an eigenspace of all stabilizers. Next, based on the syndrome, a correction is applied. The correction $r$ is always chosen in such a way that the state is returned to the codespace, i.e. the syndrome of $r$ has to match the measured syndrome of $e$. Of course, there are still many possible choices of $r$. If the combined error $er$ is equal to a stabilizer, the correction is successful since the logical state is preserved. Otherwise, the correction has failed and the logical state is corrupted. The process of choosing the correction $r$ based on the measured syndrome $S$ is called *decoding*. It should also be noted that often the correction $r$ need not physically be applied to the code, and can instead be tracked in classical memory. This trick should be used as often as possible, since the process of applying the correction can itself introduce more errors in the state if it is implemented using noisy gates.

### Logical Operators

Not every Pauli operator has to map states in $C$ to states outside of $C$. Pauli operators that leave the codespace invariant, mapping codewords to codewords, are called *logical operators*. Such Pauli operators can be seen as transforming the encoded information, since they map valid codewords to different but valid codewords. It is not hard to see from the definition of the codespace that an operator $l \in \mathcal{P}^n$ is a logical operator if and only if it commutes with all elements of the stabilizer group (see e.g. [LB13, Chapter 2.9.4]). Thus, logical operators cannot be detected by the stabilizer measurements. Expressed in the language of group characters, the logical operators $\mathscr{L}$ are exactly the annihilator of $\mathscr{S}$ when viewed as a subgroup of $\mathsf{P}^n$,

$$\mathscr{L} = \mathscr{S}^{\perp}. \tag{3.6}$$

Here, we should note that different errors or logical operators must not necessarily act on the code in a different way. In fact, any two errors that differ only by a stabilizer transform the logical state in the same way, and are hence called *logically equivalent*. This is fundamentally a quantum property. The existence of logically equivalent errors has no analogue in classical coding. This property is also referred to as *degeneracy*. It has far reaching implications for the decoding of quantum codes.

Since logically equivalent errors affect the code in the same way, it makes sense to consider the quotient group $\mathscr{L}/\mathscr{S}$. The group $\mathscr{L}/\mathscr{S}$ describes the action of the different classes of logical operators on the codespace. In the literature, $\mathscr{L}$ and $\mathscr{L}/\mathscr{S}$ are not always strictly distinguished and both are called logical operators.

We can also view the logical operators as the Pauli group of the encoded logical qubits. This is justified by the following theorem.

**Theorem 14.** *For a stabilizer code with stabilizer group $\mathscr{S}$ encoding $k$ logical qubits into $n$ physical qubits, we have*

$$|\mathscr{L}/\mathscr{S}| = 4^k \,. \tag{3.7}$$

*Proof.* Lemma 5 with $A = \mathsf{P}^n$ and $B = \mathscr{L}$ implies

$$|\mathscr{L}| = |\mathsf{P}^n/\mathscr{S}| \,, \tag{3.8}$$

and thus $|\mathscr{L}| = |\mathsf{P}^n/\mathscr{S}| = 4^n/2^r = 2^{2n-r}$. Therefore, we have $|\mathscr{L}/\mathscr{S}| = 2^{2n-r}/2^r = 4^{n-r} = 4^k$, where the last equality follows from Theorem 13. □

Since two Abelian groups are isomorphic if they have the same number of elements of each order, this actually implies $\mathscr{L}/\mathscr{S} \cong \mathsf{P}^k$. More explicitly, we can choose a basis of $\mathscr{L}$ that fulfills the commutation relations of the Pauli group (compare e.g. [Got09][section 3.4]), i.e. $\mathscr{L}$ is generated by the stabilizers and elements $\{\bar{X}_i\}_{i=1,\dots,k}$ , $\{\bar{Z}_i\}_{i=1,\dots,k}$ such that

$$\begin{aligned}
\langle \bar{X}_i, \bar{Z}_j \rangle &= (-1)^{[i=j]} \\
\langle \bar{X}_i, \bar{X}_j \rangle &= +1 \\
\langle \bar{Z}_i, \bar{Z}_j \rangle &= +1 \,.
\end{aligned}$$

These are the logical $X$- and $Z$- operators of the code. A choice of these operators also implicitly defines a basis for the codespace. For example, the logical $|0\rangle$ state is the $+1$-eigenstate of all logical $Z$-operators.

Furthermore, this implies a natural decomposition of any error $e \in \mathsf{P}^n$ (e.g. [LB13, Section 11.2.3]). We can choose for each possible syndrome a *pure error* $t \in \mathsf{P}^n$ matching this syndrome. Once these choices are fixed, each element $e \in \mathsf{P}^n$ uniquely decomposes as

$$e = t(e)s(e)l(e) \tag{3.9}$$

where $t(e)$ is the pure error matching the syndrome of $e$, $l(e)$ is one of the basis logical operators chosen above, and $s(e)$ is a stabilizer.

A rough measure of the error correction capabilities of a code is its *distance*. The distance $d$ is defined as the minimal weight of an undetectable error that transforms the logical state. Therefore,

$$d = \min_{e \in \mathscr{L} \setminus \mathscr{S}} |e| \,. \tag{3.10}$$

A code with distance $d$ can correct all Pauli errors with weight up to $\lfloor \frac{d-1}{2} \rfloor$.

We can also define the *pure distance* $d_p$ of a code, which is the minimal weight of any undetectable error, including logically trivial ones. Explicitly,

$$d_p = \min_{e \in \mathscr{L} \setminus \{I\}} |e| \,. \tag{3.11}$$

This measures up to which weight errors can necessarily be distinguished by their syndrome, independent of how they affect the logical information. It is immediate from the definitions that the pure distance $d_p$ is always smaller or equal to the distance $d$. The pure distance is not an interesting quantity in the context of QEC, but we make use of it in the context of noise estimation in Chapter 6. We also note that for classical codes, since there are no logically trivial errors, there is no distinction between distance and pure distance.

### Decoding

Choosing a good recovery for each syndrome is a difficult problem. Since the syndrome does not contain information about the logical state, some assumptions about the underlying noise process are necessary. The simplest assumption that underlies virtually every error correction strategy, is that high-weight errors are less probable than lower-weight errors. Then, it seems reasonable to return as a correction the lowest-weight error consistent with a syndrome. More generally, if the noise in each round is described by some known probability distribution $P$ over Pauli errors, one should return as a correction $r$ the most likely error consistent with the observed syndrome $S$,

$$r = \mathrm{argmax}_{e \in \mathsf{P}^n} P(e|S) \,, \tag{3.12}$$

where $P(e|S)$ is the conditional probability of the error $e \in \mathsf{P}^n$ given the syndrome $S$. This strategy is called *maximum-likelihood decoding*. It should be noted that maximum-likelihood decoding for general codes is an NP-hard problem, even in the classical case [BMv78].

While maximum-likelihood decoding is optimal for classical codes, it does not take into account the degeneracy of quantum errors. Since logically equivalent errors act on the code in exactly the same way, one should find the most likely class of

logically equivalent errors, instead of the error that is individually most likely. This decoding strategy, called *degenerate maximum-likelihood decoding*, can be written as

$$r = \mathrm{argmax}_{e \in \mathsf{P}^n} \sum_{s \in \mathscr{S}} P(es|S)\,. \tag{3.13}$$

Note that this procedure does not depend on the physical error channel $P$, but only on the *logical channel* $P_L$ which describes the probabilities of different equivalence classes of errors,

$$P_L(e) = \frac{1}{|\mathscr{S}|} \sum_{s \in \mathscr{S}} P(es)\,. \tag{3.14}$$

Thus, in principle only knowledge of the logical channel is necessary for optimal decoding. Since each equivalence class contains exponentially many errors whose probabilities need to be summed, degenerate maximum-likelihood decoding is computationally intractable. Indeed it has been shown that degenerate maximum-likelihood decoding is a #$P$-hard problem [IP15]. Therefore, in practice only approximate decoders are available.

To better understand the concept of degenerate maximum-likelihood decoding, it is useful to view it as purely operating on the logical information. Consider the decomposition given in Eq. (3.9). After some error $e \in \mathsf{P}^n$ occurred, we can always return to the codespace by applying the pure error $t(e)$ matching the syndrome. This will usually not map to the correct codeword, and thus the logical state is transformed by some logical operator. The task of degenerate maximum-likelihood decoding is then to predict the most likely logical transformation of the data. This corresponds to finding the most likely equivalence class of logical operators given the syndrome.

### Examples of Decoders

In the following, we describe some common decoding approaches. Of course, this list is not exhaustive. Besides giving a general overview of common decoding techniques, this list is meant to illustrate how knowledge of the physical noise is useful to improve error correction. Thus, we highlight how Pauli error rates can be incorporated into the decoding algorithms in order to improve their performance. The main focus is on the case of independent single-qubit Pauli errors, with possibly different rates for each qubit, since this is the setting for which most decoders are designed.

One popular class of decoders is based on *minimum-weight matchings* [Den+02; Hig21]. One tries to find the lowest-weight error consistent with the measured syndrome, which is an instance of maximum-likelihood decoding. While this does

not take into account degeneracy, it can still lead to very good logical error rates. However, for general stabilizer codes no efficient algorithm is available to construct the minimum-weight error. (In fact this is a hard problem already for general classical codes [BMv78].) The main exception is the surface code [Fow+12; Den+02], where each single $X$- or $Z$-error only affects one or two syndrome bits. One can then decode $X$- and $Z$- errors separately. For each error type, the problem can be translated into a *decoding graph* where vertices correspond to the $-1$-outcomes in the measured syndrome, and edges correspond to errors that affect the adjacent syndrome bits. The task is then to construct a minimum-weight matching in this graph, which can be done in polynomial time using the "Blossom" algorithm [Edm65]. The fact that errors affect at most two syndrome bits is important, since otherwise the problem would correspond to a hypergraph where no efficient matching algorithms are available. Error rates information can be incorporated into this decoder by weighting the edges based on the error probabilities of the corresponding qubits. The algorithm can also be adapted to the fault-tolerant setting where the components of the syndrome measurement circuits are themselves noisy, see e.g. [Hig21; HB21].

An adjacent method, first proposed in the context of surface codes as well, is *union-find decoding* [DN21]. This decoding algorithm is mainly designed to match the clock speed of first generation quantum processors, trading of some accuracy compared to minimum-weight matching for improved speed. The core idea is that the correction of errors is much easier if we know which qubits were affected by noise, while knowing that all other qubits are error free. This corresponds to the correction of so called *erasure errors*, where some qubits are lost and their state needs to be recovered based on the remaining qubits. For standard stochastic Pauli noise, however, such a support of the error is not known. The approach of the of the union-find algorithm is to first find a set of qubits which is guaranteed to cover the actual error. This is accomplished by growing clusters of erasures around each flipped syndrome bit, merging colliding clusters, until all clusters contain an even number of flipped syndrome bits. As long as the actual error is smaller than half the code distance, these clusters fully cover the actual error. Furthermore, the clusters are then themselves smaller than the code distance, and cannot support a logical operator. In this case, we can use an efficient peeling decoder for erasure errors [DZ20], and the correction will match the actual error up to stabilizers. While the original version of the decoder does not use any error rates information, it has been shown that the performance can be improved by weighting the growth in the directions of more likely errors, as shown in [HNB20]. In this work, the authors also show how union-find can be adapted to the effect of noisy measurement circuit.

A very general way of decoding stabilizer codes is based on the *belief-propagation* (BP) algorithm (see e.g. [Bab+15] for a review, and compare Section 4.3). For

this, first the so-called *Tanner graph* of the code is constructed. It consists of two types of nodes: factor nodes, which correspond to stabilizer generators, and qubit nodes. A qubit node is adjacent to a factor node if the corresponding stabilizer acts non-trivially on this qubit. The goal is to compute for each qubit $i$ the marginal probability $P_i(e|S)$ that it is affected by an error $e \in \mathsf{P}^1$ given the observed syndrome $S$. To accomplish this, certain messages are iteratively passed between factor and qubit nodes. In the end, the decoding is done by predicting for each qubit $i$ the error $\hat{e} = \mathrm{argmax}_{e \in \mathsf{P}^1} P_i(e|S)$, and checking of the total error obtained by this procedure is consistent with the observed syndrome. If it is not, either more iterations can be done or the decoding can be declared unsuccessful. Error rates information is required, since the messages from qubits to factors should be initialized based on the Pauli error rates of the qubits. While this algorithm is very successful for classical codes, it has proven difficult to adapt it to the quantum setting. The Tanner graphs of quantum codes naturally exhibit short loops, which are known to degrade the performance of this algorithm [PC08]. Still, it is possibly to decode concatenated codes exactly using a variant of this algorithm [Pou06], and there is a growing literature of ways to improve the performance for general stabilizer codes [Bab+15; LP19; Rof+20]. Belief-propagation can also be used as an initial step to reduce the number of errors, and then the remaining errors are handled by a fast decoder such as union-find.

Finally, there is the class of *tensor network decoders* [Chu21; DP18]. These map the decoding problem onto a tensor network contraction. Knowledge of the physical noise is naturally incorporated into these decoders, since very general noise models can be expressed in terms of tensor networks. This includes Pauli noise with correlated errors, and even non-Pauli noise models. Furthermore, in contrast to minimum-weight matching or union-find, these algorithms approximate *degenerate* maximum-likelihood decoding. However, tensor network contraction is computationally expensive. Thus, tensor network decoders might be more suited to probing the ultimate performance limits of a code, rather than practical error correction.

## 3.3 Subsystem Codes

A useful generalization of stabilizer codes is given by subsystem codes [Pou05; LB13]. These can be viewed as a stabilizer code where some of the logical qubits are not used to store information. These unused logical qubits are called *gauge qubits*. Since they do not store information, the corresponding logical operators can be

considered logically trivial, similar to stabilizers. Unlike stabilizers however, these logical operators do not necessarily commute.

Since the logical operators of gauge qubits can be measured without destroying the encoded information, they can sometimes be used to simplify stabilizer measurements. This is the case if a stabilizer can be written as the product of lower-weight logical operators and stabilizers. For example, in [Bra+13] the authors introduce a subsystem surface code which only requires three-qubit measurements, instead of the four-qubit measurements of the usual surface code. Furthermore, some properties of quantum codes are best expressed in the framework of subsystem codes. For example, fault-tolerant transformations between different codes (code switching) and related topological techniques such as lattice surgery can be conveniently expressed in this framework [Vui+19]. The effect of circuit-noise (see Section 3.4) can also be expressed in the language of subsystem codes [Pry20; CF21; Bac+17].

A good overview over the basic framework of subsystem codes is given in [Vui+19]. Formally, a subsystem code is defined by a *gauge group* $\mathscr{G} \subseteq \mathsf{P}^n$. Unlike the stabilizer group, we do not require $\mathscr{G} \subseteq \mathscr{G}^\perp$, i.e. gauge operators do not have to commute in $\mathcal{P}^n$. Instead, the associated stabilizer group is $\mathscr{S} = \mathscr{G}^\perp \cap \mathscr{G}$, and these are the operators that we ultimately measure. Any operator mapping the codespace to itself is called a *dressed logical operator*. Consequently, the dressed logical operators are $\mathscr{L}_d = \mathscr{S}^\perp$. A dressed logical operator acts logically non-trivial if it is not an element of the gauge group, and thus it makes sense to consider the quotient group $\mathscr{L}_d/\mathscr{G}$. We can always modify dressed logical operators such that they act trivially on the gauge qubits, resulting in *bare logical operators*. The set of all bare logical operators is $\mathscr{L}_b = \mathscr{G}^\perp$. Finally, analogous to stabilizer codes, the distance of a subsystem code is the minimal weight of a non-trivial logical operator. Thus, $d = \min_{l \in \mathscr{L}_d \setminus \mathscr{G}} |l|$.

## 3.4 Error Models

So far, we have described quantum codes focusing on Pauli errors on the data qubits. In particular, we have considered a discrete set of possible errors, corresponding to Pauli operators. Of course, in principle, one would not expect that quantum noise is discrete. Instead it should be described by general CPTP maps acting on the state space. Therefore, the question arises how far the focus on Pauli noise can be justified.

One justification is given by the *discretization of errors*. It can be shown that if a code with a given decoder corrects a set of errors $\{E\}$, corresponding to Kraus-operators of a CPTP map, then it will also correct any errors that are linear combinations of the errors $\{E\}$ [NC11, Theorem 10.2]. Since the Pauli operators are a

basis for the space of linear operators, in a sense it suffices to correct Pauli errors. For example, a stabilizer code with distance $d$ will correct all Pauli errors of weight at most $\lfloor \frac{d-1}{2} \rfloor$, and thus all errors affecting at most $\lfloor \frac{d-1}{2} \rfloor$ qubits.

However, this result does not tell us about the logical error rate of the code for a given physical noise channel, which depends on the probability of encountering uncorrectable errors in this channel. In principle, it is not sufficient to benchmark codes on Pauli channels, and furthermore decoders designed for Pauli channels are generally not optimal for other quantum channels. For example, an arbitrary quantum channel can be mapped to a Pauli channel by removing the off-diagonal elements in the Pauli basis, which is called Pauli twirl approximation. However, the logical error rate under the twirled channel can underestimate the logical error rate under the original channel [Mag+13], and furthermore a decoder designed for the twirled channel might be sub-optimal for the actual channel. Still, it is very common to benchmark codes on Pauli channels, since they can be efficiently simulated [Mar+20]. Furthermore, most decoders are designed with Pauli errors in mind. In some settings, it has been demonstrated that considering Pauli errors will not underestimate the logical error rate of the code [KG15]. There are also *honest* mappings from general noise to Pauli noise that can only overestimate, but never underestimate, the logical error rates [Mag+13].

Especially in the context of this thesis, which is not concerned with decoding for a given channel but rather with estimation of the channel itself, another important justification for considering Pauli channels arises from a method called *randomized compiling* [WE16]. The core idea is to insert random Pauli gates into the quantum circuit executed by the device. This has the effect of twirling the noise channels, thus mapping it to a Pauli channel. Pauli channels can indeed be a realistic model for the actual noise on a quantum device, provided randomized compiling is used.

Let us now describe two important classes of Pauli noise models that are used in QEC. The first are *phenomenological* noise models, which are essentially the models we have considered so far. More precisely, we consider the error correction process in discrete rounds. Before each round, a Pauli error $e \in \mathsf{P}^n$ occurs according to some distribution $P : \mathsf{P}^n \to [0, 1]$. Then, a measurement of the stabilizer generators is performed. In the simplest setting, the measurements are perfect and each return the correct syndrome bit, but we can also consider noisy measurements that return the wrong outcome with some probability (or even correlated measurement errors). This is described in more detail in Section 3.5. After each round of measurements, a correction is performed, and then the next round begins.

The above phenomenological noise model is sometimes used to benchmark codes, but it does not fully take into account all effects of the noisy hardware. In particular, the circuits used to implement the stabilizer measurements are themselves noisy and

can introduce new errors in the middle of the measurement process. Furthermore, errors propagate through these circuits and thus single errors can spread onto many qubits. This motivates the more realistic *circuit-noise model*, where each gate used in the measurement process introduces new Pauli errors, and these errors are propagated through the circuits (e.g. [DRS22]). In such a setting, one generally needs multiple rounds of syndrome measurements before a good correction can be identified. Designing good decoders for these settings is also much more challenging than for the simple phenomenological noise models, and not all decoders can be adapted in an obvious way. However, a common approach is to consider a decoding graph, or more generally a hypergraph, similar to that used by the minimum-weight decoder, see e.g. [Hig21; Spi+18]. Single circuit faults are grouped depending on their syndrome, and faults with different syndromes are considered independent. This essentially results in a phenomenological noise model on the decoding graph. While this is a very rough approximation of actual circuit-noise, this technique seems to perform reasonably well in practice [WFH11; HB21].

## 3.5 Quantum Data-Syndrome Codes

Often, the syndrome measurements used in QEC are noisy themselves and can return the wrong outcome. A simple way to deal with such errors is to repeat the syndrome measurements multiple times before decoding. Then one can try to correct measurement errors before decoding the data errors, or more generally decode the data errors based on all the repeated syndrome measurements simultaneously.

A generalization of this concept is given by quantum data-syndrome codes [Fuj14b; ALB20]. Instead of simply repeating the syndrome measurements, we encode the syndrome itself in a classical error correction code. This corresponds to redundant stabilizer measurements based on combinations of the stabilizer generators, where the combinations are specified by the classical code. We mainly use quantum data-syndrome codes as a unified language to treat both data and measurement errors at once, and are thus not too concerned with the details of their construction. As presented here, the framework applies only to phenomenological noise models, although there has been some work on extending data-syndrome codes to deal with circuit-noise [DRS22].

To define a quantum data-syndrome code, we first pick an underlying stabilizer code with stabilizer group $\mathscr{S} \subseteq \mathcal{P}^n$. Then, we choose a set of possibly redundant stabilizer generators $g_1, \dots, g_m$ that is measured in each round. An error is described by a combination of a data error $e_d \in \mathsf{P}^n$ and a measurement error $e_m \in \mathbb{F}_2^m$, as $e = (e_d, e_m)$. Here, $e_m[i] = 1$ indicates that the outcome of the $i$-th generator

measurement is flipped, and $e_m[i] = 0$ indicates that it is correct. Thus, errors are described as elements of the group $\mathsf{G}^{n,m} = \mathsf{P}^n \times \mathbb{F}_2^m$. We want to describe the corresponding measurements of the code also as elements of $\mathsf{G}^{n,m}$, such that the measurement outcomes are given by the product bicharacter on $\mathsf{G}^{n,m}$:

$$\langle (a_d, a_m), (e_d, e_m) \rangle = \langle a_d, e_d \rangle \langle a_m, e_m \rangle . \tag{3.15}$$

This allows us to treat data-syndrome codes in the same language as regular stabilizer codes. This is accomplished by the extended parity-check matrix

$$H = \begin{bmatrix} G & I_m \end{bmatrix} , \tag{3.16}$$

where the rows of $G$ are the stabilizers $g_1, \ldots, g_m$. More precisely, each generator $g_i$ is extended to an element $f_i = (g_i, \hat{i}) \in \mathsf{P}^n \times \mathbb{F}_2^m$, where $\hat{i}$ is the $i$-th standard basis vector. It is easy to see that the outcome of $i$-th generator measurement if an error $e \in \mathsf{G}^{n,m}$ occurred is then given by $\langle f_i, e \rangle$. By taking products of the outcomes of the generator measurements, we obtain the group $\mathscr{M} = \langle f_1, \ldots, f_m \rangle \subseteq \mathsf{G}^{n,m}$ of possible measurements. Note that since only the generators are physically measured, the measurements of composed stabilizers are affected by multiple measurement errors, reflected in multiple non-zero entries in the $\mathbb{F}_2$ part. Since measurement outcomes are described by a bicharacter, the set of undetectable errors is given by the annihilator $\mathscr{U} = \mathscr{M}^\perp$. These do not necessarily map the codespace of the underlying codes into itself, since measurement errors can be present.

Logical equivalence of errors is still described by the stabilizer group $\mathscr{S}$ of the underlying code. Thus, similar to subsystem codes, the available measurements and the operators describing logical equivalence do not coincide. The distance of a quantum data-syndrome code is defined as usual as the smallest weight of a logically non-trivial undetectable error, i.e.

$$d = \min_{l \in \mathscr{U} \setminus \mathscr{S}} |l| . \tag{3.17}$$

## 3.6 A Unified Description of Codes

So far, we have encountered four different kinds of codes: classical codes, stabilizer codes, subsystem codes and quantum data-syndrome codes. For our purposes, all of these can be captured by the same mathematical structure, resulting in a unified description. This description is used in Chapter 6 to prove universal results, and might also capture settings other than QEC.

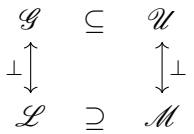



**Figure 3.1.:** An overview over the abstract setting, described by four groups. The main ingredients are a group of measurements $\mathscr{M}$ and a gauge group $\mathscr{G}$. The gauge group describes which errors are considered logically trivial. The annihilator of $\mathscr{M}$ is the group of undetectable errors $\mathscr{U} = \mathscr{M}^\perp$, and the annihilator of $\mathscr{G}$ is the group of logical operators $\mathscr{L} = \mathscr{G}^\perp$. The groups fulfill the dual inclusions $\mathscr{G} \subseteq \mathscr{U}$ and $\mathscr{M} \subseteq \mathscr{L}$. Figure from [Wag+22b].

On a basic level, we always consider a finite Abelian group $A = \prod_{i=1}^n A_i$ with a product structure, where each $A_i$ is either $\mathsf{P}^1$ or $\mathbb{F}_2$. This group comes equipped with the product bicharacter, which only takes the real values $\pm 1$ because of the form of the $A_i$. We consider a phenomenological noise model where errors are described as elements of $A$, and occur before each round of measurements. A code is described by the following four subgroups of $A$.

First is the group of *measurements* $\mathscr{M}$, which describes what kind of measurement outcomes we obtain in each round. Usually, only a set of generators is physically measured, and all other outcomes are obtained as products of generator outcomes, justifying the group structure. Measurements outcomes are given by the product bicharacter of $A$. Explicitly, the outcome of the measurement of $a \in \mathscr{M}$ when an error $e \in A$ occurred is given by $\langle a, e \rangle$. Some care needs to be taken in a standard error correction setting since in principle one measures the accumulated error and not the new error in each round, but this is solved by considering differences of syndromes, see e.g. [Wag+22b] and Section 6.2.

Second is the *gauge group* $\mathscr{G}$, which describes logical equivalence. All elements of the gauge group are considered logically trivial, and errors that differ by elements of the gauge group are logically equivalent.

Finally, we have two additional groups which are dual to the first two. The group of *undetectable errors* is $\mathscr{U} \coloneqq \mathscr{M}^\perp$. These are exactly the errors resulting in a trivial measurement outcome. On the other hand, we have the *logical operators* $\mathscr{L} \coloneqq \mathscr{G}^\perp$. These operators transform the logical information of the code and map codewords to codewords.

We still have to describe some relation between these different groups. To establish such a connection, we require that the inclusion $\mathscr{G} \subseteq \mathscr{U}$ is fulfilled. This means that logically trivial operations must also be undetectable. By taking annihilators, we see that this is equivalent to the dual inclusion $\mathscr{M} \subseteq \mathscr{L}$. An overview over this abstract description is given in Figure 3.1 (taken from [Wag+22b]).

Fundamental definitions from QEC directly transfer to the abstract setting. For example, the distance $d$ of a code is the smallest weight of a logically non-trivial undetectable error,

$$d = \min_{l \in \mathscr{U} \setminus \mathscr{G}} |l| \,, \tag{3.18}$$

while the pure distance is the smallest weight of any undetectable error,

$$d_p = \min_{l \in \mathscr{U} \setminus \{I\}} |l| \,. \tag{3.19}$$

Furthermore, if the physical errors are described by some distribution $P : A \to [0, 1]$, the logical channel $P_L$ can be defined by averaging over the gauge group,

$$P_L(e) = \frac{1}{|\mathscr{G}|} \sum_{b \in \mathscr{G}} P(eb) \,. \tag{3.20}$$

In the previous sections, the four groups describing a code have appeared under slightly different names, following the conventions used in the literature for these specific classes of codes. We now give an overview over the relations to the abstract setting.

For stabilizer codes (Section 3.2), the measurements and the gauge group coincide and are both equal to the stabilizer group $\mathscr{S}$ of the code. Consequently, the logical operators and undetectable errors also coincide. In the language of stabilizer codes, the abstract inclusion $\mathscr{G} \subseteq \mathscr{U}$ takes the form $\mathscr{S} \subseteq \mathscr{L}$, which is fulfilled since the stabilizers commute.

For subsystem codes (Section 3.3), the group of measurements is the stabilizer group $\mathscr{S}$ and the abstract gauge group is the gauge group of the subsystem code. The abstract undetectable errors are then the dressed logical operators, while the abstract logical operators are the bare logical operators. Since for subsystem codes $\mathscr{S} = \mathscr{G}^\perp \cap \mathscr{G}$, we have in particular $\mathscr{S} \subseteq \mathscr{G}^\perp$, which we recognize as the abstract inclusion $\mathscr{M} \subseteq \mathscr{L}$.

For quantum data-syndrome codes , the abstract group of measurements is defined as explained in Section 3.5, and the gauge group is the stabilizer group $\mathscr{S}$ of the underlying code. The undetectable errors and logical operators are then also as defined in Section 3.5. In particular, since the stabilizers only act on the data qubits, we have that $\mathscr{L} = \mathscr{S}^\perp = \mathscr{L}_d \times \mathbb{F}_2^m$, where $\mathscr{L}_d$ is some group of logical operators acting on the data qubits that in particular contains the stabilizers itself. Then, by the definition of $\mathscr{M}$, we get $\mathscr{M} \subseteq \mathscr{L}$, fulfilling the abstract inclusion imposed above.

# 4 Probabilistic Graphical Models

Many of the results presented in this thesis can be interpreted in the language of graphical models. Probabilistic graphical models (PGMs) are, at their core, a compact way to describe probability distributions as graphs. By using a graphical description, the independence structure of a class of distributions can be expressed in an intuitive and convenient way. Furthermore, it turns out that describing the independence structure of a distribution is, under mild assumptions, equivalent to describing a factorization of the distribution. This results in two equivalent ways of viewing PGMs. PGMs lend themselves naturally to the context of QEC, since they are concerned with the locality of probability distributions, which is also important for QEC.

PGMs have appeared in many different fields under different names. In classical error correction, PGMs can be used as a graphical description of parity-check codes, called *Tanner graph*. One of the most successful decoding algorithms is based on inference on this graph via the BP algorithm [RU08]. In statistical physics, coupled physical systems such as the Ising model are studied., which can often also be interpreted as PGMs [Pel05]. More recently, PGMs have become a widely used tool in machine learning. They appear for example in diagnostic systems or image processing problems such as segmentation or denoising, or in language problems in the form of hidden Markov models. An overview over some applications is given in [KF09] and [WJ08].

There are three main tasks associated with PGMs [KF09]. The first is *representation*: finding compact ways to represent probability distributions and their independence structure. The second is *inference*: efficiently computing marginal and conditional distributions of some variables in a model with known parameters. Finally, there is *learning*: finding the structure or parameters of the model from data.

In the context of (quantum) error correction, the first two tasks are commonly considered. Representation is usually done in the form of factor graphs (called Tanner graphs in this context), which provide a convenient way to describe error correction codes. Inference is used for decoding, leading to one of the most successful classical decoding algorithms. The learning task, however, is not part of standard error correction. Indeed, many contributions of this thesis can be interpreted as

transferring ideas from the learning of PGMs to quantum error correction, although not all of our results have originally been derived from this perspective. In Chapter 5, we see how BP and expectation-maximization (EM) can be applied to learning error channels in quantum error correction. Our results in Chapter 6 also have some connection to the field of graphical models, specifically to factor graph learning.

In this chapter, we give an overview over the most important tools from the field of graphical models for our purposes. The content of this chapter is mainly inspired by [KF09], [Bis06], [RU08], [AKN06] and [MK05].

## 4.1 Factor Graphs

We have already encountered local functions in Section 2.2.3. A factor graph is a graphical description of the factorization of a function over many variables into local functions. It is represented by a bipartite graph, with two sets of nodes: A set $V$ of variable nodes and a set $F$ of factor nodes. We identify each variable node $v \in V$ with a variable taking values in a set $A_v$. We denote as $A = \prod_v A_v$ the set of configurations of all variables, where $\prod$ denotes the Cartesian product of sets. Often, $A_v$ and $A$ are groups, but this is not necessary. Each factor node $f \in F$ is identified with a real valued function $f$ of the neighboring variables,

$$f : \prod_{v \in \mathrm{supp}(f)} A_v \to \mathbb{R} \,. \tag{4.1}$$

Here, $\mathrm{supp}(f)$ denotes the set of variable nodes neighboring $f$, which is also called the *scope* or *support* of the factor $f$. The function $f$ can be extended periodically to a function

$$f : A \to \mathbb{R} \tag{4.2}$$

of all variables, simply by ignoring all variables not in the scope of $f$. With this convention, the factor graph represents the function

$$\begin{aligned} g &: A \to \mathbb{R} \\ g &= \prod_{f \in F} f \,. \end{aligned} \tag{4.3}$$

For example, the factor graph in Figure 4.1 represents the function

$$g(v_1, v_2, v_3) = f_1(v_1, v_2, v_3) f_2(v_1, v_2) f_3(v_3) \,. \tag{4.4}$$

Usually, the function represented by a factor graph is interpreted as an unnormalized probability distribution, in which case the variable nodes correspond to random

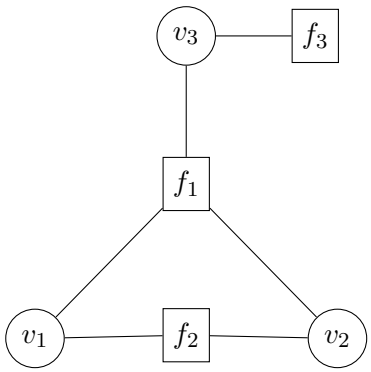



**Figure 4.1.:** An example of a factor graph.

variables. The normalization factor generally depends on the parameters and is called *partition function*. Computing the partition function is usually intractable, which is one of the major difficulties in treating PGMs. If all factors are strictly positive, i.e. each $f \in F$ only takes positive value, the distribution is often called a *Gibbs distribution*. In this case, the factor graph encodes certain independence properties of the variables. For example, the *Markov blanket* of a variable $v \in V$ is the set of variables connected to $v$ by a factor, e.g.

$$MB(v) = \{w \in V \,:\, \exists f \in F \text{ such that } w \in \operatorname{supp}(f) \text{ and } v \in \operatorname{supp}(f)\}\,. \tag{4.5}$$

Each variable is conditionally independent of all other variables given its Markov blanket. Even stronger statements are possible, but we do not make much use of these independence properties here.

## 4.2 The Canonical Factorization

The mapping from factor graphs to probability distributions is not one to one. In general, there are many different factor graphs that represent the same distribution. It can even be the case that different parameter choices on the same graph structure result in the same distribution. If different factors overlap, some of their "weight" can be traded off without changing the overall distribution. In particular, we can arbitrarily introduce new factors which are subsets of existing factors, and get a continuum of parametrization.

For factor graphs representing a Gibbs distribution, there is a special choice of representation called the *canonical factorization*. This was originally introduced as part of the famous *Hammersley-Clifford* theorem for Markov-random fields [HC71; Gri73], but an extension to factor graphs was constructed by Abbeel et al. [AKN06]. It is in a sense the finest possible factorization. Intuitively, one should not use

large factors to describe properties that can already be captured by smaller sub-factors. Thus, for a given factor graph, the canonical factorization is obtained by introducing all possible subsets of the existing factors as new factors. The functions represented by these *canonical factors* are then designed in such a way that they only capture properties that cannot be described by smaller factors. This is achieved via the *inclusion-exclusion* principle of combinatorics. Details of this construction are described in [AKN06]. We will introduce a closely related technique in Section 6.2.3, which we call *canonical moments*.

## 4.3 Belief-Propagation

An important task in the context of PGMs is *marginalization*. For models with many variables, computing the marginal distribution over some subset of variables can be very difficult. Naively, computing the marginal distribution of a variable $v$ requires summing over all possible configurations of all other variables, resulting in an exponentially large sum. For distributions represented by factor graphs however, much more efficient methods are available. One such method is called *belief-propagation* or *sum-product algorithm*. It exploits the local structure of the distribution, such that we only need to perform local sums over all configurations of variables in a factor. Thus, the computational effort scales exponentially only in the size of the largest factor, but is linear in the number of factors.

For a detailed description of BP see e.g. [Bis06] or [KF09]. We only give an informal overview here. BP iteratively passes messages between factor and variable nodes. Intuitively, the messages from a factor $f$ to a variable $v$ present the belief about the value of this variable, based on the factor $f$ and its neighborhood. It is essentially obtained by a "local marginalization", i.e. summing over all configurations of all neighbors of $f$ except $v$. In return, the messages from a variable $v$ to a factor $f$ pass on the current belief about the distribution of $v$, based on all neighboring factors except $f$. Messages are passed iteratively between factors and variables until they converge. After convergence, the marginal distribution of a variable is computed as the product of the messages from all neighboring factors.

If the factor graph is a tree, it can be shown that BP converges to the true marginals in a finite number of steps. Otherwise, BP is only an approximate algorithm. It generally performs well if the corresponding Tanner Graph does not have short loops, but can perform poorly if such short loops are present [KF09]. As explained in Section 3.2, BP can be used to construct efficient decoders for classical error correction codes. However, the presence of short loops has proven to be a problem

when applying BP to quantum codes [Bab+15]. We use BP in Chapter 5 as both a decoder and a subroutine of our estimator for concatenated quantum codes.

## 4.4 Expectation-Maximization

We now turn to the task of learning the parameters of a PGM from a set of observations of the variables. In Chapter 5, we use the ideas described in this section to develop an estimation algorithm for Pauli channels from syndrome data.

We are interested in learning the parameters, i.e. the function $f$ associated with each factor, for a model with known graph structure. Furthermore, we are interested in algorithms that can learn the parameters in the presence of hidden variables, since in the context of QEC, we can only observe the syndrome information, while the states of the qubits are hidden.

For general factor graphs, even learning with fully observed data is a difficult problem. For now, we consider the problem for a simpler class of models, which are called *Bayesian network*. A Bayesian network describes the factorization of a probability distribution into a series of conditional probabilities.

We can view a Bayesian network as a factor graph where the links between variables and factors are directed, and there are no directed cycles. Each variable may have multiple child factors, but only one parent factor. Thus, each variable has a unique set of parent variables, to which it is connected via one parent factor. The parent factor encodes the (normalized) conditional distribution of the child variable given the parent variables. Since all probabilities are normalized, no problem arises from computing the partition function. The collection of all these conditional probabilities for each factor are the parameters of the network, which we denote as $\theta$. An example of a Bayesian network is given in Figure 4.2. It encodes the probability distribution

$$P(v_1, \cdots, v_6) = f_1(v_1) f_2(v_2) f_3(v_3|v_1, v_2) f_4(v_4|v_1, v_2) f_5(v_5) f_6(v_6|v_3, v_4, v_5)\,, \quad (4.6)$$

where each factor $f_i$ is a normalized (conditional) probability distribution. Note that usually, Bayesian networks are defined without explicit factor nodes simply by connections between the variables, since the scope of each factor is uniquely specified by the parents of each variable. Here, we made the factor nodes explicit to highlight the connection to factor graphs.

Let us consider a Bayesian network where the variables are partitioned into a set $H$ of hidden variables and a set $O$ of observed variables. We want to learn the parameters of the network from a data set $D$, where each data point is a sample only of the observed variables. In this setting, we can learn the parameters using

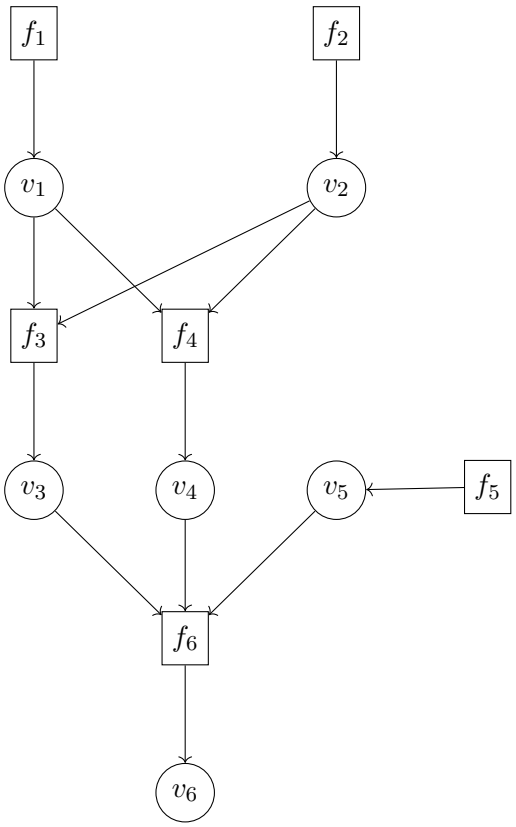



**Figure 4.2.:** An example of a Bayesian newtork.

an algorithm known as *expectation-maximization (EM)* [DLR77]. Let us provide some intuition, following the discussion in [KF09]. First, note that if we could observe values for all variables, the learning task would be straightforward. We could estimate the conditional probability distribution for each variable given its parents simply by counting how often the corresponding configuration of values appears in the data set. This leads to the idea of filling in some values for the hidden variables in each data point, but it is not clear what values to choose. If on the other hand we already knew the parameters of the network, we could compute the conditional probability of each hidden variable given the observed variables in the data and use this information to fill in the values.

The EM algorithm combines these two observations into an iterative method. We start from some guess $\theta^{(0)}$ of the parameters of the network. For each variable in the network and each data point in our data set, we compute the conditional probability distribution of this variable given the observed data point and the parameters $\theta^{(0)}$. For many models, this can be done efficiently for example using the BP algorithm. We then obtain a new estimate for the parameters by computing conditional probabilities over the data set. More precisely, the estimate in the $(i+1)$-th iteration is obtained

by first computing the *expected sufficient statistics* $M^{(i+1)}$ for each possible value $x$ of each variable $v \in V$ and each possible value $y$ of its parents $Pa(v)$ using the parameters $\theta^{(i)}$,

$$M^{(i+1)}[v, x, Pa(v), y] = \sum_{d \in D} P(v = x, Pa(v) = y | d, \theta^{(i)}) \,, \tag{4.7}$$

The name expected sufficient statistics derives from the fact that this can be seen as an expectation over all possible assignments to the hidden variables, see [KF09] for more information. To obtain the next estimate $\theta^{(i+1)}$, we normalize the expected sufficient statistics to obtain conditional probability distributions.

The above procedure is iterated for a fixed number of steps or until convergence. It can be shown that each iteration is guaranteed to increase the likelihood of the data under the model [KF09; NH98]. Thus, EM is a form of approximate maximum-likelihood estimation.

Note that the EM algorithm uses the conditional probabilities of the hidden variables given the data for the new estimate. Alternatively, one could make a hard decision by computing for each data point the most probable configuration of the hidden variables given the observed variables, and estimating the new parameters only based on this most probable configuration. This variant of EM is called *hard-assignment expectation-maximization* (HEM). By making a "hard" decision, only taking into account one configuration of the hidden variables and ignoring all other possibilities, it usually updates the parameters faster, but also throws away useful information. It is thus more prone to getting stuck on bad estimates.

# 5 Noise Estimation for Concatenated Codes

In this chapter, we develop an estimation algorithm for Pauli channels from syndrome data, based on combining belief-propagation with expectation-maximization. The algorithm is specific to the class of concatenated quantum codes, but has the advantage that it naturally allows for joint decoding and estimation. This summarizes our published work [Wag+21], which is listed in Appendix B.

First, we discuss some necessary prerequisites: The structure of concatenated codes and their maximum-likelihood decoding. Then, we describe how Pauli channels can be learned from the syndrome measurements of a concatenated code, using an approach based on PGMs. We demonstrate that this approach outperforms previous proposals, since it incorporates soft information instead of using hard decisions.

In our work [Wag+21], we also prove identifiability results for a class of codes called perfect codes. We do not discuss these here, since our subsequent works [Wag+22a] and [Wag+22b] contain strictly stronger results. These are explained in Chapter 6. Some of our technical lemmas about perfect codes might still be of independent interest.

## 5.1 Concatenated Codes and their Decoding

A concatenated code is created by encoding each qubit of one error correction code in another error correction code. In the simplest case, each of the $n$ qubits of a code is encoded again in the same code. This concatenation procedure can be repeated $r$ times, resulting in a code with $n^r$ physical qubits. We can think of the physical qubits as being divided into blocks of $n$ qubits, each corresponding to one instance of the base code. Each of these blocks encodes one logical qubit. These logical qubits can be viewed as the "physical" qubits of the next concatenation layer, and each block of $n$ logical qubits again encodes one logical qubit on the next concatenation layer. Thus, the concatenated code defines a hierarchy of logical qubits, which can be illustrated as a tree structure. An example for a base code with $5$ qubits is given in Figure 5.1. The $5$ blocks of physical qubits at the bottom encode one logical qubit

$L_1^1, \ldots, L_1^5$ each, and these $5$ logical qubits ultimately encode one logical qubit $L_2^1$ on the top layer.

This picture also suggest a natural view on the decoding of concatenated codes. An error $e \in \mathsf{P}^{n^r}$ on the physical qubits determines the logical error for each of the logical qubits in the next layer (relative to a fixed set of pure errors, see the decomposition (3.9)). These in turn determine logical errors for the next layer, and so on. The task of degenerate maximum-likelihood decoding (Section 3.2) is thus to find the most likely logical error at the root. It seems natural to implement this decoding layer-wise. First, each block of physical qubits is decoded to determining a guess of the logical error for this block. This results is an error for each of the logical qubits in the next layer, and these can be decoded in the same way. By working up the tree one obtains the logical error at the root. However, due to the hard decision on one logical error for each block, information is lost in each step. This can be avoided by instead computing the probability of each logical error given the syndrome, and propagating information about this distribution up the tree. Only at the end, the hard decision on the most likely logical error at the root is made.

This improved decoding was proposed by Poulin [Pou06], and is in fact optimal. It corresponds exactly to the BP algorithm introduced in Section 4.3. To understand the connection, we can view the tree structure illustrated in Figure 5.1 as a factor graph. Each code-block is connected to the corresponding logical qubit with a factor node. The function associated with this factor node is an indicator function enforcing two constraints: The error on the qubits n the block has to match both the logical error on the corresponding logical qubit and the observed syndrome of this block. (Thus, technically, we have a separate factor graph for each observed syndrome.) The indicator function is $1$ if the constraints are fulfilled, and $0$ otherwise. By adding an additional factor to each physical qubit, we can also encode the physical error probabilities of the qubits. All in all, the factor graph represents the (unnormalized) probability distribution over both physical and logical errors given the observed syndrome. Finding the most likely logical error corresponds to finding the marginal probability distribution of the root node, which can be accomplished using BP. Since the factor graph is a tree, BP is exact, and the marginal of the root node can be obtained after a single upwards-pass of the messages.

## 5.2 Learning Pauli channels for Concatenated Codes

Now, we turn to the problem of learning the noise affecting the code from the syndrome measurements. For simplicity, we assume that each qubit is affected by an independent Pauli channel. In principle our method is also applicable to correlated

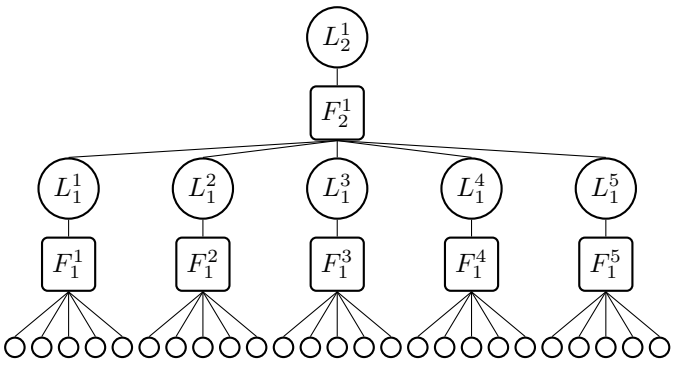



**Figure 5.1.:** Factor graph representation of a concatenated code. Figure from [Wag+21].

errors as long as correlations are only present within each block of the concatenated codes, but not between blocks.

The estimation problem is the following: We are given a data set $D$ of observed syndromes, and we are trying to find the Pauli error rates of each qubit. We assume that the error rate is constant over time, and that errors are not correlated between error correction rounds. In this case, the syndrome measurements correspond to an independent and identically distributed sample. One approach to this estimation problem is to decode each observed syndrome to decide on a corresponding physical error, and then estimate the error rate of each qubit based on these errors. In the literature, the decoding suggested in this context is usually minimum-weight decoding [HL17; Woo20; Fow+14], but if the error rates between qubits differ substantially this might be problematic. More generally, one could start with some guess of the error rates, use maximum-likelihood decoding to obtain a guess for the physical errors, and estimate a new guess of the error rates from these. We call this the *hard-assignment method*, since we decide on one concrete physical error for each syndrome.

Starting from the factor graph perspective on concatenate codes on the other hand, applying methods from the learning of PGMs seems like a natural approach. Indeed, we use the EM algorithm described in Section 4.4. Note that only the factors corresponding to the physical error rates contain learnable parameters, while other factors are fixed constraints. That is, we have one parameter $\theta[i, e]$ for each qubit $i$ and error $e \in \mathsf{P}^1$, corresponding to the probability that the error $e_i$ on the $i$-th qubit is $e$,

$$\theta[i, e] = P(e_i = e)\,. \tag{5.1}$$

In the context of concatenated codes, the EM algorithm then takes the following simple form: First, collect a set $D$ of syndrome measurements. Initialize the current value estimate of the physical Pauli error rates to some guess $\theta^{(0)}$. Then, repeat the following two steps:

1. For each syndrome $S \in D$ and all physical qubits $i$, compute the marginal probability distribution $P(e_i|S, \theta^{(k)})$ of errors on qubit $i$ given the syndrome $S$ using BP. This computation is based on the current estimate $\theta^{(k)}$ of the physical error rates. Then, compute the expected sufficient statistics $M$, which are given by
$$M[i, e] = \sum_{S \in D} P(e_i = e|S, \theta^{(k)}) \tag{5.2}$$

2. Compute a new estimate $\theta^{(k+1)}$ of the physical error rates by normalizing,
$$\theta^{(k+1)}[i, e] = \frac{M[i, e]}{\sum_{e' \in \mathsf{P}^1} M[i, e']} \tag{5.3}$$

Viewing concatenated codes as factor graphs also gives a new perspective on the hard-assignment method. Instead of just decoding once to obtain one new estimate, we could iterate the procedure similar to EM. In fact, the resulting method is exactly HEM, as described in Section 4.4. In our setting, it takes the following form:

First, collect a set $D$ of syndrome measurements. Initialize the current value estimate of the physical Pauli error rates to some guess $\theta^{(0)}$. Then, repeat the following two steps:

1. For each syndrome $S \in D$ compute the most likely error $e_{map}(S) \in \mathsf{P}^{n^r}$ based on the current estimate $\theta^{(k)}$ of the physical error rates. This can be done using the max-sum algorithm, which works similar to BP [Bis06].

2. Compute a new estimate of the physical error rates by counting,
$$\theta^{(k+1)}[i, e] = \frac{\sum_{S \in D} [e_{map}(S)_i = e]}{|D|}\,. \tag{5.4}$$

## 5.3 Numerical Results

In the previous section, we presented the EM algorithm as a method of learning Pauli error rates for concatenated codes, and as an alternative to the previously suggested HEM. We now compare the performance of these two methods, by giving an overview over the simulation results in [Wag+21].

To assess the performance, we simulate the 5-qubit code [Laf+96] concatenated with itself, subject to independent Pauli errors on each qubit. One simulation run consists of the following steps. First, for each qubit, random Pauli error rates are drawn from a Dirichlet distribution. Then, a data set of syndromes is sampled by repeatedly drawing errors from the true distribution and computing their syndrome. We then estimate the Pauli error rates of each qubit using either EM or HEM with a

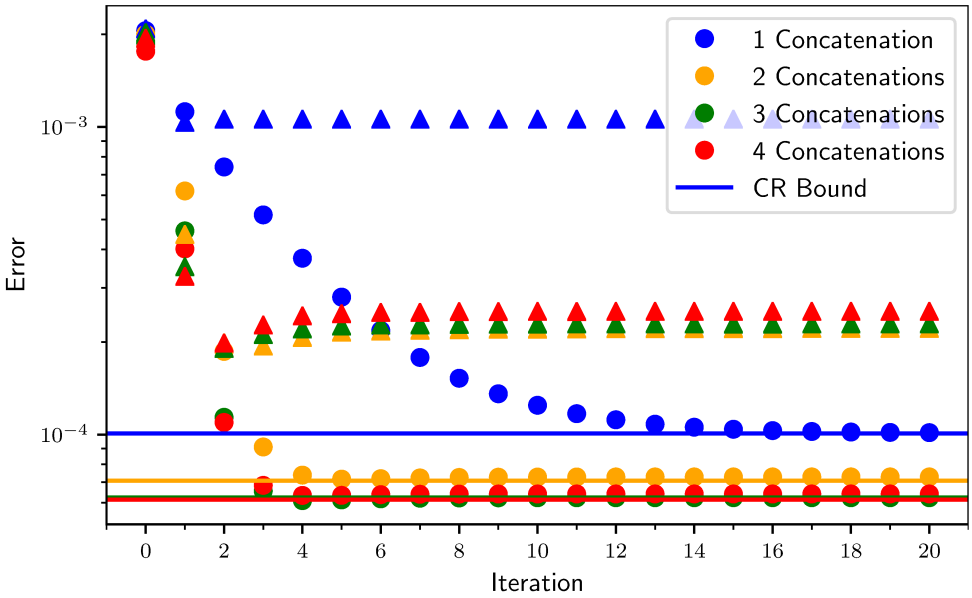



**Figure 5.2.:** Comparison of the MSE in the estimate of the Pauli-$X$ error rate of one qubit, using either EM (circles) or HEM (triangles) for the 5-qubit code concatenated with itself. The CRB for each concatenation level is indicated by a line in the matching color. A sample of $10^3$ syndromes was used in each simulation, and the MSE was determined over $10^3$ simulations. Full details of the simulations are availabe in [Wag+21].

fixed number of iterations. We compute the mean-squared error (MSE) between the estimate and the actual error rates for each qubit over many runs. Note that this includes two sources of randomness: The random draw of the actual error rates, and the random draw of the data set. We can also compare the MSE to the *Cramér-Rao bound (CRB)*, which lower bounds the MSE of any unbiased estimator.

Results for a simulation with an average error rate of $0.13$, which is close to the threshold of the 5-qubit code, are shown in Figure 5.2. Only the MSE in the Pauli-$X$ error rate of the first qubit is shown, but plots for other error rates look similar. It can be seen that the two methods achieve similar errors after the first iteration, but in subsequent iterations the error of EM is much lower. In fact, EM is optimal in the sense that it reaches the CRB, while HEM does not. A more detailed analysis in [Wag+21] suggests that HEM exhibits bias towards its initialization, which does not decrease with the number of iterations. All in all, a clear advantage of EM over the previously suggested HEM can be observed. This also manifests as an advantage when performing degenerate maximum-likelihood decoding using the estimated error rates. In [Wag+21] we demonstrate that using the estimate of EM as input for a decoder, the logical error rate is almost as good as when using the actual error rates as input (which are of course only available in simulations). Decoding

using the estimate from HEM performs much worse, and the difference increases for higher concatenation levels where more error rates need to be estimated accurately.

# 6 A General Framework for the Estimation of Pauli Channels from Syndrome Data

In this section, we give an overview over our theoretical framework for the estimation of Pauli channels from syndrome data. This framework is applicable to arbitrary stabilizer, subsystem, and data-syndrome codes, subject to very general Pauli noise. It is based on a novel combination of tools from Fourier analysis, the theory of error correction codes, and combinatorics. Using this framework, we prove two main results: The first is that the information contained in the syndrome measurements of a code is sufficient to estimate the physical Pauli channel affecting the code, provided that the channel is local in a well-defined sense. The second is that these locality conditions can be significantly relaxed if there is no need to distinguish between logically equivalent errors, as is for example the case in decoding. Furthermore, we develop an efficient estimation algorithm, which neither requires the assumption of vanishing error rates, nor the computation of intractable likelihood functions. We study the sample complexity of this algorithm, and support the theoretical results with first simulations. This section is based mostly on our works [Wag+22a] and [Wag+22b], listed in Appendix C and Appendix D, while the sample complexity bound and the simulations are new results.

## 6.1 Example: Toric Code

Before developing the general theory, let us start with a simple example. Consider the toric code subject to independently distributed Pauli-X errors on each qubit. For this scenario, an analytical formula for the error rates based on the syndrome statistics was given by Spitz et al. [Spi+18]. We give a much simpler derivation of this formula by rephrasing the problem in terms of moments instead of probabilities.

Consider a region of the toric code as depicted in Figure 6.1. We are given the probability distribution of the syndrome measurements, and our task is to find the error rate $p_4$, which is the probability of a Pauli-$X$ error on qubit $4$. We see that equivalently, we could estimate the expectation value $E(Z_4) = 1 - 2p_4$, where $Z_4$ is a random variable that takes the value $+1$ if there is no error on qubit $4$ and $-1$ if there

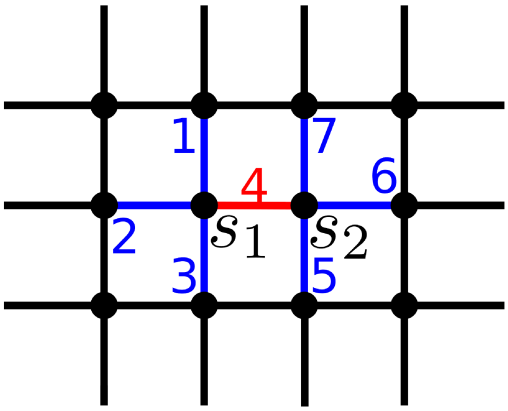



**Figure 6.1.:** A region of the toric code. Each edge represents a qubit, and each vertex represents a Pauli-Z stabilizer acting on all adjacent qubits. (Adapted from [Wag+22a].)

is an error. We use the notation $Z_4$ for this random variable since it corresponds to a hypothetical measurement of a Pauli-$Z$ operator on qubit $4$, if we were to start in a $+1$-eigenstate of this operator and then apply an error with probability $p_4$. Note however, that since the codewords are not such an eigenstate, it does not actually correspond to a physical measurement of this operator. Since errors are independent on each qubit, the expectation of a stabilizer measurement factors into a product of expectations on the adjacent qubits, resulting in the equations

$$\begin{aligned} E(s_1) &= E(Z_1)E(Z_2)E(Z_3)E(Z_4)\,, \\ E(s_2) &= E(Z_5)E(Z_6)E(Z_7)E(Z_4)\,, \\ E(s_1 s_2) &= E(Z_1)E(Z_2)E(Z_3)E(Z_5)E(Z_6)E(Z_7)\,. \end{aligned}$$

By solving this system of equations we can express the desired expectation $E(Z_4)$ only in terms of measured quantities, as

$$E(Z_4) = \pm\sqrt{\frac{E(s_1)E(s_2)}{E(s_1 s_2)}}\,. \tag{6.1}$$

Note that the choice of sign corresponds to choosing whether $p_4$ is smaller or larger than $\frac{1}{2}$. Under the reasonable assumption that $p_4$ is smaller than $\frac{1}{2}$ however, the solution is unique. Eq. (6.1) is exactly the solution derived by Spitz et al. [Spi+18], only expressed in arguably simpler terms. For an explicit calculation showing this equivalence see [Wag+22a]. A generalized version of this example, going beyond the method developed by [Spi+18], is presented in Appendix A.1.

## 6.2 The General Framework

We now give an overview of our theoretical framework and our main results in the general setting. This section closely follows our published works [Wag+22a] and [Wag+22b]. These works also contain all details and proofs that are omitted here.

We phrase our results in the unified language of Section 3.6: we describe errors and measurements as elements of a finite Abelian group $A = \prod_{i=1}^{n} A_i$, where each $A_i$ is $\mathsf{P}^1$ or $\mathbb{F}_2$. This group is equipped with the product bicharacter (2.23). A code is described by four subgroups of $A$. First, we have the group of measurements $\mathscr{M}$ and the gauge group $\mathscr{G}$. Elements of the gauge group are considered logically trivial. The annihilator of the measurement and the gauge group respectively are the group of undetectable errors $\mathscr{U} \coloneqq \mathscr{M}^{\perp}$ and the group of logical operators $\mathscr{L} \coloneqq \mathscr{G}^{\perp}$. We require that logically trivial errors are also undetectable, i.e. $\mathscr{G} \subseteq \mathscr{U}$. Dual to this, we have the inclusion $\mathscr{M} \subseteq \mathscr{L}$. An overview over these groups is given in Figure 3.1. Finally, measurements are described by the product bicharacter of $A$. Measuring $a \in \mathscr{M}$ after the error $e \in A$ occurred results in the outcome $\langle a, e\rangle = \pm 1$. We consider a phenomenological noise model where, before each round of measurements, an error $e \in A$ occurs. This error occurs according to a fixed channel $\Lambda_P$ with error distribution $P : A \to [0, 1]$. Our task is to estimate $P$, using only the syndrome measurements of the code.

Note that, depending on how exactly QEC is implemented, we are actually measuring the syndrome of the error that has accumulated over all previous rounds. This accumulated error is not distributed according to $P$. We can easily fix this by always considering the difference to the syndrome of the previous round, instead of the raw syndrome. In this way, we obtain a syndrome measurements only describing the new error in each round. (One has to be a bit more careful in the presence of measurement errors, i.e. for data-syndrome codes, for details see [Wag+22b]).

We first focus on the problem of identifiability, and ask whether the distribution $P$ can be obtained (uniquely) from only syndrome data at all. Thus, we assume that we have access to the exact syndrome statistics. In other words, we consider the infinite sample limit first. We return to the issue of practical estimation with a finite amount of samples in Section 6.3.

### 6.2.1 Moments

The data we are given consists of the probability of each syndrome under the error distribution $P$. Unfortunately, the probability of observing a given syndrome is described by a sum over all errors compatible with this syndrome, which is an exponentially large sum. Therefore, we want to avoid dealing directly with

syndrome probabilities. Instead, we describe the problem in Fourier space, which is a description in terms of moments instead of probabilities. Formally, we define the moments as the Fourier transform of the probabilities,

$$E := \mathscr{F}[P]\,. \tag{6.2}$$

Explicitly, this means that for each $a \in A$,

$$E(a) = \sum_{e \in A} \langle a, e \rangle P(e)\,. \tag{6.3}$$

Since the Fourier transform is invertible, knowing $E(a)$ for all $a \in A$ is equivalent to knowing the full distribution $P$. Furthermore, we see that for a measurement $s \in \mathscr{M}$, $E(s)$ is exactly the expectation of the measurement of $s$ over many rounds. This motivates the name "moments". More importantly, it means that, for $s \in \mathscr{M}$, we can compute $E(s)$ from the measured syndrome statistics. For general $a \in A$, however, Eq. (6.3) is just a formal definition, and we cannot measure $E(a)$. All in all, we see that the estimation task can be formulated as follows:

Given the moments $E_{\mathscr{M}} = (E(s))_{s \in \mathscr{M}}$, compute all other moments $E = (E(a))_{a \in A}$.

### 6.2.2 Local Noise

Unfortunately, we cannot expect this task to be solvable for a completely arbitrary channel $P$. The syndrome measurement do not tell us anything about the transformation of the encoded information, and thus cannot fully characterize the channel. More generally, the moments $E$ are independent from each other, and thus measuring the moments $E_{\mathscr{M}}$ will not give us any information about the remaining moments. This situation changes if we restrict attention to specific classes of channels, where relations between the moments might exist. Our goal is thus to characterize for which classes of channels estimation is possible, and for which it is not. Specifically, we focus on *local* channels, i.e. channels that do not have correlations across a large number of qubits. In the context of QEC, locality of the noise is a fundamental assumption. It implies that high-weight errors can only occur as a combination of several independent smaller errors. This is the setting where error correction has a chance to be successful. We will see that locality of the noise is also a necessary assumption for its estimation from syndromes.

We formalize our locality assumption in the following way. We describe the full channel by a set of supports $\Gamma \subseteq 2^{[n]}$. Explicitly, each $\gamma \in \Gamma$ is a region $\gamma \subseteq [n]$ of the code. These regions are allowed to overlap, and they need not be geometrically local. We assume that on each region $\gamma \in \Gamma$, there acts an independent channel $\Lambda_{P_\gamma}$

with error distribution $P_\gamma : A_\gamma \to [0,1]$. The total error in each round is obtained by combining the small errors occurring on each region $\gamma$. In other words, the total channel $\Lambda_P$ factors into a concatenation of locally supported channels, as

$$\Lambda_P = \circ_{\gamma\in\Gamma}\Lambda_{P_\gamma}\,. \tag{6.4}$$

Since concatenating channels corresponds to convolving their distributions, the complete distribution $P$ factors as a convolution. We summarize this discussion in a definition:

**Definition 15** (Local channel)**.** *A channel with distribution $P : A \to [0,1]$ is* local *if $P$ can be written in the form*

$$P = *_{\gamma\in\Gamma}\, P_\gamma\,, \tag{6.5}$$

*for a set $\Gamma \subseteq 2^{[n]}$ of local regions $\gamma \subseteq [n]$, and locally supported distributions $P_\gamma : A_\gamma \to [0,1]$.*

In Eq. (6.5), the individual distributions $P_\gamma$ are extended impulsively from $A_\gamma$ to $A$ (Chapter 2), i.e. $P_\gamma(e) = 0$ if $e \notin A_\gamma$. Physically, the factorization (6.5) means that there are many independent error mechanisms that each only affect a limited region. Formally, Eq. (6.5) is a description of the error distribution $P$ as a factor graph (Section 4.1), replacing the product for standard factor graphs with a convolution. Such convolutional factor graphs were studied in [MK05]. The idea of describing Pauli noise as factor graphs has also been used in [FW20], however using product instead of convolutional factor graphs.

The simplest example of local noise is independent single-qubit noise, where each region $\gamma$ contains only a single qubit. On the other hand, any channel can be factorized in the form (6.5) if we allow the regions $\gamma$ to be arbitrarily large, for example by using a single region of size $n$. Thus, we will describe conditions on the size of these regions that must be fulfilled for unique estimation of the channel to be possible. For the estimation of physical Pauli channels, these conditions are related to the pure distance of the code. If one is only interested in estimating noise up to logical equivalence the conditions are relaxed, and related to the regular distance of the code.

By Fourier transforming Eq. (6.5), we obtain a description of the moments as a product factor graph,

$$E = \prod_{\gamma\in\Gamma} E_\gamma\,, \tag{6.6}$$

where we denote $E_\gamma = \mathscr{F}[P_\gamma]$. The local functions $E_\gamma : A_\gamma \to [-1,+1]$ are extended periodically from $A_\gamma$ to $A$, which follows from the duality between impulsive and periodic extensions (Lemma 12).

### 6.2.3 Canonical Moments

Our next goal is to make the locality of the channel obvious in its description in terms of moments. Intuitively, there are two issues with the moments $E$. The first is that a moment $E(a)$ captures correlations not just across the full support of $a$, but also over all strict subsets of the support. Consequently, high-weight moments take non-trivial values even for local channels. The second is that the regions $\gamma$ can overlap, and in this case the factorization (6.6) is not unique. For example, a factor $E_{1,2}$ and $E_{2,3}$ both contribute to the moment $E(X_2)$, and there is a continuum of choices for $E_{1,2}(X_2)$ and $E_{2,3}(X_2)$ leading to the same overall result $E(X_2)$. Therefore, we would like to find a canonical choice of factorization.

To accomplish these goals, we introduce the *canonical moments* $F : A \to \mathbb{R}$. Intuitively $F$ is obtained as follows: from each moment $E(a)$, we divide out all correlations that are already captured by lower-weight moments $E(b)$, where $b$ is a substring of $a$. Then, only correlations across the whole support of $a$ remain. However, one has to be careful to avoid double-counting substrings in this process. We achieve this by employing the *Möbius inversion*, which is a generalization of the inclusion-exclusion principle of combinatorics [Aig07]. For this purpose, we consider the group $A = \prod_{i=1}^{n} A_i$ as a partially ordered set (poset), ordered by substring relation:

**Definition 16** (Substring relation)**.** *We say $b \in A$ is a substring of $a \in A$ if for all $i \in [n]$ either $b_i = I$ or $b_i = a_i$. In this case we write $b \leq a$.*

For example, $IXZI$ is a substring of $XXZI$, but $IZZI$ and $IIIZ$ are not. Now, we introduce the *Möbius function* $\mu$ of $A$. For our purposes, we define $\mu$ as the function fulfilling the following inversion theorem, taken from [Wag+22b]. A more general definition, additional explanation, and a proof that $\mu$ exists for locally finite posets can be found in [Aig07, Chapter 5.2], see also [Rom06] for the multiplicative version of the inversion theorem.

**Definition 17** (Möbius function and Möbius inversion)**.** *Let $S$ be a partially ordered set. The Möbius function $\mu$ of $S$ is the function $\mu : S \times S \to \mathbb{R}$ such that for any two functions $f, g : S \to \mathbb{R}$,*

$$f(t) = \prod_{s \leq t} g(s)\,, \tag{6.7}$$

*if and only if*

$$g(t) = \prod_{s \leq t} f(s)^{\mu(s,t)}\,. \tag{6.8}$$

In our setting, $\mu$ is given explicitly by

$$\begin{aligned} &\mu : A \times A \to \mathbb{R} \\ &\mu(b, a) \coloneqq \begin{cases} (-1)^{|a|-|b|}\,, \text{ if } b \leq a \\ 0\,, \text{ otherwise} \end{cases} . \end{aligned} \tag{6.9}$$

A proof of this can be found e.g. in [Wag+22b]. Finally, we define the canonical moments $F$ by

$$F(a) = \prod_{b \leq a} E(b)^{\mu(b,a)} . \tag{6.10}$$

Unpacking this definition, e.g. for a weight $3$ string $a \in A$, we would divide $E(a)$ by the moments of all the weight $2$ substrings, and then multiply the moments of the weight $1$ substrings back in. Mathematically, the canonical moments are closely related to the canonical factorization of a factor graph (Section 4.2 and [AKN06]). They have two important properties:

**Lemma 18** (Properties of canonical moments)**.**

- *The moments $E$ can be expressed by the canonical moments as*

$$E(a) = \prod_{b \leq a} F(b) . \tag{6.11}$$

- *For any $a \in A$ such that $\operatorname{supp}(a) \not\subseteq \gamma$ for all $\gamma \in \Gamma$, $F(a) = 1$.*

A proof is given in [Wag+22b]. The first statement explains the relation between canonical moments and moments, and corresponds excactly to the Möbius inversion theorem (Definition 17). The second statement shows that all canonical moments that are not supported in a region $\gamma \in \Gamma$ are trivial. This formalizes the intuition that canonical moments only capture correlations across their full support, but not correlations across a strict subset of their support. In particular, all moments beyond the "correlation length" of the channel are trivial.

All in all, we obtain that the full channel is described by a small set low-weight canonical moments, given by

$$\Gamma' = \{a \in A \,:\, \exists \gamma \in \Gamma \text{ such that } \operatorname{supp}(a) \subseteq \gamma\} . \tag{6.12}$$

From now on, we denote by $F$ the vector of non-trivial canonical moments only. The description in terms of canonical moments leads to a significant simplification of the estimation problem. While the number of moments $E$ scales exponentially in $n$, since there is one moment $E(a)$ for each $a \in A = \prod_{i=1}^{n} A_i$, the number of non-trivial

canonical moments only scales polynomially as long as the size of the regions $\gamma$ is bounded by a constant.

### 6.2.4 Identifiability of Physical Channels

As explained in Section 6.2.1, for any measurement $s \in \mathscr{M}$, we can obtain $E(s)$ from the syndrome statistics. Furthermore, in the previous section we have seen that the full physical channel is completely described by the set of non-trivial canonical moments $F$. Thus the problem of estimating the channel from the syndrome statistics can be formulated as the task of computing the (non-trivial) canonical moments $F$ from the measured moments $E_{\mathscr{M}}$. According to Lemma 18, these quantities are related by the system of equations

$$E(s) = \prod_{b \leq s} F(b)\,, s \in \mathscr{M}\,, \tag{6.13}$$

which we have to solve for the canonical moments $F$. In particular, the physical channel is identifiable from the syndrome measurements if and only if this system has a unique solution.

Let us introduce a more concise notation for such equation systems:

**Notation 19.** *For a vector $v \in \mathbb{R}^n$ and a matrix $M \in \mathbb{R}^{m \times n}$, we write $v^M$ for the $m$-element vector with entries*

$$(v^M)_i = \prod_{j=1}^{n} v_j^{M_{i,j}}\,. \tag{6.14}$$

Using this notation, the system (6.13) becomes

$$E_{\mathscr{M}} = F^D\,, \tag{6.15}$$

where $D$ is a coefficient matrix whose rows are labeled by measurements $s \in \mathscr{M}$, and whose columns are labeled by canonical moments $b \in \Gamma'$. The entries of $D$ are

$$D[s,b] = \begin{cases} 1\,, & b \leq s \\ 0\,, & \text{otherwise} \end{cases}\,. \tag{6.16}$$

Which conditions have to be imposed on the channel $P$ for the system (6.13) to have a unique solution? Certainly, if there exist two sources of errors with the same syndrome, the error rates cannot be identifiable since we cannot distinguish these errors and can trade off their rates without changing the syndrome statistics. Furthermore, we have already seen in the toric code example (Section 6.1) that the

equation system (6.13) can have multiple solutions that differ by signs. Thus, we need to restrict ourselves to one sector of the solution space. Here, we choose the sector where all moments are positive, which manifests as an assumption that error rates should be smaller than $\frac{1}{2}$. We formalize the two necessary requirements above in the following two definitions.

**Definition 20** (Physically correctable region)**.** *We say that a region $R \subseteq [n]$ of a code is* physically correctable *if it does not support any undetectable errors, i.e. any undetectable error $e \in \mathscr{U}$ supported on $R$ is the identity $I$.*

In particular, any region $R$ containing less than $\lfloor \frac{d_p - 1}{2} \rfloor$ qubits is physically correctable, where $d_p$ is the pure distance of the code. This is an easy way to roughly measure the error detection capacity of a code. However, physically correctable regions can be much larger than the pure distance.

**Definition 21** (Physically correctable channel)**.** *A channel $P : A \to [0, 1]$ of the form* (6.5) *is said to be* physically correctable *if the following two conditions are fulfilled:*

1. *For every choice of two supports $\gamma_1, \gamma_2 \in \Gamma$, the union $\gamma_1 \cup \gamma_2$ is a physically correctable region.*

2. *All moments are positive, i.e. $E(a) > 0$ for all $a \in A$.*

The first condition formalizes the discussion above, that there should be no two sources of errors with the same syndrome. Note that this does not mean that the total channel $P$ cannot support any undetectable errors. Such errors are allowed to occur frequently, but they must arise as a combination of smaller independent errors. The second condition states that error rates should be small enough. It is fulfilled in particular if $P(I) > \frac{1}{2}$, or if $P_\gamma(I) > \frac{1}{2}$ for all $\gamma \in \Gamma$.

Our first main result is that these two conditions are not only necessary, they are in fact sufficient for noise to be identifiable from syndrome statistics. No further assumptions on the noise are needed. In this sense, the situation is as good as one could reasonably hope for.

**Theorem 22** (Identifiability of physical channels)**.** *A channel $P : A \to [0, 1]$ can be uniquely estimated from the syndrome measurements of a code if and only if it is physically correctable in the sense of Definition 21.*

The proof of this result is quite involved and given in detail in [Wag+22a]. Here, we only sketch the most important ingredients.

*Proof sketch of Theorem 22.* We have already explained the "only if" direction above, so it remains to prove the "if" direction. We start from the system of equations

(6.15). Since we assume that all moments are positive, this system of equations can be transformed into a linear system by taking logarithms. It is described by the coefficient matrix $D$, whose entries are all $1$ or $0$ and related to the measurements in $\mathscr{M}$ by Eq. (6.16). Our task is to show that this linear system has a unique solution, i.e. that $D$ has full rank. Since we do not know the explicit form of the measurements in $\mathscr{M}$, we do not explicitly know the rows of $D$. Instead, we consider the matrix $D^T D$ whose rows and columns are labeled by canonical moments, i.e. elements of $\Gamma'$. Direct calculation shows that

$$D^T D[a,b] = |\{s \in \mathscr{M} \,:\, a,b \leq s\}| \,. \tag{6.17}$$

Crucially, the entries of this matrix can be calculated from global properties of the group $\mathscr{M}$, without knowing the explicit form of its elements. We compute the right-hand side of (6.17) by using local randomness properties of the code. More explicitly, it can be shown that on any physically correctable region $R$, the measurements of a code look essentially random, i.e. any substring $b \in A_R$ appears equally often in the group $\mathscr{M}$ [Del73; Fuj14b]. Since, by the assumption of physical correctability, the union of the supports of any two canonical moments is a physically correctable region, we can explicitly compute the entries of $D^T D$ based on these local randomness properties. Once the entries have been computed explicitly, we can show that $D^T D$ has full rank using a (lengthy) argument based on induction and Schur complements. □

All in all, we see that estimation of the channel $P$ can be performed by solving Eq. (6.13), which is guaranteed to have a unique solution for physically correctable channels. We have already seen one specific example in Section 6.1. There, we derived solution for the toric code, which was based on the fact that each single error only affected two syndrome bits. This can be viewed as a method to estimate the edge weights in the decoding graph of a standard minimum-weight matching decoder. Most codes however do not admit a simple minimum-weight matching decoder, since they require decoding graphs that contain hyperedges. We consider the important task of estimating the corresponding hyperedge weights in Appendix A.1, which provides another relatively simple example of our framework.

### 6.2.5 Identifiability of Logical Channels

In Section 6.2.4, we have shown that a Pauli channel can be estimated from syndrome measurements if it is physically correctable in the sense of Definition 21. However, the assumption of physical correctability can be quite restrictive. In particular, we have seen that it imposes limits on the correlations of the channel which are related

to the pure distance of the code. Unfortunately, the pure distance is constant in the code size (and often small) for topological and LDPC codes. Furthermore, some natural noise models, such as propagation of errors in the measurement circuits, are not physically correctable since they induce undetectable correlated errors. On the other hand, these undetectable errors need not be harmful for error correction, since they are logically trivial if the error correction circuits are designed well.

More generally, not all information about the physical channel is actually useful or required for error correction. As explained in Section 3.2, for decoding it suffices to know the induced logical channel. Thus, a natural question is under which conditions we are able to estimate the logical channel from syndrome data, even if we may not be able to estimate the physical channel. This question is adressed in our work [Wag+22b], of which we summarize the main points here.

Consider a local channel $P$, as defined in Definition 15, described by a set of supports $\Gamma$. Explicitly,

$$P = *_{\gamma\in\Gamma} P_\gamma \,. \tag{6.18}$$

For an abstract code as defined in Section 3.6, logical equivalence is described by the gauge group $\mathscr{G}$. Thus, we define the *logical channel* $P_L$ by averaging over this group,

$$P_L \coloneqq \frac{1}{|\mathscr{G}|} \sum_{s\in\mathscr{G}} P(es) \,. \tag{6.19}$$

We now want to adapt the framework developed in the previous sections to the logical channel $P_L$ instead of the physical channel $P$. Thus we want to describe the logical channel in terms of moments. It follows directly from the definitions that we can express the logical channel as

$$P_L = P * U_{\mathscr{G}} \,, \tag{6.20}$$

where $U_{\mathscr{G}}$ is the scaled indicator function (Notation 1) of $\mathscr{G}$. Since $\mathscr{G}^\perp = \mathscr{L}$, Lemma 10 then implies

$$\mathscr{F}[P_L] = E \cdot \Phi_{\mathscr{L}} \,, \tag{6.21}$$

where $\Phi_{\mathscr{L}}$ is the (unscaled) indicator function (Notation 1) of $\mathscr{L}$. In other words, the logical channel is fully described by moments corresponding to logical operators. Similarly to Section 6.2.1, the estimation task can thus be summarized as follows: We are given the moments $E_{\mathscr{M}} = (E(s))_{s\in\mathscr{M}}$ corresponding to the measurements, and have to compute all moments $E_{\mathscr{L}} = (E(l))_{l\in\mathscr{L}}$ corresponding to logical operators. The difference to the estimation of the physical channel is that we only need to compute a certain subset of moments, not all moments, from the available information.

As derived in Section 6.2.3, the regular moments can be described in terms of a set of canonical moments. These canonical moments capture the locality assumptions about the channel. The relation between regular and canonical moments is expressed by a coefficient matrix $D$, with rows labeled by regular moments and columns labeled by canonical moments, as in Eq. (6.15). In particular, the moments $E_{\mathscr{L}}$ are described by a sub-matrix $D_{\mathscr{L}}$, and the measurements are expressed by a sub-matrix $D_{\mathscr{M}}$. The question of identifiability reduces to a question about these coefficient matrices: We can estimate the logical channel from the available measurements if the rows of $D_{\mathscr{L}}$ linearly depend on the rows of $D_{\mathscr{M}}$.

Of course, this is not the case for arbitrary Pauli channels $P$, since all moments are in principle independent. We must again impose some restrictions on the size of the supports $\gamma \in \Gamma$. In analogy to Definition 20 and Definition 21, we define:

**Definition 23.** *(Correctable region) We say that a region $R \subseteq [n]$ of a code is* correctable *if all undetectable errors supported in this region are gauge operators, i.e. any error $e \in \mathscr{U}$ supported on $R$ is an element of $\mathscr{G}$.*

This definition is quite natural in the context of QEC: A region is correctable if it does not support any errors that can affect the encoded information. Correctable channels are now defined exactly in the same way as physically correctable channels (Definition 24), only replacing the requirement of *physically* correctable regions with just correctable regions.

**Definition 24** (Correctable channel)**.** *A channel $P : A \to [0, 1]$ of the form* (6.5) *is said to be* correctable *if the following two conditions are fulfilled:*

1. *For every choice of two supports $\gamma_1, \gamma_2 \in \Gamma$, the union $\gamma_1 \cup \gamma_2$ is a correctable region.*
2. *All moments are positive, i.e. $E(a) > 0$ for all $a \in A$.*

Despite the apparent similarity to Definition 23 and Definition 24, these are much weaker conditions than the analogues for the physical channel. Indeed, any region containing less qubits than the distance $d$ of a code is correctable. Contrast this with physically correctable regions, whose size is related to the pure distance. For LDPC codes, the distance scales with the size, while the pure distance is constant. For example, the pure distance of the toric code is 4 independent of lattice size, while the distance is exactly the linear lattice size. The conditions are also more natural for QEC, since error correction is not concerned with logically trivial (gauge) errors.

It is not hard to see that correctability is a necessary requirement for the logical channel to be identifiable from syndrome data. If the underlying physical channel $P$ is not correctable, then there exist two independent sources of errors with the

same syndrome which are not logically equivalent. In this case, one could trade off the corresponding error rate, changing the logical channel, without altering the syndrome statistics. Our main result in this section is that this is, in fact, the only condition necessary for identifiability of the logical channel. This complements the analogous result (Theorem 22) for the physical channel.

**Theorem 25** (Identifiability of logical channels)**.** *The logical channel $P_L$ induced by a channel $P : A \to [0, 1]$ can be estimated from the syndrome measurements of a code if and only if the channel $P$ is correctable in the sense of Definition 24.*

Again, we only give a rough sketch of the most important points of the proof. Details are provided in [Wag+22b].

*Proof sketch of Theorem 25.* As explained above, the logical channel is identifiable if the rows of $D_{\mathscr{L}}$ are linearly dependent on the rows of $D_{\mathscr{M}}$. Since $\mathscr{M} \subseteq \mathscr{L}$, $D_{\mathscr{M}}$ is a sub-matrix of $D_{\mathscr{L}}$ and thus it suffices to show $\operatorname{rank}(D_{\mathscr{M}}) = \operatorname{rank}(D_{\mathscr{L}})$. Similarly to the proof of Theorem 22, we reduce the problem to global properties of the group of measurements and the group of logical operators by considering the matrices $D_{\mathscr{M}}^T D_{\mathscr{M}}$ and $D_{\mathscr{L}}^T D_{\mathscr{L}}$. Ultimately, to show that $\operatorname{rank}(D_{\mathscr{M}}) = \operatorname{rank}(D_{\mathscr{L}})$, we need to compare the number of logical operators and the number of measurements supported in any given correctable region. We do this by using the *cleaning lemma* of QEC. This lemma states that for a stabilizer code, the stabilizer group and the group of logical operators look identical on any correctable region [Bra+13]. A substantially generalized version of the cleaning lemma was recently developed [KS22], which we employ to obtain the analogous result for the abstract groups of measurements and the logical operators we consider. Since the structure of these groups are the same on any correctable region, the number of logical operators or measurements supported in any given correctable region only differs by global constants. From this, we can prove the required rank equality. □

## 6.3 Concrete Estimation

So far, we have mainly discussed the question of identifiability, and focused on the infinite sample limit where the expectation values of the measurements are known exactly. However, the results presented in the previous sections can also be cast into a concrete estimator. We now describe this estimator, and consider its accuracy for a finite amount of data.

### 6.3.1 Summary of the Estimation Algorithm

In Section 6.2, we developed the system of equations (6.15), which relates the measured expectations to the canonical moments describing the channel. By replacing the exact expectations with empirical expectations in this system, we obtain an estimator for the channel. Explicitly, estimation consists of the following steps:

1. Perform $m$ syndrome measurements and use them to compute an empirical estimate $\hat{E}_{\mathscr{M}}$ of the expectations $E_{\mathscr{M}} = (E(s))_{s\in\mathscr{M}}$.

2. Insert the empirical expectations into Eq. (6.15), and solve the resulting system of equations for an estimate $\hat{F}$ of the canonical moments.

3. Insert the estimate $\hat{F}$ into Eq. (6.11) to obtain an estimate $\hat{E}$ of the moments.

4. Perform the inverse Fourier transform to obtain an estimate $\hat{P}$ of the error distribution.

In statistical terms, this is an instance of the *method of moments*, see e.g. [JC11]. We express the quantities we want to estimate in terms of the exact moments, and then insert an empirical estimate for these moments.

One issue is that the system (6.15) will generally be over-determined. Since there is some error in the empirical expectation values $\hat{E}_{\mathscr{M}}$, inserting them in place of the exact expectations $E_{\mathscr{M}}$ can results in a system that does not have an exact solution. However, we can still compute a least-squares solution as

$$\hat{F} = \operatorname{argmin}_F \frac{1}{2}\|\hat{E}_{\mathscr{M}} - F^D\|_2^2 . \tag{6.22}$$

Usually, this is done using some variant of gradient descent, which requires a choice of initialization. Fortunately, a good initialization can be easily obtained in our setting: since the system of equations (6.15) is linear after taking logarithms, we can start by computing the least-squares estimate of $\ln(\hat{F})$. This can be done e.g. via the pseudo-inverse of the coefficient matrix $D$, or one of several numerically stable alternatives. While exponentiating this solution does not result in the actual least-squares estimate of $\hat{F}$, it is generally close. Thus, we can use this as a good initialization for the minimization (6.22).

Another consideration is that the system (6.15) generally contains an exponential number of equations. For example, for a stabilizer code with $r$ stabilizer generators, there are $2^r$ equations. However, it is not necessary to consider all of these. We can simply select a number of equations proportional to the number of parameters in the noise model, as long as we are careful to preserve the rank of the coefficient matrix $D$. For example, for the toric code with independent single-qubit noise,

three equations per qubit suffice (Section 6.1). Furthermore, for the toric code, these equations can be selected by considering local regions of the code. For any given qubit, only neighboring measurements need to be considered. Thus, the error rates for each qubit can be estimated separately, each from a small system of equations. We expect that such an estimator based on local regions also works for other topological codes.

For the case of independent single-qubit errors, an implementation of the estimator described in this section is available on Github [Wag]. Furthermore, we sketch in Appendix A.1 how this estimation scheme can be used to calibrate minimum-weight decoders for arbitrary stabilizer codes.

### 6.3.2 Sample Complexity Bound

Next, we want to assess the amount of syndrome measurements we need in order to achieve good estimates of the channel. More precisely, we want to compute the *sample complexity* of our estimator. This is the required amount of samples to achieve, with a high probability of $1 - \delta$, a low error of at most $\epsilon$ in the estimate. In this section, we give a rigorous bound on the sample complexity of the estimator described in Section 6.3.1. We only consider the setting of independent single-qubit Pauli noise, where the canonical and the regular moments coincide.

We consider a code (or a local region of a code) with $n$ qubits, and denote as $\mathscr{M} \subseteq \mathsf{P}^n$ the subset of all measurements that we use for estimation. Since we might not use all measurements of the code, $\mathscr{M}$ is not necessarily a group. We denote the number of measurements we use in each round as $k = |\mathscr{M}|$.

For independent single-qubit Pauli errors, the noise is fully described by $4$ error rates per qubit. While one of these error rates is technically redundant, we include it for mathematical convenience when using the Fourier transform. We denote the vector of all error rates as $P$, and use $P_i$ for the marginal distribution of errors on the $i$-th qubit. Fourier transforming separately on each qubit yields $4$ moments per qubit, which also coincide with the canonical moments. We collect these into a parameter vector $\theta$, containing four entries per qubit. The actual noise is described by a parameter vector $\theta^*$. The existence of a "true" parameter vector $\theta^*$ is of course an idealized assumption, and does not account for a possible "model mismatch". For each measurement $s \in \mathscr{M}$ there is one expectation value $E(s)$, and we use $y = (E(s))_{s \in \mathscr{M}}$ to denote the vector of these expectations. The expectations under the actual error distribution are denoted as $y^*$. By measuring a set of $m$ syndromes, we obtain an empirical estimate $\hat{y}$ of $y^*$.

Finally, we will assume that the moments are all bounded from below by some constant $\beta > 0$, in particular $y^*[s] > \beta$ for all $s \in \mathscr{M}$, and also $\theta^* > \beta$ element-wise.

This corresponds to an upper bound on the error rates. We discuss this condition in more detail later.

The actual expectations and parameters are related by the system of equations (6.15), which in the current notation reads

$$y^* = g(\theta^*) \coloneqq (\theta^*)^D \,. \tag{6.23}$$

As described in Section 6.3.1, we obtain an estimate $\hat{\theta}$ of the parameters by a least-squares minimization,

$$\hat{\theta} = \mathrm{argmin}_{\theta \in \mathcal{D}} \frac{1}{2} \|\hat{y} - g(\theta)\|_2^2 \,, \tag{6.24}$$

where we only need to optimize over the domain

$$\mathcal{D} = \{\theta \in \mathbb{R}^{4n} \,:\, \beta \leq \theta_i \leq 1 \,\forall i \in [4n]\} \,, \tag{6.25}$$

because we assumed a lower bound $\beta$ on the moments.

We are interested in bounding the error in $\hat{\theta}$ as a function of the number of samples $m$. We start by bounding the error in the empirical estimate $\hat{y}$, using the *Hoeffding's bound* [Hoe63]. Since the outcome of each measurement is $\pm 1$, the Hoeffding bound states that, for each $s \in \mathscr{M}$, the probability that the empirical expectation deviates more than $\epsilon$ from the actual expectation $y^*$ is bounded as

$$P(|\hat{y}[s] - y^*[s]| > \epsilon) < 2 \exp\left(-\frac{m\epsilon^2}{2}\right) \,. \tag{6.26}$$

Combining this with a standard *union bound* over all of the $k$ measurements $s \in \mathscr{M}$, we obtain that

$$P(\|\hat{y} - y^*\|_\infty > \epsilon) < 2k \exp\left(-\frac{m\epsilon^2}{2}\right) \,. \tag{6.27}$$

In other words, to guarantee that the error $\|\hat{y} - y^*\|_\infty$ is smaller than $\epsilon$ with probability at least $1 - \delta$, we need at least

$$m \geq \frac{2}{\epsilon^2} \ln\left(\frac{2k}{\delta}\right) \tag{6.28}$$

samples. Since $\|\hat{y} - y^*\|_2 \leq \sqrt{k}\|\hat{y} - y^*\|_\infty$, the required number of samples to control the 2-norm error is instead

$$m \geq \frac{2k}{\epsilon^2} \ln\left(\frac{2k}{\delta}\right) \,. \tag{6.29}$$

We want to translate the bound on the error in $y$ into a bound on the error in $\theta$. This requires analyzing the stability of the least-squares solution (6.24). A detailed

analysis, based on basic tools from the pertubation theory of optimization problems [BS00], can be found in Appendix A.2. The result is the following bound on the error in $\theta$.

**Lemma 26.** *Assume that both $y^* \geq \beta > 0$ and $\theta^* \geq \beta$ element-wise, and that there exists a lower bound $0 < \tilde{\beta} \leq \beta - 4\|y^* - \hat{y}\|_2$. Then the error in $\theta$ is bounded by the error in $y$ as*

$$\|\hat{\theta} - \theta^*\|_2 \leq \frac{2}{\tilde{\beta}^2} \frac{\sigma_{max}(D)}{\sigma_*(D)^2} \|\hat{y} - y^*\|_2 \,. \tag{6.30}$$

*Here, $\sigma_{max}(D)$ denotes the maximal singular value of $D$, and $\sigma_*(D)$ denotes the minimal non-zero singular value of $D$.*

While strictly speaking, this result assumes that $D$ is full-rank, a similar bound also holds if $D$ is rank-deficient, provided we project onto the part of $\theta$ that can still be uniquely estimated, which is the orthogonal complement of the kernel of $D$. This is discussed in detail in Appendix A.2. Furthermore, note that the lower bound $\tilde{\beta}$ can be related to the given lower bound $\beta$. For a given number of samples, an estimate of $\|y^* - \hat{y}\|_2$, and thus of $\tilde{\beta}$, can be obtained from Hoeffding's bound (6.26). If the number of samples is large, $\tilde{\beta} \approx \beta$.

Combining Lemma 26 with Eq. (6.29), we see that the we can obtain an error $\|\hat{\theta} - \theta^*\|_2 < \epsilon$ with probability at least $1 - \delta$ using

$$m \geq \frac{8k}{\tilde{\beta}^4 \epsilon^2} \frac{\sigma_{max}(D)^2}{\sigma_*(D)^4} \ln\left(\frac{2k}{\delta}\right) \tag{6.31}$$

samples.

Let us consider a scheme where we estimate the error rates of each qubit $i$ separately from a local region, with measurements $\mathscr{M}_i$ and coefficient matrix $D_i$. In each region, the coefficient matrix is not full rank, but we assume that the parameters of qubit $i$ can be uniquely determined from region $i$ (see also the discussion in Appendix A.2). In this case we can apply Eq. (6.31) individually for the parameters of each qubit. Since the inverse Fourier transform is proportional to an isometry, we furthermore have $\|\hat{P}_i - P_i^*\|_2 = \frac{1}{2}\|\hat{\theta}_i - \theta_i^*\|_2$, where $\theta_i$ is the 4-component vector of moments for the $i$-th qubit. Furthermore, standard norm inequalities give $\|\hat{P}_i - P_i^*\|_1 \leq 2\|\hat{P}_i - P_i^*\|_2$. To bound the probability of the error being at most $\epsilon$ in all regions simultaneously, we can again use a union bound. All in all, this yields the following result:

**Lemma 27.** *Assume that both $y^* \geq \beta > 0$ and $\theta^* \geq \beta$ element-wise, and that there exists a lower bound $0 < \tilde{\beta} \leq \beta - 4\|y^* - \hat{y}\|_2$. For independent single-qubit noise, we*

*can estimate the marginal error distribution $P_i$ of each qubit $i$ simultaneously with error $\|\hat{P}_i - P_i^*\|_1 < \epsilon$ with probability $1-\delta$ from*

$$m \geq \frac{8k}{\tilde{\beta}^4 \epsilon^2} \frac{\sigma_{max}^2}{\sigma_*^4} \ln\left(\frac{2kn}{\delta}\right) \tag{6.32}$$

*syndrome measurements, where $k = \max_{i\in[n]} |\mathscr{M}_i|$, $\sigma_{max} = \max_{i\in[n]} \sigma_{max}(D_i)$ and $\sigma_* = \min_{i\in[n]} \sigma_*(D_i)$.*

We see that our algorithm exhibits the essentially optimal scaling $\mathcal{O}\left(\frac{1}{\epsilon^2}\log(\frac{1}{\delta})\right)$ for estimation with additive error. A proof that this scaling is optimal can e.g. be found in [AB99b, Lemma 5.1], where it is shown that estimating the bias of a coin requires at least this many samples, and estimating the bias of a coin can be seen as a special case of learning Pauli error rates.

The constants in this bound, however, merit some discussion. Notably, the sample complexity depends on a lower bound $\beta$ on the relevant moments. This is similar to the situation in randomized benchmarking for Pauli channels, discussed by Flammia and Wallman [FW20]. (In their work, the moments are called Pauli fidelities.) If we assume that the error rate on each qubit is lower bounded as $P_i(I) > \frac{1}{2} + \frac{\gamma}{2}$, the moments of each qubit are bounded from below by $\gamma$. Then, if the errors between qubits are indeed independent, the moment of a measurement with weight $w$ scales as $\gamma^w$. Thus, in principle $\beta$ could scale exponentially in the size of the code. However, as long as estimation is possible from stabilizers with constant weight, $\gamma^w$ is lower bounded by a constant and does not influence the scaling of the sample complexity. This is in particular the case if the estimation is possible from regions of bounded size, which we expect for topological codes. Still, in practice, it will be advantageous to find a system of equations for each region containing mostly low-weight measurements.

We further note that the bound given in Lemma 27 is similar to the sample complexity bound derived for the problem of factor graph learning by Abbeel et al. [AKN06]. This is not surprising, since we have seen that the estimation problem we consider is closely related to factor graphs. However, a fundamental difference is that we consider a convolutional factor graph, not a product factor graph. This introduces additional complications that are not present in [AKN06]. In particular, we have to consider how errors propagate through the Fourier transform. This is relatively easy for independent single-qubit noise, where we can Fourier transform on each qubit separately, and furthermore canonical and regular moments coincide. For more complicated noise models, a more detailed analysis of the interplay between the canonical factorization and the Fourier transform will be required. In particular, the factorization into canonical moments does not directly correspond to a factorization into a convolution of probability distributions, since

naively Fourier transforming individual canonical moments does not result in positive distributions. Thus, we first have to compute the regular moments, which can introduce additional error since the Möbius inversion might be ill-conditioned. A different perspective on the same problem is that the canonical moments are in principle unbounded, which complicates the error analysis and sample complexity bounds. Another fundamental difference to [AKN06] is that we cannot measure our variables directly, and instead only have access to the syndrome measurements. This manifests as a term $\frac{\sigma_{max}^2}{\sigma_*^4}$ in our bound (6.32), which is related to the conditioning of the coefficient matrix. Nevertheless, the strong analogy to factor graph learning suggests that estimating correlated Pauli noise will likely require an amount of samples that scales exponentially in the "correlation length", i.e. in the size of the largest factor. It furthermore provides an avenue to bound the error also in case of model mismatch, since it might be possible to transfer bounds from [AKN06] to our setting.

Finally, while the bound in Lemma 27 tells us which features need to be considered when designing an estimator or coefficient matrix in practice, no attempt was made to optimize the constants. It is possible that these could be further improved with a more detailed analysis.

### 6.3.3 Simulations

To understand if an advantage in decoding accuracy can be gained from the estimation of error channels using syndrome data, we now present some first simulations in a simple setting. More details about the procedures and simulation parameters can be found in Appendix A.3.

We consider a surface code with 2 rough and 2 smooth boundaries encoding $1$ logical qubit, as illustrated in Figure 6.2. We subject this code to phenomenological single-qubit Pauli noise, with possibly different Pauli error rates on each qubit. To model the variation between qubits, we use random $T1$- and $T2$-coherence times, with distributions roughly based on the data provided by Tannu and Qureshi [TQ19] for the IBM-Q20. We convert these $T1$- and $T2$-times into Pauli channels via the Pauli-twirl approximation for an amplitude-phase damping channel [Etx+21].

Our goal is to examine the effect of error rate information on the logical error rate of a decoder. To achieve this, we perform simulations that consist of the following steps. We first draw a random Pauli channel for each qubit, as described above. We then sample a dataset of $10^4$ syndromes by randomly sampling Pauli errors and computing their syndromes. From this, we obtain an estimate of the error rate of each qubit using a least-squares estimator as described in Section 6.3.1, with some slight improvements described in Appendix A.3. The estimated error rates are used

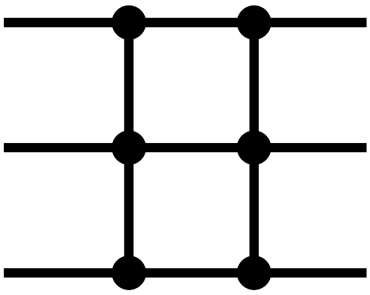


**Figure 6.2.:** A distance 3 surface code with 2 rough and 2 smooth boundaries. Edges represent qubits, vertices represent Pauli-$Z$ stabilizers acting on their neighboring qubits, and faces represent Pauli-$X$ stabilizers acting on their neighboring qubits. In total there are 13 qubits and 6 of each type of stabilizer.

as parameters of the tensor-network decoder developed by Chubb [Chu21]. This decoder implements a very good approximation of degenerate maximum-likelihood decoding for phenomenological noise, and is thus essentially optimal for our noise model. To obtain the approximate logical error rate of this decoder, we sample $10^4$ syndromes from the actual distribution, decode using the estimated error rates, and check how many logical errors occurred.

Results are shown in Figure 6.3, for different code distances. Each box-plot represents $80$ simulations. Note that the simulations contain two sources of randomness: randomly drawn error rates, and randomly drawn syndrome data. Thus, the boxes should not be interpreted as traditional error bars or confidence intervals, which quantify variance only over the random sampling of the dataset. We compare with two other methods to supply error rates to the decoder. One is using the exact error rates ("perfect knowledge"), which are only available in simulations. This gives a lower bound on the achievable logical error rate. The other is to always use the average error rates, which does not use any knowledge about the specific realization of the channel ("averaged channel"). Note that even the averaged channel still contains some information about the general noise level of each qubit, which a tensor network decoder can use. Thus, this is a somewhat stricter benchmark than e.g. a matching decoder with uniform weights, which uses no error rates information at all. We compute the averaged channel as an empirical average over $10^7$ realizations of the twirled Pauli channel. Note that the averaged channel can not be computed by first averaging the coherence times and then twirling, as done in [Etx+21].

We see that estimating the error rates from $10^4$ syndromes results in a clear improvement in logical error rate compared to decoding with the averaged channel. The improvement is higher for larger code distances, because there are more qubits and thus using approximate error rates for each qubit leads to larger errors. For $d = 13$, the improvement is by a factor of almost $2$.

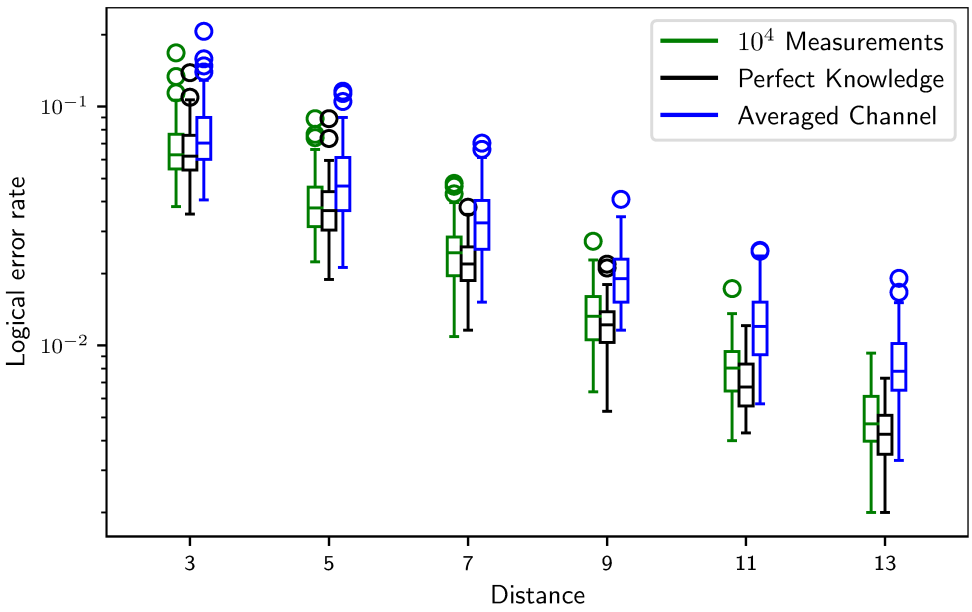



**Figure 6.3.:** Logical error rate of a tensor-network decoder based on either estimated, exact or averaged error rates. Each box-plot represents 80 simulations, as described in the main text. The boxes extend from the lower to the upper quartile of the data, with a line at the median. The whiskers extend to the last data point within 1.5 inter-quartile ranges from the box. Outliers beyond that are shown individually.

We suspect that estimation of error rates plays an even more important role when the effects of circuit-noise are taken into account, since currently error rates are dominated by multi-qubit gates (e.g. the CNOT-gate). The error rates of these gates also vary significantly, as shown in [TQ19]. However, we leave more comprehensive simulations for future research.

# 7 Conclusion

> *There is room for words on subjects, other than last words.*
>
> **— Robert Nozick**

Efficient device characterization is of fundamental importance for quantum computation. It is needed both to verify the correct working of a device, as well as for the calibration of its components. Detailed characterization of noise can be very helpful for QEC in particular, since both codes and decoders can be tailored accordingly.

In this thesis, we have considered the estimation of quantum noise from the syndrome measurements of a QEC code. From a characterization perspective, this efficiently uses already available data to obtain additional information about a device. From a QEC perspective, it is interesting to exploit information contained in the syndromes that is neglected in standard correction schemes. A possible application is the calibration of decoders without destructive measurements, perhaps even in the presence of time-varying noise.

Previous approaches to this estimation problem [Fow+14; HL17; Woo20; FB16; Fuj14a; Com+14; Spi+18] suffer from a lack of theoretical underpinning, limited applicability, and are often intractable for general codes and noise. We provide a stronger foundation for such schemes by developing a comprehensive theoretical framework for this estimation task. In particular, we precisely describe under which conditions syndrome data is sufficient to characterize the noise. Our framework encompasses both exact estimation of the physical channel, as well as estimation up to logical equivalence. The main result, given in Theorem 22 and Theorem 25, is that a channel is identifiable as long as it is correctable by the code. Thus, at least for phenomenological noise models, the situation is as good as one could reasonably hope. We complement these fundamental results with an efficient estimation algorithm. In contrast to many previous approaches, our algorithm neither relies on the assumption of vanishing error rates, nor does it require dealing with intractable likelihood-functions. Instead, it is based on a Fourier approach which greatly simplifies the estimation problem.

We hope that the theory we developed in this thesis can serve as a fundamental framework for further research. Naturally, there remain many interesting questions to be explored. First of all, our rigorous results were derived only for phenomeno-

logical noise models. While this might be a reasonable approximation for quantum communication or storage scenarios, other scenarios require the study of more realistic noise models. Of particular importance are the more fine-grained circuit-noise models used in the analysis of fault-tolerance. These introduce additional complications that are not readily dealt with in our framework. In particular, one has to take into account that the measurement circuits themselves introduce additional errors, and some of these occur to late to be reliably detected. Perhaps a treatment of such late errors following the analysis of Delfosse et al. [DRS22] might prove fruitful. One could also attempt to transfer our results via a mapping from circuit-noise to effective phenomenologial noise on an enlarged code, such as the ones given by Pryadko [Pry20] or Chubb and Flammia [CF21]. We note, however, that most decoders also do not take into account all the details of a full circuit-level model. The most common approximation is to use a decoding graph, and for these our methods straightforwardly apply (Appendix A.1). Ultimately, we expect that the applicability of our framework is coupled to the applicability of decoders, and the development of decoders using more fine-grained approximations will naturally yield new noise estimation schemes within these approximations. Whether a full circuit-level model can be estimated from syndrome data remains an interesting open question. In a similar vein, it would also be interesting to consider the estimation of more general quantum channels, beyond stochastic Pauli noise.

Within our existing framework, there are also several possible research directions. Our estimator is essentially based on convolutional factor graph learning [MK05; MKF04], which is currently not well understood. In particular, it would be fruitful to better characterize the interplay between the two main components of our framework, which are the canonical factorization and the Fourier transform. Naively Fourier transforming the canonical moments does not yield valid probability distributions, which means the canonical moments do not correspond to a factorization of the original distribution as a convolution. This complicates e.g. the analysis of error bounds, see also the discussion in Section 6.3.2. It would be interesting to develop methods that avoid this issue, or an improved canonical factorization into a proper convolution of probability distributions.

In conclusion, we hope that our results highlight the potential of using syndrome data for characterization, and provide a basis for the development of new approaches to both characterization and QEC.

# Bibliography


[AKN06] Pieter Abbeel, Daphne Koller, and Andrew Y. Ng. “Learning Factor Graphs in Polynomial Time and Sample Complexity”. In: *Journal of Machine Learning Research* 7.64 (2006), pp. 1743–1788. arXiv: 1207.1366 (cit. on pp. 32–34, 51, 62, 63).

[AB99a] Dorit Aharonov and Michael Ben-Or. “Fault-Tolerant Quantum Computation With Constant Error Rate”. June 1999. arXiv: quant-ph/9906129 (cit. on p. 1).

[Aig07] Martin Aigner. “A Course in Enumeration”. Springer-Verlag Berlin Heidelberg, 2007 (cit. on p. 50).

[AB99b] Martin Anthony and Peter L. Bartlett. “Neural Network Learning: Theoretical Foundations”. Cambridge University Press, 1999 (cit. on p. 62).

[ALB20] Alexei Ashikhmin, Ching-Yi Lai, and Todd A. Brun. “Quantum Data-Syndrome Codes”. In: *IEEE Journal on Selected Areas in Communications* 38.3 (2020), pp. 449–462. arXiv: 1907.01393 (cit. on p. 26).

[Bab+15] Zunaira Babar, Panagiotis Botsinis, Dimitrios Alanis, Soon X. Ng, and Lajos Hanzo. “Fifteen Years of Quantum LDPC Coding and Improved Decoding Strategies”. In: *IEEE Access* 3 (2015), pp. 2492–2519 (cit. on pp. 16, 22, 23, 35).

[Bac+17] Dave Bacon, Steven T. Flammia, Aram W. Harrow, and Jonathan Shi. “Sparse Quantum Codes From Quantum Circuits”. In: *IEEE Transactions on Information Theory* 63.4 (2017), pp. 2464–2479 (cit. on p. 24).

[BMv78] Elwyn R. Berlekamp, Robert J. McEliece, and Henk C. A. van Tilborg. “On the inherent intractability of certain coding problems (Corresp.)” In: *IEEE Transactions on Information Theory* 24.3 (1978), pp. 384–386 (cit. on pp. 20, 22).

[Bis06] Christopher M. Bishop. “Pattern Recognition and Machine Learning (Information Science and Statistics)”. Springer-Verlag Berlin Heidelberg, 2006 (cit. on pp. 32, 34, 42).

[Blu+17] Robin Blume-Kohout, John K. Gamble, Erik Nielsen, et al. “Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography”. In: *Nature Communications* 8.1 (Feb. 2017). arXiv: 1605.07674 (cit. on p. 1).

[BS00] J. Frédéric Bonnans and Alexander Shapiro. “Pertubation Analysis of Optimization Problems”. 1st ed. Springer New York, NY, 2000 (cit. on pp. 61, 83, 84, 88).

[BHT98] Gilles Brassard, Peter Hoyer, and Alain Tapp. “Quantum counting”. In: *Automata, Languages and Programming*. Springer Berlin Heidelberg, 1998, pp. 820–831. arXiv: quant-ph/9805082 (cit. on p. 1).

[Bra+13] Sergey Bravyi, Guillaume Duclos-Cianci, David Poulin, and Martin Suchara. “Subsystem surface codes with three-qubit check operators”. In: *Quantum Information & Computation* 13.11-12 (2013), pp. 963–985. arXiv: 1207.1443 (cit. on pp. 24, 57).

[BE21] Nikolas P. Breuckmann and Jens N. Eberhardt. “Quantum Low-Density Parity-Check Codes”. In: *PRX Quantum* 2.4 (Oct. 2021). arXiv: 2103.06309 (cit. on p. 1).

[CTV17] Earl T. Campbell, Barbara M. Terhal, and Christophe Vuillot. “Roads towards fault-tolerant universal quantum computation”. In: *Nature* 549.7671 (Sept. 2017), pp. 172–179. arXiv: 1612.07330 (cit. on p. 1).

[Cao+19] Yudong Cao, Jonathan Romero, Jonathan P. Olson, et al. “Quantum Chemistry in the Age of Quantum Computing”. In: *Chemical Reviews* 119.19 (2019), pp. 10856–10915 (cit. on p. 1).

[Che+22] Edward H. Chen, Theodore J. Yoder, Youngseok Kim, et al. “Calibrated Decoders for Experimental Quantum Error Correction”. In: *Physical Review Letters* 128.11 (Mar. 2022). arXiv: 2110.04285 (cit. on pp. 4, 5, 79).

[Che+21] Zijun Chen, Kevin J. Satzinger, Juan Atalaya, et al. “Exponential suppression of bit or phase errors with cyclic error correction”. In: *Nature* 595.7867 (July 2021), pp. 383–387. arXiv: 2102.06132 (cit. on p. 4).

[Chu21] Christopher T. Chubb. “General tensor network decoding of 2D Pauli codes”. 2021. arXiv: 2101.04125 (cit. on pp. 23, 64, 92).

[CF21] Christopher T. Chubb and Steven T. Flammia. “Statistical mechanical models for quantum codes with correlated noise”. In: *Annales de l’Institut Henri Poincaré D* 8.2 (May 2021), pp. 269–321. arXiv: 1809.10704 (cit. on pp. 24, 68).

[Com+14] Joshua Combes, Christopher Ferrie, Chris Cesare, et al. “In-situ characterization of quantum devices with error correction”. May 2014. arXiv: 1405.5656 (cit. on pp. 4, 67).

[DP18] Andrew S. Darmawan and David Poulin. “Linear-time general decoding algorithm for the surface code”. In: *Physical Review E* 97.5 (May 2018). arXiv: 1801.01879 (cit. on p. 23).

[DN21] Nicolas Delfosse and Naomi H. Nickerson. “Almost-linear time decoding algorithm for topological codes”. In: *Quantum* 5 (Dec. 2021), p. 595. arXiv: 1709.06218 (cit. on p. 22).

[DRS22] Nicolas Delfosse, Ben W. Reichardt, and Krysta M. Svore. “Beyond Single-Shot Fault-Tolerant Quantum Error Correction”. In: *IEEE Transactions on Information Theory* 68.1 (Jan. 2022), pp. 287–301. arXiv: 2002.05180 (cit. on pp. 26, 68).

[DZ20] Nicolas Delfosse and Gilles Zémor. “Linear-time maximum likelihood decoding of surface codes over the quantum erasure channel”. In: *Physical Review Research* 2 (3 July 2020), p. 033042. arXiv: 1703.01517 (cit. on p. 22).

[Del73] Philippe Delsarte. “Four fundamental parameters of a code and their combinatorial significance”. In: *Information and Control* 23.5 (1973), pp. 407–438 (cit. on p. 54).

[DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. “Maximum Likelihood from Incomplete Data via the EM Algorithm”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 39.1 (1977), pp. 1–38 (cit. on p. 36).

[Den+02] Eric Dennis, Alexei Kitaev, Andrew Landahl, and John Preskill. “Topological quantum memory”. In: *Journal of Mathematical Physics* 43.9 (Sept. 2002), pp. 4452–4505. arXiv: quant-ph/0110143 (cit. on pp. 21, 22).

[Edm65] Jack Edmonds. “Paths, Trees, and Flowers”. In: *Canadian Journal of Mathematics* 17 (1965), pp. 449–467 (cit. on p. 22).

[Eis+20] Jens Eisert, Dominik Hangleiter, Nathan Walk, et al. “Quantum certification and benchmarking”. In: *Nature Reviews Physics* 2.7 (June 2020), pp. 382–390. arXiv: 1910.06343 (cit. on p. 1).

[Etx+21] Josu Etxezarreta Martinez, Patricio Fuentes, Pedro Crespo, and Javier Garcia-Frias. “Time-varying quantum channel models for superconducting qubits”. In: *npj Quantum Information* 7.1 (July 2021), p. 115 (cit. on pp. 1, 63, 64, 90).

[Fey82] Richard P. Feynman. “Simulating physics with computers”. In: *International Journal of Theoretical Physics* 21.6 (June 1982), pp. 467–488 (cit. on p. 1).

[FO21] Steven T. Flammia and Ryan O’Donnell. “Pauli error estimation via Population Recovery”. In: *Quantum* 5 (Sept. 2021), p. 549. arXiv: 2105.02885 (cit. on p. 5).

[FW20] Steven T. Flammia and Joel J. Wallman. “Efficient Estimation of Pauli Channels”. In: *ACM Transactions on Quantum Computing* 1.1 (Dec. 2020). arXiv: 1907.12976 (cit. on pp. 5, 49, 62).

[FB16] Jan Florjanczyk and Todd A. Brun. “In-situ Adaptive Encoding for Asymmetric Quantum Error Correcting Codes”. Dec. 2016. arXiv: 1612.05823 (cit. on pp. 4, 67).

[Fow+12] Austin G. Fowler, Matteo Mariantoni, John M. Martinis, and Andrew N. Cleland. “Surface codes: Towards practical large-scale quantum computation”. In: *Phys. Rev. A* 86 (3 Sept. 2012), p. 032324 (cit. on p. 22).

[Fow+14] Austin G. Fowler, D. Sank, J. Kelly, R. Barends, and John M. Martinis. “Scalable extraction of error models from the output of error detection circuits”. May 2014. arXiv: 1405.1454 (cit. on pp. 3, 41, 67).

[Fuj14a] Yuichiro Fujiwara. “Instantaneous quantum channel estimation during quantum information processing”. May 2014. arXiv: 1405.6267 (cit. on pp. 4, 67).

[Fuj14b] Yuichiro Fujiwara. “Ability of stabilizer quantum error correction to protect itself from its own imperfection”. In: *Physical Review A* 90.6, 062304 (Dec. 2014), p. 062304. arXiv: 1409.2559 (cit. on pp. 26, 54).

[FH13] W. Fulton and J. Harris. “Representation Theory: A First Course”. Graduate Texts in Mathematics. Springer New York, 2013 (cit. on pp. 7, 17).

[Ghi+22] Biswash Ghimire, Thomas Wagner, Hermann Kampermann, and Dagmar Bruß. “Quantum Grid States and Hybrid Graphs”. 2022. arXiv: 2207.09826 (cit. on pp. 5, 149).

[Got97] Daniel Gottesman. “Stabilizer Codes and Quantum Error Correction”. 1997. arXiv: quant-ph/9705052 (cit. on p. 16).

[Got09] Daniel Gottesman. “An Introduction to Quantum Error Correction and Fault-Tolerant Quantum Computation”. Apr. 2009 (cit. on pp. 15, 19).

[Gri73] G. R. Grimmett. “A Theorem about Random Fields”. In: *Bulletin of the London Mathematical Society* 5.1 (1973), pp. 81–84 (cit. on p. 33).

[Gro96] Lov K. Grover. “A Fast Quantum Mechanical Algorithm for Database Search”. In: *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*. STOC ’96. Philadelphia, Pennsylvania, USA: Association for Computing Machinery, 1996, pp. 212–219. arXiv: quant-ph/9605043 (cit. on p. 1).

[HC71] J. M. Hammersley and P. Clifford. “Markov fields on finite graphs and lattices”. 1971 (cit. on p. 33).

[HFW20] Robin Harper, Steven T. Flammia, and Joel J. Wallman. “Efficient learning of quantum noise”. In: *Nature Physics* 16.12 (Dec. 2020), pp. 1184–1188. arXiv: 1907.13022 (cit. on p. 5).

[HYF21] Robin Harper, Wenjun Yu, and Steven T. Flammia. “Fast Estimation of Sparse Quantum Noise”. In: *PRX Quantum* 2 (1 Feb. 2021), p. 010322. arXiv: 2007.07901 (cit. on p. 5).

[HHL09] Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. “Quantum Algorithm for Linear Systems of Equations”. In: *Physical Review Letters* 103.15 (Oct. 2009). arXiv: 0811.3171 (cit. on p. 1).

[Hel+22] Jonas Helsen, Ingo Roth, Emilio Onorati, Albert H. Werner, and Jens Eisert. “General Framework for Randomized Benchmarking”. In: *PRX Quantum* 3.2 (June 2022). arXiv: 2010.07974 (cit. on p. 2).

[Hig21] Oscar Higgott. “PyMatching: A fast implementation of the minimum-weight perfect matching decoder”. 2021. arXiv: 2105.13082 (cit. on pp. 21, 22, 26, 79).

[HB21] Oscar Higgott and Nikolas P. Breuckmann. “Subsystem Codes with High Thresholds by Gauge Fixing and Reduced Qubit Overhead”. In: *Physical Review X* 11.3 (Aug. 2021). arXiv: 2010.09626 (cit. on pp. 22, 26, 79).

[Hoe63] Wassily Hoeffding. “Probability Inequalities for Sums of Bounded Random Variables”. In: *Journal of the American Statistical Association* 58.301 (1963), pp. 13–30 (cit. on p. 60).

[HNB20] Shilin Huang, Michael Newman, and Kenneth R. Brown. “Fault-tolerant weighted union-find decoding on the toric code”. In: *Physical Review A* 102.1 (July 2020). arXiv: 2004.04693 (cit. on p. 22).

[HL17] Ming-Xia Huo and Ying Li. “Learning time-dependent noise to reduce logical errors: real time error rate estimation in quantum error correction”. In: *New Journal of Physics* 19.12 (Dec. 2017), p. 123032. arXiv: 1710.03636 (cit. on pp. 3, 41, 67).

[IP15] Pavithran Iyer and David Poulin. “Hardness of Decoding Quantum Stabilizer Codes”. In: *IEEE Transactions on Information Theory* 61.9 (2015), pp. 5209–5223. arXiv: 1310.3235 (cit. on p. 21).

[JC11] Joao Jesus and Richard E. Chandler. “Estimating functions and the generalized method of moments”. In: *Interface Focus* 1.6 (Sept. 2011), pp. 871–885 (cit. on pp. 58, 91).

[KS22] Gleb Kalachev and Sergey Sadov. “A linear-algebraic and lattice-theoretical look at the Cleaning Lemma of quantum coding theory”. In: *Linear Algebra and its Applications* 649 (Sept. 2022), pp. 96–121. arXiv: 2204.04699 (cit. on pp. 7, 8, 57).

[KG15] Amara Katabarwa and Michael R. Geller. “Logical error rate in the Pauli twirling approximation”. In: *Scientific Reports* 5.1 (Sept. 2015), p. 14670 (cit. on p. 25).

[KR21] Martin Kliesch and Ingo Roth. “Theory of Quantum System Certification”. In: *PRX Quantum* 2.1 (Jan. 2021). arXiv: 2010.05925 (cit. on p. 1).

[Kni+08] E. Knill, D. Leibfried, R. Reichle, et al. “Randomized benchmarking of quantum gates”. In: *Physical Review A* 77.1 (Jan. 2008). arXiv: 0707.0963 (cit. on p. 2).

[KF09] Daphne Koller and Nir Friedman. “Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning”. The MIT Press, 2009 (cit. on pp. 31, 32, 34, 36, 37).

[Kri+22] Sebastian Krinner, Nathan Lacroix, Ants Remm, et al. “Realizing repeated quantum error correction in a distance-three surface code”. In: *Nature* 605.7911 (May 2022), pp. 669–674. arXiv: 2112.03708 (cit. on p. 4).

[Laf+96] Raymond Laflamme, Cesar Miquel, Juan Pablo Paz, and Wojciech Hubert Zurek. “Perfect Quantum Error Correcting Code”. In: *Phys. Rev. Lett.* 77 (1 July 1996), pp. 198–201. arXiv: quant-ph/9602019 (cit. on p. 42).

[LB13] Daniel A. Lidar and Todd A. Brun. “Quantum Error Correction”. Cambridge University Press, 2013 (cit. on pp. 15, 18, 19, 23).

[LP19] Ye-Hua Liu and David Poulin. “Neural Belief-Propagation Decoders for Quantum Error-Correcting Codes”. In: *Physical Review Letters* 122.20, 200501 (May 2019), p. 200501. arXiv: 1811.07835 (cit. on p. 23).

[Mag+13] Easwar Magesan, Daniel Puzzuoli, Christopher E. Granade, and David G. Cory. “Modeling quantum noise for efficient testing of fault-tolerant circuits”. In: *Physical Review A* 87.1 (Jan. 2013). arXiv: 1206.5407 (cit. on p. 25).

[MK05] Yongyi Mao and F.R. Kschischang. “On factor graphs and the Fourier transform”. In: *IEEE Transactions on Information Theory* 51.5 (2005), pp. 1635–1649 (cit. on pp. 7, 10–13, 32, 49, 68).

[MKF04] Yongyi Mao, Frank R Kschischang, and Brendan J Frey. “Convolutional factor graphs as probabilistic models”. In: *Proceedings of the 20th conference on Uncertainty in artificial intelligence*. 2004, pp. 374–381. arXiv: 1207.4136 (cit. on p. 68).

[Mar+20] Josu Etxezarreta Martinez, Patricio Fuentes, Pedro M. Crespo, and J. Garcia-Frias. “Approximating Decoherence Processes for the Design and Simulation of Quantum Error Correction Codes on Classical Computers”. In: *IEEE Access* 8 (2020), pp. 172623–172643 (cit. on p. 25).

[NH98] Radford M. Neal and Geoffrey E. Hinton. “A View of the Em Algorithm that Justifies Incremental, Sparse, and other Variants”. In: *Learning in Graphical Models*. Ed. by Michael I. Jordan. Dordrecht: Springer Netherlands, 1998, pp. 355–368 (cit. on p. 37).

[NC11] Michael A. Nielsen and Isaac L. Chuang. “Quantum Computation and Quantum Information: 10th Anniversary Edition”. 10th. USA: Cambridge University Press, 2011 (cit. on pp. 15–17, 24).

[Pel05] Alessandro Pelizzola. “Cluster variation method in statistical physics and probabilistic graphical models”. In: *Journal of Physics A: Mathematical and General* 38.33 (2005), R309 (cit. on p. 31).

[Pou05] David Poulin. “Stabilizer Formalism for Operator Quantum Error Correction”. In: *Physical Review Letters* 95 (23 Dec. 2005), p. 230504. arXiv: quant-ph/0508131 (cit. on p. 23).

[Pou06] David Poulin. “Optimal and efficient decoding of concatenated quantum block codes”. In: *Physical Review A* 74 (5 Nov. 2006), p. 052333. arXiv: quant-ph/0606126 (cit. on pp. 23, 40).

[PC08] David Poulin and Yeojin Chung. “On the Iterative Decoding of Sparse Quantum Codes”. In: *Quantum Information & Computation* 8.10 (Nov. 2008), pp. 987–1000. arXiv: 0801.1241 (cit. on p. 23).

[Pry20] Leonid P. Pryadko. “On maximum-likelihood decoding with circuit-level errors”. In: *Quantum* 4 (Aug. 2020), p. 304. arXiv: 1909.06732 (cit. on pp. 24, 68).

[RU08] Tom Richardson and Rüdiger Urbanke. “Modern Coding Theory”. Cambridge University Press, 2008 (cit. on pp. 31, 32).

[Rof+20] Joschka Roffe, David R. White, Simon Burton, and Earl Campbell. “Decoding across the quantum low-density parity-check code landscape”. In: *Physical Review Research* 2.4 (Dec. 2020). arXiv: 2005.07016 (cit. on p. 23).

[Rom06] Steven Roman. “Field Theory”. Springer, New York, 2006 (cit. on p. 50).

[Rud90] Walter Rudin. “Fourier Analysis on Groups”. John Wiley & Sons, Ltd, 1990 (cit. on p. 10).

[Sho94] Peter W. Shor. “Algorithms for quantum computation: discrete logarithms and factoring”. In: *Proceedings 35th Annual Symposium on Foundations of Computer Science*. 1994, pp. 124–134 (cit. on p. 1).

[Spi+18] Stephen T. Spitz, Brian Tarasinski, Carlo W. J. Beenakker, and Thomas E. O’Brien. “Adaptive Weight Estimator for Quantum Error Correction in a Time-Dependent Environment”. In: *Advanced Quantum Technologies* 1.1 (2018), p. 1870015. arXiv: 1712.02360 (cit. on pp. 4, 5, 26, 45, 46, 67, 79, 81).

[Ste96] Andrew Steane. “Multiple-particle interference and quantum error correction”. In: *Proceedings of the Royal Society London A* 452 (1996), pp. 2551–2577 (cit. on p. 81).

[TQ19] Swamit S. Tannu and Moinuddin K. Qureshi. “Not All Qubits Are Created Equal: A Case for Variability-Aware Policies for NISQ-Era Quantum Computers”. In: *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*. ASPLOS ’19. Providence, RI, USA: Association for Computing Machinery, 2019, pp. 987–999. arXiv: 1805.10224 (cit. on pp. 1, 63, 65, 91).

[Ter99] Audrey Terras. “Fourier Analysis on Finite Groups and Applications”. London Mathematical Society Student Texts. Cambridge University Press, 1999 (cit. on pp. 7, 9, 11).

[Vui+19] Christophe Vuillot, Lingling Lao, Ben Criger, et al. “Code deformation and lattice surgery are gauge fixing”. In: *New Journal of Physics* 21.3 (2019), p. 033028. arXiv: 1810.10037 (cit. on p. 24).

[Wag] Thomas Wagner. “MomentEstimator”. `https://github.com/TWagner2/MomentEstimator` (cit. on pp. 59, 92).

[Wag+21] Thomas Wagner, Hermann Kampermann, Dagmar Bruß, and Martin Kliesch. “Optimal noise estimation from syndrome statistics of quantum codes”. In: *Physical Review Research* 3 (1 Mar. 2021), p. 013292. arXiv: 2010.02243 (cit. on pp. 6, 39, 41–43, 95).

[Wag+22a] Thomas Wagner, Hermann Kampermann, Dagmar Bruß, and Martin Kliesch. “Pauli channels can be estimated from syndrome measurements in quantum error correction”. In: *Quantum* 6 (Sept. 2022), p. 809. arXiv: 2107.14252 (cit. on pp. 6, 39, 45–47, 53, 109, 133).

[Wag+22b] Thomas Wagner, Hermann Kampermann, Dagmar Bruß, and Martin Kliesch. “Learning logical quantum noise in quantum error correction”. Sept. 2022. arXiv: 2209.09267 (cit. on pp. 6, 12, 28, 39, 45, 47, 50, 51, 55, 57, 133).

[WJ08] Martin J. Wainwright and Michael I. Jordan. “Graphical Models, Exponential Families, and Variational Inference”. In: *Foundations and Trends in Machine Learning* 1.1–2 (2008), pp. 1–305 (cit. on p. 31).

[WE16] Joel J. Wallman and Joseph Emerson. “Noise tailoring for scalable quantum computation via randomized compiling”. In: *Physical Review A* 94 (5 Nov. 2016), p. 052325. arXiv: 1512.01098 (cit. on pp. 5, 25).

[WFH11] David S. Wang, Austin G. Fowler, and Lloyd C. L. Hollenberg. “Surface code quantum computing with error rates over 1%”. In: *Physical Review A* 83 (2 Feb. 2011), p. 020302 (cit. on pp. 26, 79).

[War+21] Matthew Ware, Guilhem Ribeill, Diego Ristè, et al. “Experimental Pauli-frame randomization on a superconducting qubit”. In: *Phys. Rev. A* 103 (4 Apr. 2021), p. 042604. arXiv: 1803.01818 (cit. on p. 5).

[Woo20] James R Wootton. “Benchmarking near-term devices with quantum error correction”. In: *Quantum Science and Technology* 5.4 (Aug. 2020), p. 044004 (cit. on pp. 3, 41, 67).

# Glossary

**QEC** quantum error correction

**PGM** probabilistic graphical model

**BP** belief-propagation

**EM** expectation-maximization

**HEM** hard-assignment expectation-maximization

**MSE** mean-squared error

**CRB** Cramér-Rao bound

**poset** partially ordered set

# Appendix A

## A.1 Example: Estimating Edge Weights for General Matching Decoders

In Chapter 6, we have developed a very general theory for the estimation of error rates from syndrome statistics of error correction codes. We now give another example, which is of practical relevance. It is a very direct generalization of the toric code case described in Section 6.1.

For the toric or surface code, each error affects at most two syndrome bits. Thus, the toric admits a standard minimum-weight matching decoder. For this setting, Spitz et al. [Spi+18] already give an explicit formula to estimate the weights of the edges from the syndrome statistics. However, for most codes, single errors affect more than two syndrome bits. In this case, one can consider a more general matching decoder, replacing the edges in the decoding graph with hyperedges. We now give a way to estimate these hyperedge weights from syndrome data. This problem was considered in the context of decoder calibration in [Che+22], but only approximate solutions were given. To the best of our knowledge, so far no exact solution for this case has been derived in the literature.

Explicitly, a general matching decoder computes a maximum-likelihood decoding by considering a simplified error model represented by a hypergraph. For examples of ordinary matching decoders see e.g. [Hig21; WFH11; HB21], and for a hypergraph version [Che+22]. Hyperedges represent sets of errors with the same syndrome, and have a weight that corresponds to the probability that an error from this set occurs. Vertices represents bits of the measured syndrome, and an error on an hyperedge flips all vertices contained in the hyperedge. Errors on different hyperedges are assumed to be independent. Usually, the weight of each hyperedge is obtained by summing over many fault mechanisms with the same syndrome, using the simplifying assumption that all these fault mechanisms are independent. We consider the case where the underlying noise model is unknown and want to instead estimate the weight of each edge from syndrome data.

To simplify notation, we map this problem on a classical error correction code with $n$ bits. Each hyperedge corresponds to one bit, and each vertex is a parity-check acting on the bits associated with the incident hyperedges. We denote the set of

vertices as $H$, since it corresponds to a set of parity-checks generating the dual code. We can now apply the general estimator from Section 6.3.1, in this case to a classical code. Because we only have single-bit errors, the regular and the canonical moments coincide.

More explicitly, for each bit $i$, we introduce a random variables $Z_i$, where $Z_i = +1$ if no error is present on bit $i$ and $Z_i = -1$ if there is an error. Because errors are independent, we see that the expectation of a measurement $s$ is given by

$$E(s) = \prod_{i \in \mathrm{supp}(s)} E(Z_i)\,. \tag{A.1}$$

Note that this holds for all elements of the dual code, not just for the generators. This is a specific instance of the general system (6.13).

For this instance, we can give an explicit solution which directly generalized the solution for the toric code. We construct this solution in terms of the set $H$ of parity-checks, which generates the dual code. Any subset $A \subseteq H$ defines an element $s_A = \prod_{g \in A} g$ of the dual code, i.e. a measurement. For a bit $i$, we denote with $\mathcal{N}(i)$ the set of parity-check generators acting on it,

$$\mathcal{N}(i) = \{g \in H \,:\, g_i \neq I\}\,. \tag{A.2}$$

Equivalently, we can view this as the syndrome of an error on bit $i$. The core of our solution is the following lemma.

**Lemma 28.** *For any subset $A \subseteq H$ we have*

$$\prod_{B \subseteq A} E(s_B)^{(-1)^{|B|+1}} = \prod_{i : A \subseteq \mathcal{N}(i)} E(Z_i)^{2^{|A|-1}}\,, \tag{A.3}$$

*Proof.* By inserting Eq. (A.1) we obtain

$$\begin{aligned}\prod_{B \subseteq A} E(s_B)^{(-1)^{|B|}} &= \prod_{B \subseteq A} \prod_{i \in \mathrm{supp}(s_B)} E(Z_i)^{(-1)^{|B|}} = \prod_{i \in [n]} \prod_{B \subseteq A \,:\, i \in \mathrm{supp}(s_B)} E(Z_i)^{(-1)^{|B|}} \\ &= \prod_{i \in [n]} E(Z_i)^{\sum_{B \subseteq A \,:\, i \in \mathrm{supp}(s_B)} (-1)^{|B|}}\,.\end{aligned}$$

For a given $i$, we can split $A$ into the two disjoint sets $A_i = A \cap \mathcal{N}(i)$ and $A_i^c = A \setminus A_i$. Then, we obtain

$$\sum_{B \subseteq A \,:\, i \in \mathrm{supp}(s_B)} (-1)^{|B|} = \sum_{B' \subseteq A_i \,:\, |B'| \text{ is odd}} (-1)^{|B|'} \sum_{C' \subseteq A_i^c} (-1)^{|C|'} = -2^{|A_i|-1} \sum_{C' \subseteq A_i^c} (-1)^{|C|'}\,. \tag{A.4}$$

Here, the first equality follows because $i \in \mathcal{N}(s_B)$ only if $B$ contains an odd number of elements from $A_i$, and the last equality follows by counting the number of odd subsets of $A_i$. The remaining sum can be evaluated by splitting the terms by weight,

$$\sum_{C' \subseteq A_i^c} (-1)^{|C|'} = \sum_{w=0}^{|A_i^c|} (-1)^w \binom{|A_i^c|}{w} = [|A_i^c| = 0] = [A = A_i] = [A \subseteq \mathcal{N}(i)] \ , \quad \text{(A.5)}$$

where the second equality is a well known property of binomial coefficients and the last steps are by definition of $A_i$ and $A_i^c$. Inserting this back into the previous equations proves the lemma. □

The right hand side of Eq. (A.3) is a product over all single-bit errors whose syndrome includes the given syndrome $A$. The left hand side is the product of odd subsets of $A$ divided by the product of even subsets of $A$. If the right hand side only contains one term, i.e. there is only one single-bit error whose syndrome includes $A$, we directly obtain an equation for this expectation value. If the right hand side contains multiple terms, we first compute all of them except one, using the other equations. Then, we solve for the remaining unknown. Thus, we can estimate the expectations for all bits by repeating the following steps, assuming that the code has at least distance 3:

1. Find a bit whose neighborhood is not contained in that of any bit whose expectation value has not yet been calculated. This is always possible, since otherwise there must exist two bits with the same neighborhood and thus with the same syndrome, and the code has at most distance 2.

2. Solve for the expectation of this bit using Eq. (A.3).

This is essentially a hierarchy of estimation steps: We start with bits which have a large neighborhood, and work our way down to bits with small neighborhoods. The estimation algorithm proposed by Spitz et al. [Spi+18] is a special case of this method, for the situation where every neighborhood is of size at most 2.

To give a simple example with larger neighborhoods, we can consider the well known Steane code [Ste96] subject to only Pauli-$X$ errors. This is not a circuit-noise model, but illustrates the use of Eq. (A.3). The Tanner graph of this code is shown in figure Figure A.1. It consists of 7 bits and 3 parity-check generators, labeled $s_1, s_2, s_3$. If errors on the bits are independent, we have the 7 equations:

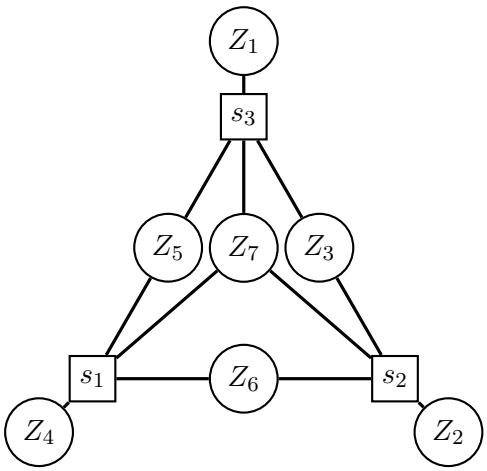



**Figure A.1.:** A representation of the Steane code with one type of errors. The circles correspond to bits, and the rectangles to parity-checks.

$$
\begin{aligned}
E(s_1) &= E(Z_4)E(Z_5)E(Z_7)E(Z_7) \\
E(s_2) &= E(Z_2)E(Z_3)E(Z_6)E(Z_7) \\
E(s_3) &= E(Z_1)E(Z_3)E(Z_5)E(Z_6) \\
E(s_1s_2) &= E(Z_2)E(Z_3)E(Z_4)E(Z_5) \\
E(s_1s_3) &= E(Z_1)E(Z_3)E(Z_4)E(Z_6) \\
E(s_2s_3) &= E(Z_1)E(Z_2)E(Z_5)E(Z_6) \\
E(s_1s_2s_3) &= E(Z_1)E(Z_2)E(Z_4)E(Z_7)\,.
\end{aligned}
$$

We see that $Z_7$ has the largest syndrome, as it neighbors all 3 generators. We can explicitly calculate

$$
E(Z_4)^4 = \frac{E(s_1)E(s_2)E(s_3)E(s_1s_3s_3)}{E(s_1s_2)E(s_1s_3)E(s_2s_3)}\,, \tag{A.6}
$$

which is one instance of Eq. (A.3). Next, we can calculate $E(Z_3), E(Z_5)$ and $E(Z_6)$ using the result for $E(Z_4)$. For example,

$$
E(Z_3)^2E(Z_4)^2 = \frac{E(s_2)E(s_3)}{E(s_2s_3)}\,. \tag{A.7}
$$

Finally, we can compute the remaining expectations $E(Z_1), E(Z_2)$ and $E(Z_4)$ using all previous results.

## A.2 Error Propagation

Here, we consider error propagation in the least-squares estimator of Section 6.3.2 and give a proof of Lemma 26. For ease of exposition, we first derive error bounds under the assumption that the coefficient matrix $D$ is full rank. We then explain how the arguments can be adapted if $D$ is rank-deficient. Let us quickly mention some notation: Scalar functions applied to vectors are always to be read element-wise, and similarly for comparisons between vectors: $a > b$ if $a_i > b_i$ for all $i$.

### A.2.1 Full Rank Coefficient Matrix

We start by introducing the log-space variables $x \coloneqq \ln(\theta)$. In these variables, the function $g(\theta) = \theta^D$ corresponds to $f(x) = \exp(Dx)$. For any $y \in \mathbb{R}^k$, we can consider the cost function

$$F_y = \frac{1}{2}\|f(x) - y\|_2^2 \,. \tag{A.8}$$

As in Section 6.3.2, we denote as $y^*$ the vector of ideal expectations of the measurements under the actual error distribution, and as $\hat{y}$ the vector of empirically measured expectations. Thus, the ideal cost function is $F_{y^*}$, and the cost function we actually optimize is $F_{\hat{y}}$. Since we assume that $D$ has full rank, $F_y$ has a unique minimizer for any $y \in \mathbb{R}^k$. We denote the minimizer of $F_{y^*}$ as $x^*$, and the minimizer of $F_{\hat{y}}$ as $\hat{x}$. Since $y^*$ corresponds to the ideal equation system, we assume that it has an exact solution fulfilling $f(x^*) = y^*$. However, the perturbed system might not have a solution, and thus $f(\hat{x}) \neq \hat{y}$ in general. Our goal is to bound the error $\|\hat{x} - x^*\|_2$ in terms of the error $\|\hat{y} - y^*\|_2$. To do this, we will view $F_{\hat{y}}$ as a perturbed version of $F_{y^*}$ and apply [BS00][Proposition 4.32]. In our setting, the proposition can be stated as follows.

**Lemma 29.** *Assume the following two conditions are fulfilled:*

- *There exists a neighborhood $\mathcal{N}$ of $x^*$ and a growth constant $\alpha > 0$, such that for all $x \in \mathcal{N}$ the following* second order growth condition *holds,*

$$F_{y*}(x) \geq F_{y^*}(x^*) + \alpha\|x - x^*\|_2^2 \,. \tag{A.9}$$

- *The difference function $G = F_{\hat{y}} - F_{y^*}$ is* Lipschitz continous *on $\mathcal{N}$ with Lipschitz constant $\kappa$.*

*Then, we have the following error bound for the minimizers $\hat{x}$ and $x^*$ of $F_{\hat{y}}$ and $F_{y^*}$,*

$$\|\hat{x} - x^*\|_2 \leq \frac{\kappa}{\alpha} \,. \tag{A.10}$$

For simplicity we make the somewhat idealized assumption that $\hat{x}$ minimizes the cost function exactly, but similar bounds can be derived if $\hat{x}$ is only an approximate minimizer [BS00][Proposition 4.32].

To apply Lemma 29, we need to find a suitable neighborhood $\mathcal{N}$ and compute the corresponding growth constant $\alpha$ and the Lipschitz constant $\kappa$. Since second order growth and Lipschitz continuity are related to second and first derivatives respectively, we first compute all relevant derivatives:

$$\frac{\partial f_k}{\partial x_i}(x) = D_{k,i} f_k(x)\,, \tag{A.11}$$

$$\frac{\partial^2 f_k}{\partial x_i \partial x_j} = D_{k,i} D_{k,j} f_k(x)\,, \tag{A.12}$$

$$\frac{\partial F_y}{\partial x_i} = \sum_k (f_k(x) - y_k) D_{k,i} f_k(x)\,, \tag{A.13}$$

$$\frac{\partial^2 F_y}{\partial x_i \partial x_j} = \sum_k D_{k,i} D_{k,j} f_k(x)(2 f_k(x) - y_k)\,. \tag{A.14}$$

For any vector $a \in \mathbb{R}^n$, we introduce the notation $\underline{a}$ for the $n \times n$ matrix with $a$ on the diagonal and $0$ everywhere else. With this notation, we can express the Jacobian $J_x$ at $x$ and the Hessian $H_x$ at $x$ of the functions more concisely as

$$J_x f = \underline{f(x)} D\,, \tag{A.15}$$

$$J_x F_y = (f(x) - y)^T \underline{f(x)} D\,, \tag{A.16}$$

$$H_x F_y = D^T A_y(x) D\,, \tag{A.17}$$

where we defined

$$A_y(x) \coloneqq \underline{f(x)}(2\underline{f(x)} - \underline{y})\,. \tag{A.18}$$

In particular, note that $H_x F_y$ is positive-definite if $A_y(x)$ is positive-definite.

Remember from Section 6.3.2 that we assume a lower bound $\beta$ on all moments, meaning that $y^* \geq \beta$ and $x^* \geq \ln(\beta)$. In particular, it suffices to optimize over the domain

$$\mathcal{D} = \{x \in \mathbb{R}^{4n} \,:\, \ln(\beta) \leq x \leq 0)\}. \tag{A.19}$$

We also assume that the error in $y^*$ is not too large compared to $y^*$,

$$\|\hat{y} - y^*\|_2 \leq \frac{\beta}{4}\,. \tag{A.20}$$

We now choose an appropriate region $\mathcal{N}$ for Lemma 29. With $\tau \coloneqq \|f(\hat{x}) - y^*\|_2$, we define

$$\mathcal{N} \coloneqq \{x \in \mathcal{D} \,:\, \|f(x) - y^*\|_2 \leq \tau\} = \{x \in \mathcal{D} \,:\, F_{y^*}(x) \leq 2\tau^2\}\,. \tag{A.21}$$

By definition, $\mathcal{N}$ must contain both $x^*$ and $\hat{x}$.

Let us first show that $\mathcal{N}$ is a convex set. We can bound $\tau$ by a simple triangle inequality,

$$\tau = \|f(\hat{x}) - y^*\|_2 \leq \|f(\hat{x}) - \hat{y}\|_2 + \|\hat{y} - y^*\|_2\,. \tag{A.22}$$

Furthermore, since $\hat{x}$ minimizes $F_{\hat{y}}$, we have

$$\|f(\hat{x}) - \hat{y}\|_2 \leq \|f(x^*) - \hat{y}\|_2 = \|y^* - \hat{y}\|_2\,. \tag{A.23}$$

This yields

$$\tau \leq 2\|\hat{y} - y^*\|\,. \tag{A.24}$$

By the bounded error assumption Eq. (A.20) and the lower bound $y^* \geq \beta$, we get $\tau \leq \frac{\beta}{2} \leq \frac{\min_i y_i^*}{2}$. In particular, $f(x) \geq \frac{y^*}{2}$ element-wise for all $x \in \mathcal{N}$. Then, the matrix $A_{y^*}(x)$ defined in Eq. (A.18) is positive-definite on $\mathcal{N}$, and it follows that $H_x F_{y^*}$ is also positive-definite on $\mathcal{N}$. Thus $F_{y^*}$ is convex on $\mathcal{N}$. On the other hand, since $\mathcal{N} = \{x \in \mathcal{D} \,:\, F_{y^*}(x) \leq 2\tau^2\}$, we see that $\mathcal{N}$ is a level-set of a convex function and thus itself convex.

Now, we can calculate the growth constant. We have already seen above that $F_{y^*}$ is convex on $\mathcal{N}$. To obtain a lower bound on the growth constant, we need to lower bound the Hessian, i.e. show strong convexity. We use the following well-known fact:

**Lemma 30.** *Let $g : \mathbb{R}^n \supseteq U \to \mathbb{R}$ be a function on an open convex set $U$. If $H_x g \geq cI$ for all $x \in U$, and the gradient of $g$ vanishes at $x_0 \in U$, then*

$$g(x) \geq g(x_0) + \frac{c}{2}\|x - x_0\|_2^2\,. \tag{A.25}$$

Note that the fact that $\mathcal{N}$ is closed, and not open, is not a problem, since it is a ball and thus we can consider its interior and extend the inequality to the boundary by continuity. Consequently, the growth constant $\alpha$ is bounded by $\frac{1}{2}$ times the smallest eigenvalue of $H_x F_{y^*}$ on $\mathcal{N}$. From Eq. (A.17), and since $A_{y^*}$ is positive-definite on $\mathcal{N}$, the smallest eigenvalue of $H_x F_{y^*}(x)$ is the smallest squared singular value $\sigma_{min}^2(\sqrt{A_{y^*}(x)}D)$ of $\sqrt{A_{y^*}(x)}D$. By the properties of singular values, we can bound this as

$$\sigma_{min}^2(\sqrt{A_{y^*}(x)}D) \geq \sigma_{min}^2(\sqrt{A_{y^*}(x)})\sigma_{min}^2(D)\,. \tag{A.26}$$

Furthermore, by the definition Eq. (A.18) of $A_{y^*}(x)$,

$$\sigma_{min}^2(\sqrt{A_{y^*}(x)}) \geq \min_{x\in\mathcal{N}} \min_i (2f_i(x) - y_i^*) f_i(x)\,. \tag{A.27}$$

By the definitions Eq. (A.21) of $\mathcal{N}$ and Eq. (A.22) of $\tau$, we have $f_i(x) \geq y_i^* - \tau$. Thus

$$\begin{aligned}
&\min_{x\in\mathcal{N}} \min_i (2f_i(x) - y_i^*) f_i(x) \\
&\geq \min_i (2(y_i^* - \tau) - y_i^*)(y_i^* - \tau) \\
&= \min_i (y_i^* - 2\tau)(y_i^* - \tau) \geq \min_i (y_i^* - 2\tau)^2 \geq (\beta - 2\tau)^2 \geq (\beta - 4\|\hat{y} - y^*\|_2)^2\,,
\end{aligned} \tag{A.28}$$

where the last inequality is due to Eq. (A.24). All in all, we obtain the bound

$$\alpha \geq \frac{(\beta - 4\|\hat{y} - y^*\|_2)^2 \sigma_{min}^2(D)}{2}\,. \tag{A.29}$$

It remains to bound the Lipschitz constant $\kappa$ of the difference function $G = F_{\hat{y}} - F_{y^*}$ on the region $\mathcal{N}$. From Eq. (A.16) we immediately obtain,

$$J_x G = (y^* - \hat{y})^T (\underline{f(x)} D)\,. \tag{A.30}$$

Since $N$ is convex, the Lipschitz constant $\kappa$ of $G$ on $\mathcal{N}$ is bounded by the maximal operator norm of $J_x G$,

$$\kappa \leq \max_{x\in\mathcal{N}} \|J_x G\|_2 \leq \max_{x\in\mathcal{N}} \|\underline{f(x)}\|_2 \|D\|_2 \|(y^* - \hat{y})^T\|_2\,, \tag{A.31}$$

where we used that the operator norm is sub-multiplicative. The operator norm of $D$ is its largest singular value $\sigma_{max}(D)$. Furthermore,

$$\|\underline{f(x)}\|_2 = max_i |f_i(x)| \leq 1\,, \tag{A.32}$$

where we used the fact that $x \leq 0$. For simplicity we use a relatively crude bound here, but one might be able to improve it slightly by exploiting the definition of $\mathcal{N}$. Finally, the operator norm of $(y^* - \hat{y})^T$ is simply the 2-norm of $(y^* - \hat{y})$. All in all we obtain,

$$\kappa \leq \sigma_{max}(D) \|y^* - \hat{y}\|_2\,. \tag{A.33}$$

Inserting the bounds Eq. (A.29) and Eq. (A.33) into Lemma 29 yields,

**Lemma 31.** *For $\beta \leq y^* \leq 0$ and $\ln(\beta) \leq x^* \leq 0$, and assuming $\|\hat{y} - y^*\|_2 \leq \frac{\beta}{4}$, the error $\|\hat{x} - x^*\|_2$ in the minimizers $\hat{x}$ and $x^*$ of $F_{\hat{y}}$ and $F_{y^*}$ is bounded as*

$$\|\hat{x} - x^*\|_2 \leq \frac{2}{(\beta - 4\|\hat{y} - y^*\|_2)^2} \frac{\sigma_{max}(D)}{\sigma_{min}(D)^2} \|\hat{y} - y^*\|_2 \,. \tag{A.34}$$

To convert this into a bound on the error in $\theta = e^x$, note that since $x < 0$, we have $\theta = e^x < 1$. Viewing $\theta$ as a function of $x$, we can bound the corresponding Lipschitz constant by $\|J_x \theta\|_2 = \|\underline{\theta(x)}\|_2 \leq 1$, where we used that the operator norm of a diagonal matrix is the absolute value of its largest entry. Thus,

$$\|\hat{\theta} - \theta^*\|_2 \leq \|\hat{x} - x^*\|_2 \,. \tag{A.35}$$

If we furthermore assume some absolute lower bound $\tilde{\beta} \leq \beta - 4\|\hat{y} - y^*\|$, we obtain the following lemma:

**Lemma 32.** *Let $\beta > 0$, $y^*$ be a vector with $\beta \leq y_i^* \leq 1$, and $D$ a coefficient matrix with entries $1$ and $0$ and full rank. Assume that the ideal minimization problem,*

$$\theta^* = \mathrm{argmin}_{\theta \,:\, \beta \leq \theta_i \leq 1} \|y^* - \theta^D\|_2 \,, \tag{A.36}$$

*has an exact solution $\theta^*$ fulfilling $y^* = (\theta^*)^D$. Let $0 < \tilde{\beta} \leq \beta$ and $\hat{y}$ be a vector fulfilling $0 < \tilde{\beta} \leq \beta - 4\|\hat{y} - y^*\|_2$. Then the solution $\hat{\theta}$ of the perturbed minimization problem,*

$$\hat{\theta} = \mathrm{argmin}_{\theta \,:\, \beta \leq \theta_i \leq 1} \|\hat{y} - \theta^D\|_2 \,, \tag{A.37}$$

*fulfills*

$$\|\hat{\theta} - \theta^*\|_2 \leq \frac{2}{\tilde{\beta}^2} \frac{\sigma_{max}(D)}{\sigma_{min}(D)^2} \|\hat{y} - y^*\|_2 \,. \tag{A.38}$$

### A.2.2 Rank Deficient Coefficient Matrix

As explained in Section 6.3.1, one might want to consider small regions of a code at a time, and only estimate some parameters in each region. For example, for independent single qubit noise one could estimate the channel for each qubit separately by considering only neighboring measurements and qubits. In this case, the coefficient matrix $D$ for a given region is not full rank, but we are only interested in estimating some of the parameters in this region. The remaining parameters might not be uniquely determined from the measurements of this region, and are instead estimated separately. We will call the first kind of parameters the "estimable parameters" of a region, and the second kind the "non-estimable parameters" of a region. When computing the least-squares solution, we still need to optimize over the non-estimable parameters to obtain a proper solution for the estimable

parameters, but we discard the non-estimable parameters after the optimization is finished. As before, we want to bound the error in our estimate by the error in the measured expectations, but this time we only need to consider the error in the estimable parameters.

From now on, we consider a fixed region with coefficient matrix $D$. Formally, the estimable parameters correspond to standard basis vectors in $\ker(D)^\perp$, while the non-estimable parameters correspond to all remaining standard basis vectors. Our method is to first obtain any solution for the full parameter vector of a region, and then project onto the estimable parameters. Therefore, we start by bounding the error in the full parameter vector. As in the previous section, our estimate $\hat{x}$ of the log-space parameters $x = \ln(\theta)$ is obtained as the minimizer of the cost function

$$F_{\hat{y}}(x) = \frac{1}{2}\|f(x) - \hat{y}\|_2^2\,, \tag{A.39}$$

with $f(x) = \exp(Dx)$. We view this as an approximation to the ideal cost function

$$F_{y^*} = \frac{1}{2}\|f(x) - y^*\|_2^2\,. \tag{A.40}$$

In contrast to the previous section, this problem does not have a unique solution, even for the ideal cost function defined by $y^*$. We denote the set of solutions of the ideal problem as $S^*$. We still assume that the ideal system $f(x) = y^*$ admits exact solutions, i.e. $F_{y^*}(x^*) = 0$ for all $x^* \in S^*$. Similar to the ideal cost function, the perturbed cost function $F_{\hat{y}}$ does not have a unique minimizer. We accept any minimizer $\hat{x}$ as a solution of the optimization problem. In this notation, our task is to bound

$$\operatorname{dist}(\hat{x}, S^*) = \min_{x^* \in S^*} \|\hat{x} - x^*\|_2\,. \tag{A.41}$$

Fortunately, we can again use [BS00][Proposition 4.32], but in a slightly more general version:

**Lemma 33.** *Assume the following two conditions are fulfilled:*

- *There exists a neighborhood $\mathcal{N}$ of $S^*$ and a growth constant $\alpha > 0$ such that for all $x \in \mathcal{N}$ the following* second order growth condition *holds,*

$$F_{y*}(x) \geq F^* + \alpha \operatorname{dist}(x, S^*)^2\,, \tag{A.42}$$

  *where $F^*$ is the minimal value of $F_{y^*}$.*

- *The difference function $G = F_{\hat{y}} - F_{y^*}$ is* Lipschitz continous *on $\mathcal{N}$ with Lipschitz constant $\kappa$.*

*Then, we have the following error bound for the minimizers $\hat{x}$ and $x^*$ of $F_{\hat{y}}$ and $F_{y^*}$,*

$$\|\hat{x} - x^*\|_2^2 \leq \frac{\kappa}{\alpha}\,. \tag{A.43}$$

As before, we define $\mathcal{D} = \{x \in \mathbb{R} \,:\, \ln(\beta) \leq x \leq 0)\}$, $\tau \coloneqq \|f(\hat{x}) - y^*\|_2$ and

$$\mathcal{N} \coloneqq \{x \in \mathcal{D} \,:\, F_{y^*}(x) \leq 2\tau^2\}\,. \tag{A.44}$$

Again $\mathcal{N}$ is a convex set.

The calculation of the growth constant $\alpha$ is similar to the previous section, but requires some conceptual modifications. Since we assume that the ideal problem can be solved exactly, i.e. we have $f(x^*) = y^*$ for all $x^* \in S^*$, the solution set $S^*$ is an affine space parallel to $\ker(D)$. We can consider the projection operator $\Pi_{S^*}$ onto $S^*$. Then, $\operatorname{dist}(x, S^*) = \|x - \Pi_{S^*}(x)\|_2$ for any $x \in \mathcal{N}$. Since the minimal value of $F_{y^*}$ is $0$, the second order growth condition can be written as

$$F_{y^*}(x) \geq \alpha \|x - \Pi_{S^*}(x)\|_2^2\,. \tag{A.45}$$

For any $x$, we can define the normal vector $h(x) \coloneqq \frac{x - \Pi_{S^*}(x)}{\|x - \Pi_{S^*}(x)\|_2}$. Since $S^*$ is an affine space parallel to $\ker(D)$, it holds $h(x) \perp \ker(D)$. To show Eq. (A.45), it is sufficient to show that the function

$$F_{y^*}(t) = F_{y^*}(\Pi_{S^*}(x) + h(x)t)\,, \tag{A.46}$$

is strongly convex on the interval $(0, \|x - \Pi_{y^*}(x)\|_2)$ for any $x \in \mathcal{N}$. Note that derivatives of $F_{y^*}(t)$ correspond to directional derivatives of $F_{y^*}(x)$.

From the above discussion, we conclude that in order to show the second order growth condition, it is sufficient to bound the second order directional derivatives of $F_{y^*}$ in directions $h$ orthogonal to $\ker(D)$. The second order directional derivative in direction $h$ is given by $h^T H_x F_{y^*} h$, and we have (Eq. (A.17))

$$H_x F_{y^*} = D^T A_{y^*}(x) D\,. \tag{A.47}$$

As in the previous section, we assume the error bound $\|\hat{y} - y^*\| \leq \frac{\beta}{4}$, which implies that $A_{y^*}(x)$ is positive-definite for all $x \in \mathcal{N}$. It follows that $\ker(H_x F_{y^*}) = \ker(D)$, independent of $x$. Thus, we have to bound the eigenvalues of $h^T H_x F_{y^*} h$ in directions $h \perp \ker(H_x F_{y^*})$. This corresponds to bounding the smallest non-zero eigenvalue,

which is the square of the smallest non-zero singular value $\sigma_*(\sqrt{A_{y^*(x)}}D)$, over $x \in \mathcal{N}$. Since $\sigma_*(\sqrt{A_{y^*(x)}}D) \geq \sigma_*(\sqrt{A_{y^*}}x)\sigma_*(D)$, we obtain

$$\alpha \geq \sigma_*^2(D) \min_{x \in \mathcal{N}} \sigma_{min}^2 \sqrt{A_{y^*}(x)}\,, \tag{A.48}$$

where we used the fact that $A_{y^*}(x)$ is positive-definite to replace $\sigma_*(\cdot)$ with $\sigma_{min}(\cdot)$. The second term was already bounded in Eq. (A.28) in the previous section. This concludes the calculation of the growth constant $\alpha$.

The calculation of the Lipschitz constant $\kappa$ is analogous to the previous section, and we obtain the same result as before. Thus, all in all, we obtain the same bounds on the error in the full parameter vector $x$ as in the previous section, but replacing the minimal singular value with the minimal non-zero singular value. Furthermore, projecting onto the estimable part of $x$ only decreases the error. We can propagate the error in the estimable part of $x$ through the exponential, as before, and obtain a bound on the error in $\theta$. Note that this last step uses the fact that the estimable part is unique. For the full parameter vector, the exponential will distort the solution manifold and the closest point in $S^*$ to a given point $x$ might change after applying the exponential. All in all, we obtain the following bound:

**Lemma 34.** *Let $\beta > 0$, $y^*$ be a vector with $\beta \leq y^* \leq 1$, and $D$ a possibly rank-deficient coefficient matrix with entries $1$ and $0$. Assume that the ideal minimization problem,*

$$\theta^* = \mathrm{argmin}_{\theta\,:\,\beta \leq \theta_i \leq 1} \|y^* - \theta^D\|_2\,, \tag{A.49}$$

*has an exact solution set $S^*$ fulfilling $y^* = (\theta^*)^D$ for all $\theta^* \in S^*$. Let $\hat{y}$ be a vector fulfilling $0 < \tilde{\beta} \leq \beta - 4\|\hat{y} - y^*\|_2$. Then any solution $\hat{\theta}$ of the perturbed minimization problem,*

$$\hat{\theta} = \mathrm{argmin}_{\theta\,:\,0 < \theta_i \leq 1} \|\hat{y} - \theta^D\|_2\,, \tag{A.50}$$

*fulfills*

$$\|\Pi_{\ker(D)^\perp}(\hat{\theta}) - \Pi_{\ker(D)^\perp}(\theta^*)\|_2 \leq \frac{2}{\tilde{\beta}^2} \frac{\sigma_{max}(D)}{\sigma_*(D)^2} \|\hat{y} - y^*\|_2\,, \tag{A.51}$$

*where $\Pi_{\ker(D)^\perp}$ is the projection onto $\ker(D)^\perp$.*

## A.3 Details of Simulations

In this appendix, we provide some more details about the simulations in Section 6.3.3.

First, let us describe exactly the distribution of error rates that was used. The procedure is based on [Etx+21]. For each qubit, times $T1$ and $T_\phi$ were drawn

from Gaussian distributions truncated at $0$. The means were set to $\mu_{T_1} = 80\mu s$ and $\mu_{T_\phi} = 57\mu s$, and the standard deviations to $\sigma_{T_1} = 35\mu s$ and $\sigma_{T_\phi} = 26\mu s$. These parameters were inspired by the distributions measured in [TQ19]. From $T_1$ and $T_\phi$, we calculate a $T_2$ time via

$$\frac{1}{T_2} = \frac{1}{2T_1} + \frac{1}{T_\phi}\,. \tag{A.52}$$

Then, the twirled Pauli-channel was computed, with error rates

$$\begin{aligned} p_I &= 1 - p_X - p_Z - p_Y\,, \\ p_X &= p_Y = \frac{1}{4}\left(1 - \exp(-\frac{t}{T_1})\right)\,, \\ p_Z &= \frac{1}{4}\left(1 + \exp(-\frac{t}{T_1} - 2\exp(-\frac{t}{T_2}))\right)\,. \end{aligned}$$

Here, we set $t = 5\mu s$.

The estimation procedure is based on the method of moments estimator described in Section 6.3.1. The theory of the generalized method of moments [JC11] however suggests a slight improvement, which we describe in the following.

We use the same notation as Appendix A.2: $\theta$ denotes the vector of parameters, i.e. the $3$ moments per qubit in a region, $y$ denotes the moments of the stabilizers in a region, and $g(\theta) = \theta^D$ with the coefficient matrix $D$. We use a hat for the empirically measured values, and a star for the actual values. For example, if we have $n$ measurements $s_1, \ldots, s_n$ of a stabilizer $s \in \mathscr{M}$, then $\hat{y}[s] = \frac{1}{n}\sum_{i=1}^{n} s_i$. On the other hand $y^*[s] = E(s)$, where $E$ denotes the expectation under the actual error distribution.

From [JC11], we see that instead of the cost function

$$F(\theta) = \frac{1}{2}\|\hat{y} - f(\theta)\|_2^2\,, \tag{A.53}$$

one should optimally consider the weighted cost function

$$F_W(\theta) = \frac{1}{2}(\hat{y} - f(\theta))^T W(\hat{y} - f(\theta))\,, \tag{A.54}$$

where $W$ is a weighting matrix. The asymptotically optimal estimator is obtained by setting $W = V^{-1}$ , where $V$ is the true covariance matrix of $y^*$, i.e. for $s, t \in \mathscr{M}$,

$$V[s,t] = E((s - E(s))(t - E(t)) = E(st) - E(s)E(t) = y^*[st] - y^*[s]y^*[t]\,. \tag{A.55}$$

We do not have direct access to $V$. Note however that if $y^*$ were known, all entries of $V$ could be computed from $y^*$. We can thus use the following iterative procedure. Start with the weight matrix $W_0 = I$, and minimize the cost function (A.54) to

obtain an estimate $\hat{\theta}_0$ of the parameters. Use this to compute $\hat{y}_0 = g(\hat{\theta}_0)$, and from this compute a new weighting matrix $W_1 = V_1^{-1}$ via Eq. (A.55). Then, minimize (A.54) with the new weights. In principle, this procedure can be repeated many times, but we finish after minimizing with $W_1$.

Finally, we also project all estimated moments to be larger than $\beta = 0.5$, and all measured moments to be larger than $\beta^w$, where $w$ is the stabilizer weight.

After estimation, the decoding was done using the tensor-network decoder presented in [Chu21]. This decoder has two approximation parameters, $\chi$ and $\tau$. We set $\chi = 20$ and $\tau = 60$.

Our implementation of the estimator can be found on Github [Wag]. The scripts that were used to start the specific simulations presented here are given in Appendix A.4.

## A.4 Code Listing

Here,we list the Julia scripts that were used for the simulations in Section 6.3.3. These scripts are to be used with the full code available on Github [Wag]. "Path-to-main" needs to be replaced with the path to the Main.jl file.

```
include("Path-to-main")
n_procs = Distributed.nprocs()
Filepath = ARGS[1]
n_simulation = parse(Int,ARGS[2])
l = parse(Int,ARGS[3])
project = parse(Bool,ARGS[4])
muT1 = parse(Float64, ARGS[5])
muTphi = parse(Float64,ARGS[6])
sigmaT1 = parse(Float64, ARGS[7])
sigmaTphi = parse(Float64, ARGS[8])
t = parse(Float64, ARGS[9])
beta=parse(Float64, ARGS[10])
n_step=parse(Int,ARGS[11])
SamplePerSimulation=parse(Bool,ARGS[12])
regularize=parse(Bool,ARGS[13])
fullcovariance=parse(Bool,ARGS[14])

n_test=parse(Int,ARGS[15])
χ=parse(Int,ARGS[16])
τ=parse(Int,ARGS[17])
Decode_actual =parse(Bool,ARGS[18])
n_estimate = [parse(Int,a) for a in ARGS[19:end]]

Debug=false

```

```
@info "Parameters" n_simulation l project muT1 muTphi sigmaT1
    sigmaTphi t n_estimate beta n_step SamplePerSimulation
    regularize fullcovariance χ τ Decode_actual

if Debug
    @everywhere global_logger(ConsoleLogger(stderr,Logging.Debug))
else
    @everywhere global_logger(ConsoleLogger(stderr,Logging.Info))
end

Code = qeccgraph_surfacecode_regular(l)
LocalEstimator = Estimator_lsq_optim(β=beta, n_step=n_step)
ChannelSampler = ChannelSampler_TVAPDTwirled(t=t,μ_T1 = muT1, μ
    _Tphi = muTphi, σ_T1 = sigmaT1, σ_Tphi = sigmaTphi,
    SamplePerSimulation = SamplePerSimulation)
if regularize
    #Estimate the mean and variance of the moments for given
        channel params and use it to regularize, in principle this
        could also be computed analytically by averaging the TVAPD(
        T1,T2) channel over T1,T2
    SampleP = sample_TVAPD_twirled(t,muT1, sigmaT1, muTphi,
        sigmaTphi, 10^5)
    SampleMom = momentsfromrates(SampleP)[2:end,:]
    Mean = dropdims(mean(SampleMom;dims=2);dims=2)
    if !fullcovariance
  Var = dropdims(var(SampleMom;dims=2);dims=2) #We could also
      estimate the full covariance matrix instead but it will be
      singular because it only has 2 parameters
  Regularizer = TiledRegularizer_L2_Repeating(Mean,(1 ./ Var))
    else
  Cov = cov(SampleMom')
        Regularizer = TiledRegularizer_L2_Repeating(Mean, pinv(Cov)
            )
    end
    @info "Regularizer Weight" Regularizer.Wt
else
    Regularizer = nothing
end

EstimatorParams = SimulationParameters_estimator(C = Code,
    f_neighborhops = surfacecode_hops, LocalEstimator=LocalEstimator
    ,ChannelSampler=ChannelSampler,Regularizer=Regularizer,
    n_simulations=n_simulation,n_estimate=n_estimate,project=project
    )
Params = SimulationParameters_EstimateAndDecode(EstimatorParams=
    EstimatorParams,n_test=n_test,χ=χ,τ=τ,Decode_actual=
    Decode_actual)

```

```
estimateanddecode_simulation(Filepath, Params)
```

**Listing A.1:** Script used for estimation and decoding

The simulations presented in Section 6.3.3 used the following parameters:

```
nsimulation=80
lvalues=(3 5 7 13 11 9)
project=1
muT1="80e-6"
muTphi="57e-6"
sigmaT1="35e-6"
sigmaTphi="26e-6"
t="5e-6"
beta="0.5"
nstep=2
SamplePerSimulation=1
regularize=0
fullcovariance=0
nestimate=(10000)

ntest=10000
chi=20
tau=60
decodeactual=1
```

**Listing A.2:** Parameters of the Simulations

This script was used to decode with the averaged channel, where the Filepath should point to the .hdf5 file generated from the estimation and decoding script:

```
include("Path-to-main")
n_procs = Distributed.nprocs()
Filepath = ARGS[1]
Debug=false

@info "Parameters" n_procs Filepath
if Debug
    @everywhere global_logger(ConsoleLogger(stderr,Logging.Debug))
else
    @everywhere global_logger(ConsoleLogger(stderr,Logging.Info))
end

decodewithaveragechannel_surfacecode_simulation(Filepath)
```

**Listing A.3:** Script used for decoding with averaged channel

# B

# Optimal Noise Estimation from Syndrome Statistics of Quantum Codes

| | |
|---|---|
| Title: | Optimal Noise Estimation<br>from Syndrome Statistics of Quantum Codes |
| Authors: | Thomas Wagner, Hermann Kampermann,<br>Dagmar Bruß and Martin Kliesch |
| Journal: | Physical Review Research |
| Publication status: | Published |
| Contribution by TW: | First author (input approx. 85%) |

This publication corresponds to reference [Wag+21]. A summary of its contents is presented in chapter 5.

The research objective was jointly devised by MK and me. The project was regularly discussed by all authors. I developed all analytical results and their proofs, with some suggestions by my co-authors. In particular, MK suggested to focus on *local* identifiability and to approach this problem using the inverse function theorem. The proofs were checked by all co-authors. I implemented all simulations and analyzed their results. I wrote the initial draft of the manuscript, with significant contributions by MK. The manuscript was then proofread and improved by all my co-authors.

# Optimal noise estimation from syndrome statistics of quantum codes

Thomas Wagner,* Hermann Kampermann, Dagmar Bruß, and Martin Kliesch
*Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany*

(Received 27 October 2020; accepted 3 March 2021; published 31 March 2021)

Quantum error correction allows to actively correct errors occurring in a quantum computation when the noise is weak enough. To make this error correction competitive information about the specific noise is required. Traditionally, this information is obtained by benchmarking the device before operation. We address the question of what can be learned from only the measurements done during decoding. Such estimation of noise models was proposed for surface codes, exploiting their special structure, and in the limit of low error rates, also for other codes. However, so far it has been unclear under what general conditions noise models can be estimated from the syndrome measurements. In this work, we derive a general condition for identifiability of the error rates. For general stabilizer codes, we prove identifiability under the assumption that the rates are small enough. Without this assumption, we prove a result for perfect codes. Finally, we propose a practical estimation method with linear runtime for concatenated codes. We demonstrate that it outperforms other recently proposed methods and that the estimation is optimal in the sense that it reaches the Cramér-Rao bound. Our method paves the way for practical calibration of error corrected quantum devices during operation.

DOI: 10.1103/PhysRevResearch.3.013292

## I. INTRODUCTION

Quantum error correction is an essential ingredient in quantum computing schemes. When employing active quantum error correction via stabilizer codes, the decoding can be significantly improved if information about the error rates of all qubits is available. In contrast to traditional benchmarking before operation, a new approach is to estimate error rates online from the syndrome statistics of the code itself [1–7]. It should be stressed that the syndrome statistics is the only information that can be measured without destroying the encoded information. As pointed out by Fowler *et al.* [2], this results in a noise model that is directly applicable for the decoder. Furthermore, it allows for the tracking of time-varying error rates [3,6]. Experimentally, online optimization of control parameters in a nine-qubit superconducting quantum processor has been demonstrated in a Google experiment [8].

However, apart from the work of Spitz *et al.* [6], there has been very little theoretical investigation of the estimation problem, see Sec. IV A for a detailed discussion. For example, it is not clear for what combinations of noise models and codes the unknown parameters are identifiable from the syndrome statistics. Evidently, some restrictions must apply since estimating completely general noise would require measurements which destroy the logical state. For some codes and noise models, including surface codes with independent Pauli noise on each qubit, the analytical method developed by [6] proves parameter identifiability. On the other hand, for many other important codes such as the five-qubit code [9], the Steane code [10], and more general color codes [11], this method is not applicable.

In this work, we address this question by deriving a general condition for parameter identifiability, and using it to explicitly prove results for the five-qubit code and the Steane code. Furthermore, we introduce an explicit error rates estimator, similar to techniques employed in classical distributed source coding [12], for concatenated codes and simulate it on the concatenated five-qubit code. This estimator outperforms previously proposed methods [2–4] in this setting, because it does not require the assumption of very low error rates.

### Stabilizer codes

Let us introduce our notation while briefly summarizing stabilizer codes. The Pauli group $\mathcal{P}_n$ on $n$ qubits is the group of *Pauli strings* generated by the Pauli operators $\{X, Y, Z, I\}$ with phases,

$$\mathcal{P}_n = \left\{ \epsilon \bigotimes_{i=1}^{n} e_i \mid \epsilon \in \{\pm 1, \pm \mathrm{i}\}, e_i \in \{I, X, Y, Z\} \right\}. \quad (1)$$

The Pauli group modulo phases,

$$\mathsf{P}_n = \mathcal{P}_n / \{\pm 1, \pm \mathrm{i}\}, \quad (2)$$

is called the *effective Pauli group*. We denote the $i$th tensor factor of $\boldsymbol{e} \in \mathsf{P}_n$ as $e_i$. The Pauli operator acting as $e \in \mathsf{P}_1$ on qubit $i$ and as the identity elsewhere is denoted $\boldsymbol{e}^{(i)} \in \mathsf{P}_n$. A stabilizer code encoding $k = n - l$ qubits is defined by a commutative subgroup $\mathcal{S}$ of $\mathcal{P}_n$ with generators $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_l$

*thomas.wagner@uni-duesseldorf.de

[13]. The code space is the simultaneous +1 eigenspace of the generators. Phases are generally not important for quantum error correction, so we consider data errors as elements of the effective Pauli group. For an error $\boldsymbol{e} \in \mathsf{P}_n$, we define the *syndrome* $\boldsymbol{S}(\boldsymbol{e}) \in \mathbb{F}_2^l$ entrywise by

$$\boldsymbol{S}(\boldsymbol{e})_i := \begin{cases} 0, & \text{if } \boldsymbol{g}_i \text{ and } \boldsymbol{e} \text{ commute in } \mathcal{P}_n, \\ 1, & \text{if } \boldsymbol{g}_i \text{ and } \boldsymbol{e} \text{ anticommute in } \mathcal{P}_n. \end{cases} \tag{3}$$

To correct an error $\boldsymbol{e} \in \mathsf{P}_n$, a recovery $\boldsymbol{r} \in \mathsf{P}_n$ is applied based on the measured syndrome. Since errors that only differ by stabilizers act equivalently on the encoded information, the recovery is successful if the equivalence class $[\boldsymbol{er}]$ is trivial, i.e., $[\boldsymbol{er}] = [I] \in \mathsf{P}_n/\mathcal{S}$.

## II. IDENTIFIABILITY CONDITIONS

We consider a stabilizer code with $n$ qubits and $l$ stabilizer generators. Let us first define identifiability. Given is a parameterized noise model, mapping a vector of error rates $\boldsymbol{\theta}$ to a vector $(\mathbb{P}[\boldsymbol{E}]_{\boldsymbol{E}\in\mathsf{P}_n})$ specifying the probability $\mathbb{P}[\boldsymbol{E}]$ for each error $\boldsymbol{E} \in \mathsf{P}_n$. This induces the map $\mathcal{M} : \boldsymbol{\theta} \mapsto (\mathbb{P}[\boldsymbol{S}])_{\boldsymbol{S}\in\mathbb{F}_2^l}$, mapping a parameter vector to the corresponding syndrome statistics via

$$\mathbb{P}[\boldsymbol{S}] = \sum_{\boldsymbol{E}\in\mathsf{P}_n : \boldsymbol{S}(\boldsymbol{E})=\boldsymbol{S}} \mathbb{P}[\boldsymbol{E}], \tag{4}$$

where $\mathbb{P}[\boldsymbol{S}]$ is the induced probability of observing the syndrome $\boldsymbol{S} \in \mathbb{F}_2^l$. Error rates are identifiable from the syndrome statistics if the map $\mathcal{M}$ is injective. This will usually not be the case, due to symmetry around error rates of 0.5. However, we can still hope that the parameters are at least identifiable if we restrict to some region in the space of parameters $\boldsymbol{\theta}$.

*Definition 1 (Local identifiability).* We say that error rates are *locally identifiable* at $\boldsymbol{\theta}$ if there exists $\varepsilon > 0$ such that the map $\mathcal{M}$ is injective on the ball $B_\varepsilon(\boldsymbol{\theta}) := \{\boldsymbol{\theta}' \mid \|\boldsymbol{\theta}' - \boldsymbol{\theta}\|_2 < \varepsilon\}$.

For ease of exposition, we will focus in this section on independent single qubit Pauli noise, which is a simple but widely studied error model. A substantial generalization of proposition 2 and theorem 3 to much more general error models, including measurement errors, can be found in Sec. IV B. For now let us assume that errors on the $i$th qubit of the code are modeled by the Pauli channel

$$\rho \mapsto \left(1 - \theta_X^i - \theta_Y^i - \theta_Z^i\right)\rho + \theta_X^i X\rho X + \theta_Y^i Y\rho Y + \theta_Z^i Z\rho Z. \tag{5}$$

with $\theta_X^i, \theta_Z^i, \theta_Y^i \in [0, 1]$ such that $\theta_X^i + \theta_Z^i + \theta_Y^i \leqslant 1$. The parameter vector $\boldsymbol{\theta}$ for this error model is given by the error rates $(\theta_e^i)_{i\in\{1,\ldots,n\}, e\in\{X,Y,Z\}}$ of all nontrivial single qubit errors. By the inverse function theorem, local identifiability at $\boldsymbol{\theta}$ holds if and only if the Jacobian matrix $J = D_{\boldsymbol{\theta}}\mathcal{M}$ at $\boldsymbol{\theta}$ has full (column) rank. We will label the rows of the Jacobian by syndromes $\boldsymbol{S}$ and the columns by parameters $\theta_e^i$, and denote entries with square brackets, e.g. as $J[\boldsymbol{S}, \theta_e^i]$. In the limit of low error rates, it is intuitive that identification of error rates is possible since a syndrome always arises from the matching single qubit error, and no combined errors occur. Thus the only requirement is that the single errors can be identified from the syndromes. This just means that the code has distance at least 3, i.e., only trivial codes are excluded. This leads to the estimators proposed in Refs. [2–4]. We confirm this intuition by calculating the Jacobian of $\mathcal{M}$ and checking its rank:

*Proposition 2 (Identifiability for small error rates).* For a quantum code subject to independent single qubit Pauli noise, error rates are locally identifiable at $\boldsymbol{\theta} = \boldsymbol{0}$ if and only if $\boldsymbol{S}(\boldsymbol{e}) \neq \boldsymbol{S}(\boldsymbol{e}')$ for every choice of two different single qubit errors $\boldsymbol{e}, \boldsymbol{e}'$.

A proof is provided in Sec. IV B. Our first central result is an identifiability condition without the assumption of low rates. This establishes a connection between local identifiability and the posterior distribution of errors for each qubit.

*Theorem 3 (General identifiability condition).* Consider a quantum code subject to independent single qubit Pauli noise. Assume that all error rates are nonzero and that $\mathbb{P}[\boldsymbol{S}] > 0$ for all syndromes $\boldsymbol{S} \in \mathbb{F}_2^l$. Then error rates are locally identifiable at $\boldsymbol{\theta}$ if and only if the matrix $\tilde{J}$ with entries

$$\tilde{J}\left[\boldsymbol{S}, \theta_e^i\right] = \frac{\mathbb{P}[E_i = e|\boldsymbol{S}]}{\mathbb{P}[E_i = e]} - \frac{\mathbb{P}[E_i = I|\boldsymbol{S}]}{\mathbb{P}[E_i = I]} \tag{6}$$

has full column rank. Here, $\mathbb{P}[E_i = e|\boldsymbol{S}]$ is the conditional probability that the $i$th qubit is affected by the error $e \in \mathsf{P}_1$ given that the observed syndrome is $\mathbb{P}[\boldsymbol{S}]$.

The proof is provided in Sec. IV B.

### Identifiability for perfect codes

We demonstrate the analytical application of theorem 3 by considering the class of *perfect codes*.

*Definition 4* (*Perfect single error correcting quantum code* [14]). A quantum code $C$ on $n$ qubits is called a *perfect single error correcting* code if there is a bijection between the set of nontrivial single qubit errors and the set of nontrivial syndromes, i.e., there exists a bijective map

$$f : \{\boldsymbol{e}^{(i)} \mid e \in \{X, Y, Z\},\ i \in \{1, \ldots, n\}\} \to \mathbb{F}_2^l \setminus \{\boldsymbol{0}\}. \tag{7}$$

These codes are called "perfect" because they saturate the (quantum) Hamming bound. A well known example of such a code is the five-qubit code [9]. Other families of perfect codes are cyclic Hamming codes [15] and a class of twisted codes [16]. The main result of this section is that error rates for such codes are locally identifiable around the points of equal rates, even for high error rates. This provides another concrete class of codes where identification of error rates is possible.

*Theorem 5 (Identifiability for perfect codes).* Let $C$ be a perfect single error correcting quantum code on $n$ qubits subject to independent single qubit Pauli noise. Then the error rates are locally identifiable around any point $\boldsymbol{\theta}$ with equal error rates, i.e., if there exists $p \in (0, 1)$ such that $\theta_e^i = p$ for all $i$ and all $e \in \{X, Y, Z\}$.

Note that the condition on $\boldsymbol{\theta}$ above does *not* mean that we restrict ourselves to a simple single parameter model. We still allow all estimated error rates to vary individually, but require that the actual error rates are close to being equal. In order to prove theorem 5 via theorem 3, we have to check the rank of the matrix $\tilde{J}$ given in (6). Using Bayes theorem, we can express its entries as

$$\tilde{J}\left[\boldsymbol{S}, \theta_e^i\right] = \frac{\mathbb{P}[\boldsymbol{S} \mid E_i = e] - \mathbb{P}[\boldsymbol{S} \mid E_i = I]}{\mathbb{P}[\boldsymbol{S}]}. \tag{8}$$

The key insight, which might be of independent interest, is that most of the conditional probabilities in this expression are equal.

*Lemma 6.* Consider a perfect single error correcting code on $n$ qubits subject to independent single qubit noise where all error rates are equal. Let $e, e' \in \mathsf{P}_1$. Then for any syndrome $\boldsymbol{S} \in \mathbb{F}_2^l \setminus \{\mathbf{0}\}$ and qubit $i$ such that $\boldsymbol{S} \neq \boldsymbol{S}(\boldsymbol{e}^{(i)})$ and $\boldsymbol{S} \neq \boldsymbol{S}((\boldsymbol{e}')^{(i)})$, we have

$$\mathbb{P}[\boldsymbol{S} \mid E_i = e] = \mathbb{P}[\boldsymbol{S} \mid E_i = e']. \tag{9}$$

The proof is provided in Sec. IV B. This lemma immediately implies that if $\boldsymbol{S} \neq \mathbf{0}$ and $\boldsymbol{S} \neq \boldsymbol{S}(e_i)$, then $\tilde{J}[\boldsymbol{S}, \theta_e^i] = 0$. We can ignore the case $\boldsymbol{S} = \mathbf{0}$ due to normalization, and in the case $\boldsymbol{S} = \boldsymbol{S}(e_i)$ we have $\tilde{J}[\boldsymbol{S}, \theta_e^i] \neq 0$. Thus the columns of $\tilde{J}$ are linearly independent unit vectors, i.e., $\tilde{J}$ has full rank. This proves theorem 5.

The arguments of the proof straightforwardly generalize to other noise models as long as the perfect code condition is fulfilled, i.e., there is a bijection between syndromes and elementary errors. For example one could consider simple noise models where Pauli $X$ and Pauli $Z$ errors occur independently. The rates of such a model are locally identifiable on the Steane code around points of equal rates, since the Steane code reduces to the classical Hamming code when only one type of errors is considered. The Hamming code is known to be a perfect code. Estimation of such a model on the Steane code was considered in Ref. [3]. Thus theorem 5 also provides a theoretical background for the results presented there.

## III. NUMERICAL ESTIMATION METHOD

In this section, we complement the previous results with a practical estimation method, which is based on the combination of belief propagation (BP) and expectation maximization (EM). In the limit of low error rates, methods based on "hard assignments" were proposed independently by [2–4]. They use either the recovery output by a ("hard") decoder or the lowest weight error corresponding to a syndrome. Inspired by techniques from classical distributed source coding [12], we instead consider an estimation method that uses the full information about the distribution of errors given a certain syndrome, by combining a "soft" decoder [17] with the expectation maximization algorithm [18,19].

### A. Belief propagation

Let us briefly summarize concatenated codes and their maximum-likelihood decoding [17]. We consider independent single qubit Pauli errors. A concatenated quantum code is obtained by encoding each qubit of a quantum code again in the same code. This defines a tree structure, where the logical qubit of a code block is a "physical qubit" in the next layer, as illustrated in Fig. 1 for the five-qubit code. We can view this as a graphical representation of the probability distribution over all possible errors given the measured syndrome, called a *factor graph*. The root node represents the total logical error. Maximum-likelihood decoding is done by computing its marginal distribution to find the most likely logical operator. Computation of marginal probabilities is efficiently possible using the BP algorithm (see e.g., Ref. [20]). BP works by passing messages along the edges of the graph. To compute the marginal of the root node it suffices to pass messages upwards, starting from the leaves. Doing an additional downwards pass, we can also calculate the marginals of the leaf nodes, i.e., the distribution of errors on a qubit given the measured syndrome. The computational effort of this method scales linearly in the number of qubits.

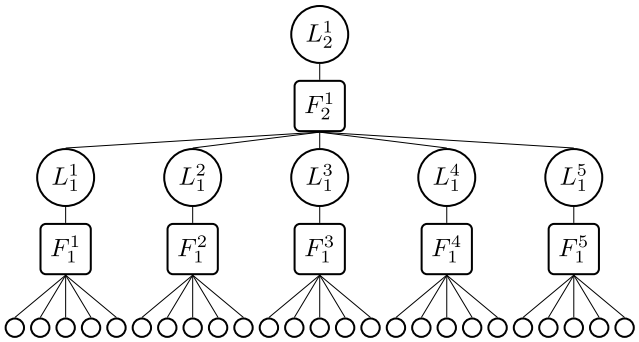



FIG. 1. The factor graph representation of the two times concatenated five-qubit code. The circles depict variable nodes representing (logical) errors, i.e., each variable takes values in $\mathsf{P}_1$. The squares depict factor nodes representing the stabilizer checks.

### B. Expectation maximization

Starting from an initialization $\boldsymbol{\theta}^{(0)}$ of the estimated error rates and given a set $D$ of measured syndromes, we can calculate a new estimate of the error rates using the EM algorithm, i.e., iterating the following steps until convergence.

(1) Expectation step. Compute the *expected sufficient statistics*

$$M_E(e_i|\boldsymbol{\theta}^{(k)}) = \sum_{\boldsymbol{S}\in D} \mathbb{P}[E_i = e|\boldsymbol{S}, \boldsymbol{\theta}^{(k)}],$$

based on the current estimate $\boldsymbol{\theta}^{(k)}$ of the error rates.

(2) Maximization step. Compute a new estimate $\boldsymbol{\theta}^{(k+1)}$ of the error rates by normalizing the expected sufficient statistics:

$$(\theta^{(k+1)})_e^i = \frac{M_E[e_i|\boldsymbol{\theta}^{(k)}]}{\sum_{e'\in\mathsf{P}_1} M_E\left[e_i'|\boldsymbol{\theta}^{(k)}\right]}. \tag{10}$$

Computationally, the main effort is in calculating the conditional probabilities needed for the expectation step. The key point is that this can be done efficiently using BP. In an online estimation setting, the first iteration of EM introduces almost no overhead, since the marginals calculated during decoding can be used. Further iterations require redecoding of the syndromes and are thus roughly as expensive as decoding. We will also compare our estimator with the "hard assignment" method [2–4], which is the best known scalable method. We extend this method slightly by allowing for multiple iterations. It can then be expressed as a variant of EM, called hard assignment expectation maximization (HEM) (see Ref. [19]). It consists of iterating the steps:

(1) For each syndrome $\boldsymbol{S} \in D$, compute the most likely error

$$\boldsymbol{E}_{\mathrm{map}}(\boldsymbol{S}) = \arg\max_{\boldsymbol{E}\in\mathsf{P}_n}(\mathbb{P}[\boldsymbol{E}|\boldsymbol{S}, \boldsymbol{\theta}^{(k)}]).$$

(2) Obtain the new error rates by counting how often each single qubit error appears:

$$(\theta^{(k+1)})^i_e = \frac{\sum_{\boldsymbol{S}\in D} \delta_{E_{\mathrm{map}}(\boldsymbol{S})_i,e}}{|D|}.$$

Here, $\delta$ is the Kronecker-delta.

Instead of the marginals, only the most likely error for each syndrome is considered. It can be computed efficiently using the max-sum algorithm which works similar to BP, see e.g., Ref. [20].

### C. Numerical results

In the following, we present a numerical comparison of our estimator (EM) and the "hard assignment" estimator (HEM). In light of our previous identifiability results, we consider the five-qubit code, concatenated with itself, subject to independent depolarizing noise with error rate $p$ on each qubit. Extending the method to a phenomenological noise model with measurement errors is straightforward, and some results are shown in Sec. IV C. We initialize the algorithm randomly around the actual rates, with a precision controlled by a real parameter $\alpha$ (higher is more accurate). To be precise, for each qubit $i$, we sample error rates $\boldsymbol{\theta}^i$ from a Dirichlet distribution

$$\mathbb{P}[\boldsymbol{\theta}^i] = \frac{1}{B(\boldsymbol{\alpha})} \prod_{e\in \mathsf{P}_1} \left(\theta^i_e\right)^{\alpha_e}, \tag{11}$$

where $\alpha_I = (1-3p)\alpha$, $\alpha_X = \alpha_Y = \alpha_Z = p\alpha$ and $B(\alpha)$ is a normalization constant. Such an initialization could be obtained from previous benchmarking or an educated guess. We then run the estimator for $n_{\mathrm{it}}$ iterations on a data set of $n_{\mathrm{est}}$ syndromes generated from the actual distribution. Using a fixed initialization and random actual error rates was also tested for $\alpha = 20$ and $n_{\mathrm{est}} = 1000$ and did not significantly change the mean squared error (MSE) of the estimate of the parameter vector. We chose $p = 0.13$, which is close to the threshold of the code [17,21], both because we are interested in the regime of high error rates and because estimating logical error rates is difficult in the regime of low rates. A comparison of logical error rates before and after the estimation, using a relatively bad initialization, is shown in Fig. 2. We also compare with the "perfect knowledge decoder" that is given knowledge of the actual error rates. Logical error rates were estimated by decoding $10^5$–$10^6$ random errors, except for the perfect knowledge decoder where $10^8$ random errors were used. A clear improvement is observed even after one iteration, and for five iterations, EM was able to reach close to optimal error rates, while HEM showed no further improvement after the first iteration. We also confirmed that the MSE of the EM estimator is optimal in the sense that it reaches the Cramér-Rao bound, which lower bounds the MSE of any unbiased estimator (Fig. 3). The HEM estimator showed significantly higher MSE. Finally, we note that since it is a form of maximum-likelihood estimation, we expect the estimator to be robust in case of model misspecification [22]—quantifying the robustness is left for future research.

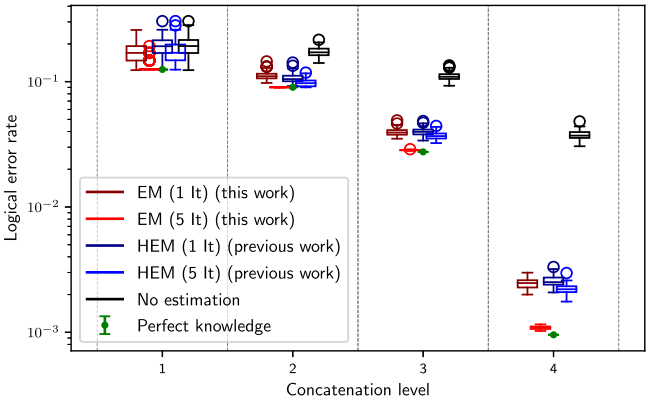



FIG. 2. Logical error rate of the maximum likelihood decoder. Each point is a box-plot including 100 runs with random initializations and estimation data, except for the perfect knowledge decoder, where the error bars indicate a 95% Clopper-Pearson confidence interval. For each concatenation level, the box-plots are shown from left to right in the same order as in the legend, while the dot represents the perfect knowledge decoder. The boxes extend from the lower to the upper quartile of values, with a line at the median. The whiskers extend to the last data point within 1.5 interquartile ranges of the box in each direction. Outliers beyond this are shown individually as circles. The parameters were $p = 0.13$, $\alpha = 20$, and $n_{\mathrm{est}} = 10^3$.

## IV. DETAILS AND PROOFS

In this section, we provide further details and generalizations on some topics, as well as all the proofs that were previously omitted. Furthermore, we present more extensive numerical tests of our estimator.

### A. Analytical solution under a conditional independence assumption

Spitz *et al.* [6] have derived an analytical solution of the estimation problem for certain models. Here, we rederive this solution in a slightly more general setting and discuss the

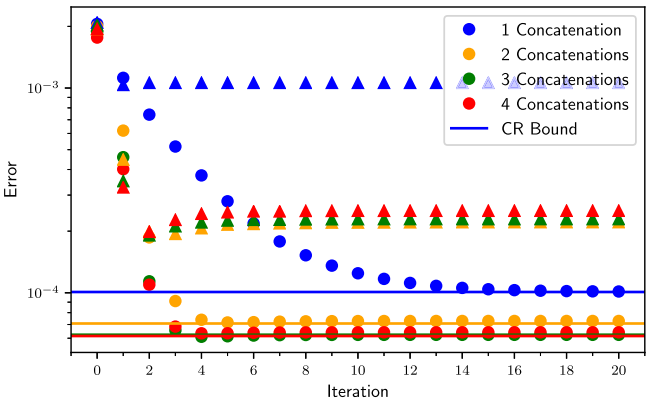



FIG. 3. Comparison of the MSE in $\theta^1_X$ between EM (circles, this work) and HEM (triangles, previous work). The Cramér-Rao bound for each concatenation level is indicated by a line in the matching color. The parameters were $p = 0.13$, $\alpha = 20$, and $n_{\mathrm{est}} = 1000$.

underlying assumptions and limitations by giving examples of quantum codes that cannot be treated in this way.

The estimation method is considered for a circuit noise model, where errors can affect each part of the error correction circuit, including measurements.

*Definition 7 (Independent binary circuit noise).* Let $\{X_q\}_{q=1,\ldots,m}$ denote a collection of (multi-qubit) Pauli errors, where each error may affect one or multiple sites in the error detection circuit. Under *independent binary circuit noise*, each error occurs independently, and the error $X_q$ occurs with probability $\theta_q$.

The errors in $\{X_q\}_{q=1,\ldots,m}$ will also be referred to as *elementary errors*.

In such a model, the errors can be treated as binary variables, where $X_q = 1$ with probability $\theta_q$ and $X_q = 0$ with probability $1 - \theta_q$. Furthermore, the outcomes of the stabilizer generator measurements can be denoted by binary variables $S_i$, where $S_i = 1$ if the total error anticommutes with the $i'$th generator and $S_i = 0$ otherwise.

Then, the rates of errors that affect multiple stabilizers can be estimated using the following proposition.

*Proposition 8.* Consider a stabilizer code subject to independent binary circuit noise. Let $S_1, S_2$ be two syndrome bits and $X$ be an elementary error such that the following three conditions are fulfilled:

(1) $\mathbb{P}[S_1 = S_2 \mid X] = \mathbb{P}[S_1 = S_2]$;

(2) $\mathbb{P}[S_i = 1 \mid X] = \mathbb{P}[S_i = 0 \mid \bar{X}]$ for $i = 1, 2$;

(3) $S_1 \perp S_2 | X$, i.e., $S_1$ is conditionally independent of $S_2$ given $X$,

where $\bar{X} = 1 - X$. Then

$$\mathbb{P}[X = 1]\mathbb{P}[X = 0] = \frac{\mathbb{E}[S_1S_2] - \mathbb{E}[S_1]\mathbb{E}[S_2]}{1 - 2\mathbb{E}[S_1 \oplus S_2]}, \tag{12}$$

where $\oplus$ is addition modulo 2 and $\mathbb{E}[\,\cdot\,]$ denotes expectation values.

The idea is that the correlation between $S_1$ and $S_2$ gives us the rate of the error $X$. Note that the first two conditions are automatically fulfilled for any error $X$ that anticommutes with both $S_1$ and $S_2$. The third condition however is interesting. It essentially states that $X$ is the only elementary error in our noise model that affects both $S_1$ and $S_2$.

*Proof of proposition* 8. Since the syndromes are binary variables, we have

$$\begin{aligned}\mathbb{E}[S_1S_2] - \mathbb{E}[S_1]\mathbb{E}[S_2] &= \mathbb{P}[S_1 = 1, S_2 = 1] \\ &\quad - \mathbb{P}[S_1 = 1]\mathbb{P}[S_2 = 1].\end{aligned}$$

This can be rewritten using the law of total probability and the independence of the errors $X_i$,

$$\begin{aligned}&\mathbb{E}[S_1S_2] - \mathbb{E}[S_1]\mathbb{E}[S_2] \\ &\quad = \sum_{X=0,1} \mathbb{P}[S_1 = 1, S_2 = 1 \mid X]\mathbb{P}[X] \\ &\qquad - \sum_{X,X'=0,1} \mathbb{P}[S_1 = 1 \mid X]\mathbb{P}[S_2 = 1 \mid X']\mathbb{P}[X]\mathbb{P}[X'].\end{aligned}$$

Now we regroup the second term,

$$\begin{aligned}&\mathbb{E}[S_1S_2] - \mathbb{E}[S_1]\mathbb{E}[S_2] \\ &\quad = \sum_{X=0,1} \mathbb{P}[S_1 = 1, S_2 = 1 \mid X]\mathbb{P}[X] \\ &\qquad - \sum_{\substack{X,X'=0,1 \\ X=X'}} \mathbb{P}[S_1 = 1 \mid X]\mathbb{P}[S_2 = 1 \mid X']\mathbb{P}[X]\mathbb{P}[X'] \\ &\qquad - \sum_{\substack{X,X'=0,1 \\ X\neq X'}} \mathbb{P}[S_1 = 1 \mid X]\mathbb{P}[S_2 = 1 \mid X']\mathbb{P}[X]\mathbb{P}[X'] \\ &\quad = \sum_{X=0,1} \mathbb{P}[S_1 = 1, S_2 = 1 \mid X]\mathbb{P}[X] \\ &\qquad - \sum_{X=0,1} \mathbb{P}[S_1 = 1 \mid X]\mathbb{P}[S_2 = 1 \mid X]\mathbb{P}[X]\mathbb{P}[X] \\ &\qquad - \sum_{X=0,1} \mathbb{P}[S_1 = 1 \mid X]\mathbb{P}[S_2 = 1 \mid \bar{X}]\mathbb{P}[X]\mathbb{P}[\bar{X}].\end{aligned}$$

Finally, we use assumptions 3, 2, and 1 in this order to finish the calculation,

$$\begin{aligned}&\mathbb{E}[S_1S_2] - \mathbb{E}[S_1]\mathbb{E}[S_2] \\ &\quad = \mathbb{P}[X = 1]\mathbb{P}[\bar{X} = 1] \\ &\qquad \times \left( \sum_{X=0,1} \mathbb{P}[S_1 = 1 \mid X]\mathbb{P}[S_2 = 1 \mid X] \right. \\ &\qquad \left. - \sum_{X=0,1} \mathbb{P}[S_1 = 1 \mid X]\mathbb{P}[S_2 = 1 \mid \bar{X}] \right) \\ &\quad = \mathbb{P}[X = 1]\mathbb{P}[\bar{X} = 1] \\ &\qquad \times (\mathbb{P}[S_1 = S_2 \mid X = 1] - \mathbb{P}[S_1 \neq S_2 \mid X = 1]) \\ &\quad = \mathbb{P}[X = 1]\mathbb{P}[\bar{X} = 1](1 - 2\mathbb{P}[S_1 \neq S_2]) \\ &\quad = \mathbb{P}[X = 1]\mathbb{P}[\bar{X} = 1](1 - 2\mathbb{E}[S_1 \oplus S_2]),\end{aligned}$$

where we also used $\mathbb{P}[\bar{X}] = 1 - \mathbb{P}[X]$. The derived identity is equivalent to (12). ■

Errors that only affect a single stabilizer can be estimated once the other rates are known, using the following proposition.

*Proposition 9.* Let $S$ be a stabilizer and let $\{X_1, \ldots, X_k\}$ be the set of all elementary errors in our noise model that anticommute with $S$. Then,

$$\prod_{i=1}^{k} (1 - 2\mathbb{P}[X_i = 1]) = (1 - 2\mathbb{E}[S]). \tag{13}$$

*Proof.* By assumption, the outcome of measuring $S$ is completely determined by the errors $X_1, \ldots, X_k$. Therefore

$$(1 - 2\mathbb{E}[S]) = 1 - 2\mathbb{P}[X_1 \oplus \cdots \oplus X_k = 1].$$

Since the elementary errors are independent we can factor out one of them,

$$(1-2\mathbb{E}[S]) = 1-2\Bigg(\mathbb{P}[X_1=1]\mathbb{P}\Bigg[\bigoplus_{i=2}^{k} X_i = 0\Bigg] + \mathbb{P}[X_1=0]\mathbb{P}\Bigg[\bigoplus_{i=2}^{k} X_i = 1\Bigg]\Bigg).$$

Using that $\mathbb{P}[X_1=0] = 1-\mathbb{P}[X_1=1]$, we obtain

$$\begin{aligned}
&(1-2\mathbb{E}[S])\\
&= 1-2\Bigg(\mathbb{P}[X_1=1]\Bigg(1-\mathbb{P}\Bigg[\bigoplus_{i=2}^{k} X_i = 1\Bigg]\Bigg) + (1-\mathbb{P}[X_1=1])\mathbb{P}\Bigg[\bigoplus_{i=2}^{k} X_i = 1\Bigg]\Bigg)\\
&= 1-2\Bigg(\mathbb{P}[X_1=1]\Bigg(1-2\mathbb{P}\Bigg[\bigoplus_{i=2}^{k} X_i = 1\Bigg]\Bigg) + \mathbb{P}\Bigg[\bigoplus_{i=2}^{k} X_i = 1\Bigg]\Bigg)\\
&= 1-2\mathbb{P}[X_1=1] + 4\mathbb{P}[X_1=1]\mathbb{P}\Bigg[\bigoplus_{i=2}^{k} X_i = 1\Bigg] - 2\mathbb{P}\Bigg[\bigoplus_{i=2}^{k} X_i = 1\Bigg]\\
&= (1-2\mathbb{P}[X_1=1])\Bigg(1-2\mathbb{P}\Bigg[\bigoplus_{i=2}^{k} X_i = 1\Bigg]\Bigg).
\end{aligned}$$

The claim now follows by induction. ■

If e.g. the rates of $X_2, \ldots, X_k$ are already determined by using the estimation from the previous section, proposition 9 can be used to estimate $X_1$.

The estimation using propositions 8 and 9 is in closed form, however there are some limitations. First of all, the assumption of binary noise is relatively restrictive. For example, such a model does not include the commonly used depolarizing noise, since the probability of a Pauli $Y$ error is not the product of the probabilities of $X$ and $Z$ errors. It is possible to work around this problem to some extent by modeling depolarizing noise as independent $X, Z$ and $Y$ errors with some effective rates, which works for low error rates. The second problem is that one only considers correlations between pairs of stabilizers, but not higher order correlations. This is generally not sufficient to fully characterize a code. For example, considering the well known five-qubit code subject to independent Pauli noise on each qubit, there are 15 parameters to be estimated (the probabilities of each of the three nontrivial Pauli errors for each of the five qubits), while the two propositions provide at best $\binom{4}{2}+4 = 10$ equations. However, we have shown that it is possible to estimate error rates of this code at least in certain parameter regimes (theorem 5). Furthermore, proposition 8 requires that one can find pairs of stabilizers that are only correlated by a single elementary error. It is not always possible to find such pairs. As an example, consider the seven qubit Steane code subject to only independent Pauli $X$ errors on each qubit. The stabilizers of this code are illustrated in Fig. 4.

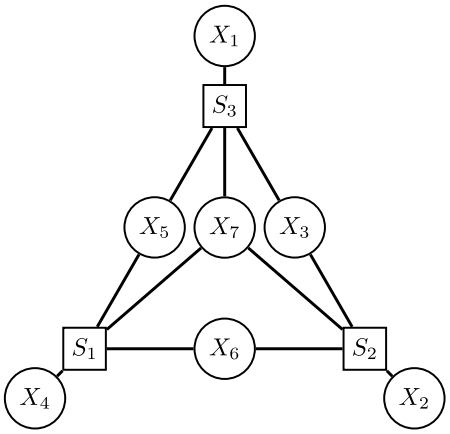



FIG. 4. Errors $X_i$ and stabilizers $S_i$ for the seven qubit Steane code with only $X$ errors (only the three relevant stabilizers are shown). A connection between an error and a stabilizer means that they anticommute.

We see that because of the central error node $X_7$, there are no two stabilizers that are connected only through a single elementary error. Therefore we cannot apply proposition 8 here. However, theorem 5 implies that parameters of this model are identifiable at least in a certain regime. Note that similar problems occur for color codes, since the Steane code is the smallest example of a color code [11].

### B. Generalized identifiability results

In this section, we provide generalized versions of proposition 2 and theorem 3 as well as their proof. Furthermore, we provide the proof of theorem 5.

#### 1. Formal definition of error model

We consider a quite general error model that includes independent single qubit Pauli noise as a special case. There are two main underlying assumptions. The first is that errors on the data qubits and syndrome bits are stochastic Pauli errors and bit flips, which is common in the treatment of quantum error correction codes. The second is that there is some independence between different kinds of errors, which is both of fundamental importance for error correction and often physically reasonable. The first assumption implies that errors can be modeled as elements of the group $\mathcal{E}_n^l = \mathsf{P}_n \times \mathbb{F}_2^l$, where the first component represents a Pauli error on the data and the second component represents a bit flip on the measured syndrome. The product in this group is thus $(\boldsymbol{e}, \boldsymbol{f}), (\boldsymbol{e}', \boldsymbol{f}') \mapsto (\boldsymbol{e}\boldsymbol{e}', \boldsymbol{f} \oplus \boldsymbol{f}')$ and the identity element is $I = (I_{\mathsf{P}_n}, \mathbf{0})$. The syndrome of $(\boldsymbol{e}, \boldsymbol{f}) \in \mathcal{E}_n^l$ is $\boldsymbol{S}(\boldsymbol{e}) \oplus \boldsymbol{f}$.

*Definition 10 (decomposable error model).* Let $N_1, \ldots, N_m \subset \mathcal{E}_n^l$ be disjoint error sets and $I \notin N_i$, $\forall\, i \in \{1, \ldots, m\}$. For each $i \in \{1, \ldots, m\}$, let $\boldsymbol{\theta}^i = (\theta_{\boldsymbol{e}}^i)_{\boldsymbol{e} \in N_i \cup \{I\}}$ be a probability vector over $N_i \cup \{I\}$, and define $\boldsymbol{\theta} = (\theta_{\boldsymbol{e}}^i)_{i \in \{1,\ldots,m\}, \boldsymbol{e} \in N_i}$ by grouping together all these probability vectors and excluding the rates of trivial

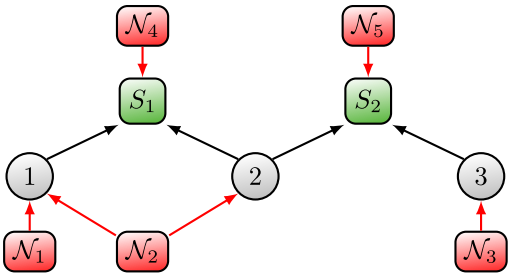



FIG. 5. Representation of a simple decomposable error model on the repetition code. The circles represent the three qubits of the code. The green boxes represent the two stabilizer generators $S_1 = Z \otimes Z \otimes I$ and $S_2 = I \otimes Z \otimes Z$. The noise model decomposes into channels that act independently, as illustrated by the red boxes. For example, the channel $\mathcal{N}_1$ applies an $X$ error to the first qubit with some probability $\theta^1_{X\otimes I\otimes I}$. $\mathcal{N}_2$ applies the error $X \otimes X \otimes I$ with probability $\theta^2_{X\otimes X\otimes I}$ and the error $I \otimes X \otimes I$ with probability $\theta^2_{I\otimes X\otimes I}$. $\mathcal{N}_4$ flips the outcome of the measurement of $S_1$ with probability $\theta^4_{(1,0)}$.

errors. An error model is *decomposable* with error sets $N_1, \ldots, N_m$ and parameters $\boldsymbol{\theta}$ if errors from the different sets occur independently, i.e., the probability of a given error combination $\boldsymbol{X} \in ((N_1 \cup \{I\}) \times \cdots \times (N_m \cup \{I\}))$ is

$$\mathbb{P}[\boldsymbol{X}] = \prod_{i=1}^{m} \prod_{\boldsymbol{e}\in N_i} \left(\theta^i_{\boldsymbol{e}}\right)^{\delta_{X_i,\boldsymbol{e}}} \left(\theta^i_I\right)^{\delta_{X_i,I}}, \tag{14}$$

where $\delta$ is the Kronecker-delta and $\theta^i_I = 1 - \sum_{\boldsymbol{e}\in N_i} \theta^i_{\boldsymbol{e}}$.

An example of such a model is given in Fig. 5. There, the error sets would be $N_1 = \{X \otimes I \otimes I\}, N_2 = \{X \otimes X \otimes I, I \otimes X \otimes I\}, N_3 = \{I \otimes I \otimes X\}, N_4 = \{(1, 0)\}, N_5 = \{(0, 1)\}$. (Since errors here either only act on data qubits or only on syndrome bits we omitted the other trivial part of the errors.) For independent single-qubit Pauli noise the error sets would be given by $N_i = \{\boldsymbol{X}^{(i)}, \boldsymbol{Z}^{(i)}, \boldsymbol{Y}^{(i)}\}$. We will refer to the elements of the individual error sets as *elementary errors*. Since there can be overlap between the supports of the different error channels, we often consider the vector $\boldsymbol{X} \in ((N_1 \cup \{I\}) \times \cdots \times (N_m \cup \{I\}))$, containing all the elementary errors that occurred. The combined error $\boldsymbol{E} \in \mathsf{P}_n$ on the qubits and syndrome bits is then the product of all elementary errors that occurred, i.e., $\boldsymbol{E} = \prod_i X_i$. For independent single qubit Pauli noise $\boldsymbol{X}$ and $\boldsymbol{E}$ coincide. The map $\mathcal{M} : \boldsymbol{\theta} \mapsto (\mathbb{P}[\boldsymbol{S}])_{\boldsymbol{S}\in\mathbb{F}_2^l}$ introduced in Sec. II can now be written as

$$\mathbb{P}[\boldsymbol{S}] = \sum_{\boldsymbol{X}:\boldsymbol{S}(\boldsymbol{X})=\boldsymbol{S}} \mathbb{P}[\boldsymbol{X}]. \tag{15}$$

Our identifiability conditions can now be straightforwardly generalized by considering the elementary errors as the new "single qubit errors." We also note that in the presence of measurement errors, it might be appropriate to include redundant stabilizer measurements such that the length $l$ of a syndrome is larger than the number of stabilizer generators [23–25]. Our results also apply to such a scheme.

### 2. *Proof of proposition 2*

Explicitly, proposition 2 is generalized as follows.

*Proposition 11.* Consider a quantum code subject to a decomposable error model with error sets $N_1, \ldots, N_m$ and parameters $\boldsymbol{\theta}$. Then the parameters of the channel are locally identifiable at $\boldsymbol{\theta} = \boldsymbol{0}$ if and only if $\boldsymbol{S}(\boldsymbol{e}) \neq \boldsymbol{S}(\boldsymbol{e}')$ for every choice of two different elementary errors $\boldsymbol{e}, \boldsymbol{e}' \in \bigcup_{i=1}^m N_i$.

*Proof.* We have to show that the map $\mathcal{M}$ defined in Sec. IV B 1 is locally invertible at $\boldsymbol{0}$. The probability of the error $\boldsymbol{E} \in \mathcal{E}^l_n$ is

$$\mathbb{P}[\boldsymbol{E}] = \sum_{\boldsymbol{X}\,:\,\boldsymbol{E}(\boldsymbol{X})=\boldsymbol{E}} \mathbb{P}[\boldsymbol{X}], \tag{16}$$

where $\mathbb{P}[\boldsymbol{X}]$ is given in (14). The probability of observing syndrome $\boldsymbol{S}$ is

$$\mathbb{P}[\boldsymbol{S}] = \sum_{\boldsymbol{E}\in\mathcal{E}^l_n:\boldsymbol{S}(\boldsymbol{E})=\boldsymbol{S}} \mathbb{P}[\boldsymbol{E}]. \tag{17}$$

Thus the map $\mathcal{M}$ decomposes as $\mathcal{M} = \mathcal{T} \circ g$, where

$$g : \boldsymbol{\theta} \mapsto (\mathbb{P}[\boldsymbol{E}])_{\boldsymbol{E}\in\mathcal{E}^l_n}, \tag{18}$$

describes the distribution of total errors, and

$$\mathcal{T} : (\mathbb{P}[\boldsymbol{E}])_{\boldsymbol{E}\in\mathcal{E}^l_n} \mapsto (\mathbb{P}[\boldsymbol{S}])_{\boldsymbol{S}\in\mathbb{F}_2^l}. \tag{19}$$

describes the probability of each syndrome. Since $\mathcal{T}$ is linear, we have

$$\begin{aligned} J &:= D_{\boldsymbol{\theta}}\mathcal{M} = D_{g(\boldsymbol{\theta})}\mathcal{T} \circ D_{\boldsymbol{\theta}} g \\ &= \mathcal{T} \circ D_{\boldsymbol{\theta}} g. \end{aligned} \tag{20}$$

We begin by calculating the derivative of $\mathbb{P}[\boldsymbol{X}]$,

$$\frac{\partial \mathbb{P}[\boldsymbol{X}]}{\partial \theta^i_{\boldsymbol{e}}} = \delta_{X_i,\boldsymbol{e}} \mathbb{P}[\boldsymbol{X}_{-i}] - \delta_{X_i,I} \mathbb{P}[\boldsymbol{X}_{-i}], \tag{21}$$

where $\boldsymbol{X}_{-i}$ denotes $\boldsymbol{X}$ without the $i$th component. Since we consider $\boldsymbol{\theta} = 0$, $\mathbb{P}[\boldsymbol{X}_{-i}]$ is zero if $X_j \neq I$ for any $i \neq j$. Thus

$$\frac{\partial \mathbb{P}[\boldsymbol{X}]}{\partial \theta^i_{\boldsymbol{e}}} = \begin{cases} +1, & X_i = \boldsymbol{e} \text{ and } X_j = I \; \forall j \neq i, \\ -1, & X_i = I \text{ and } X_j = I \; \forall j \neq i, \\ 0, & \text{otherwise.} \end{cases} \tag{22}$$

We then have

$$\begin{aligned} D_{(\boldsymbol{\theta}=\boldsymbol{0})} g\left[\boldsymbol{E}, \theta^i_{\boldsymbol{e}}\right] &= \frac{\partial \mathbb{P}[\boldsymbol{E}]}{\partial \theta^i_{\boldsymbol{e}}} = \sum_{\boldsymbol{X}:\boldsymbol{E}(\boldsymbol{X})=\boldsymbol{E}} \frac{\partial \mathbb{P}[\boldsymbol{X}]}{\partial \theta^i_{\boldsymbol{e}}} \\ &= \begin{cases} +1, & \boldsymbol{E} = \boldsymbol{e}, \\ -1, & \boldsymbol{E} = I, \\ 0, & \text{otherwise,} \end{cases} \end{aligned} \tag{23}$$

where the last line follows because there is always at most one nonzero summand, since the different error sets are by definition disjoint. Therefore the derivative of $g$ has a very simple form:

$$D_{(\boldsymbol{\theta}=\boldsymbol{0})} g = \begin{pmatrix} -1 & \cdots & -1 \\ \boldsymbol{u}_{\boldsymbol{e}_1} & \cdots & \boldsymbol{u}_{\boldsymbol{e}_k} \end{pmatrix}, \tag{24}$$

where $\boldsymbol{u}_{\boldsymbol{e}_i}$ denotes the unit vector associated to the corresponding elementary error $\boldsymbol{e}_i$, and $k = \sum_{i=1}^m |N_i|$. Since the error sets $N_1, \ldots, N_m$ are disjoint, i.e., there are no duplicate elementary errors, this matrix has $k$ independent columns and thus full column rank. As long as no two elementary errors have the same syndrome, the images of these columns under $\mathcal{T}$ are

again linearly independent. Then $D_{(\boldsymbol{\theta}=\mathbf{0})}\mathcal{M} = \mathcal{T} \circ D_{(\boldsymbol{\theta}=\mathbf{0})}g$ has full column rank, and the inverse function theorem completes the proof. ■

#### 3. Proof of theorem 3

Our general version of theorem 3 can be stated as follows.

*Theorem 12.* Consider a quantum code subject to a decomposable error model with error sets $N_1, \ldots, N_m$ and parameters $\boldsymbol{\theta}$. Assume that $\theta^i_{\boldsymbol{e}} > 0$ for all $i$ and all $\boldsymbol{e} \in N_i$, and that $\mathbb{P}[\boldsymbol{S}] > 0$ for all syndromes $\boldsymbol{S} \in \mathbb{F}_2^l$. Then the error rates are locally identifiable at $\boldsymbol{\theta}$ if and only if the matrix $\tilde{J}$ with entries

$$\tilde{J}\big[\boldsymbol{S}, \theta^i_{\boldsymbol{e}}\big] = \frac{\mathbb{P}[X_i = \boldsymbol{e}|\boldsymbol{S}]}{\mathbb{P}[X_i = \boldsymbol{e}]} - \frac{\mathbb{P}[X_i = I|\boldsymbol{S}]}{\mathbb{P}[X_i = I]} \tag{25}$$

has full column rank.

*Proof.* We have to show that the map $\mathcal{M}$ defined in Sec. IV B 1 is locally invertible at $\boldsymbol{\theta}$. Since we assume that the rates of all errors and syndromes are strictly greater than 0, the $\mathcal{M}$ will be locally invertible at $\boldsymbol{\theta}$ if and only if the entrywise logarithm $\ln(\mathcal{M})$ is locally invertible at $\boldsymbol{\theta}$. Thus we consider the derivative of the logarithm-likelihood $\ln(\mathbb{P}[\boldsymbol{S}])$ for each syndrome. Remember that the probability of a syndrome $\boldsymbol{S}$ can be expressed as

$$\mathbb{P}[\boldsymbol{S}] = \sum_{\boldsymbol{X}:\boldsymbol{S}(\boldsymbol{X})=\boldsymbol{S}} \mathbb{P}[\boldsymbol{X}], \tag{26}$$

where $\mathbb{P}[\boldsymbol{X}]$ is given in (14). As in the proof of proposition 11, we compute the derivative

$$\begin{aligned}\frac{\partial \mathbb{P}[\boldsymbol{X}]}{\partial \theta^i_{\boldsymbol{e}}} &= \delta_{X_i,\boldsymbol{e}}\mathbb{P}[\boldsymbol{X}_{-i}] - \delta_{X_i,I}\mathbb{P}[\boldsymbol{X}_{-i}] \\ &= \delta_{X_i,\boldsymbol{e}}\frac{\mathbb{P}[\boldsymbol{X}]}{\theta^i_{\boldsymbol{e}}} - \delta_{X_i,I}\frac{\mathbb{P}[\boldsymbol{X}]}{\theta^i_I} \\ &= \delta_{X_i,\boldsymbol{e}}\frac{\mathbb{P}[\boldsymbol{X}]}{\mathbb{P}[X_i = \boldsymbol{e}]} - \delta_{X_i,I}\frac{\mathbb{P}[\boldsymbol{X}]}{\mathbb{P}[X_i = I]}.\end{aligned}$$

Using the fact that

$$\mathbb{P}[\boldsymbol{S}] = \sum_{\boldsymbol{X}|\boldsymbol{S}(\boldsymbol{X})=\boldsymbol{S}} \mathbb{P}[\boldsymbol{X}], \tag{27}$$

we obtain

$$\begin{aligned}&\frac{\partial \ln(\mathbb{P}[\boldsymbol{S}])}{\partial \theta^i_{\boldsymbol{e}}} \\ &= \frac{1}{\mathbb{P}[\boldsymbol{S}]}\sum_{\boldsymbol{X}:\boldsymbol{S}(\boldsymbol{X})=\boldsymbol{S}}\left(\delta_{X_i,\boldsymbol{e}}\frac{\mathbb{P}[\boldsymbol{X}]}{\mathbb{P}[X_i = \boldsymbol{e}]} - \delta_{X_i,I}\frac{\mathbb{P}[\boldsymbol{X}]}{\mathbb{P}[X_i = I]}\right) \\ &= \frac{1}{\mathbb{P}[\boldsymbol{S}]}\left(\frac{\sum_{\boldsymbol{X}:\boldsymbol{S}(\boldsymbol{X})=\boldsymbol{S}}\delta_{X_i,\boldsymbol{e}}\mathbb{P}[\boldsymbol{X}]}{\mathbb{P}[X_i = \boldsymbol{e}]} - \frac{\sum_{\boldsymbol{X}:\boldsymbol{S}(\boldsymbol{X})=\boldsymbol{S}}\delta_{X_i,I}\mathbb{P}[\boldsymbol{X}]}{\mathbb{P}[X_i = I]}\right) \\ &= \frac{\mathbb{P}[X_i = \boldsymbol{e}|\boldsymbol{S}]}{\mathbb{P}[X_i = \boldsymbol{e}]} - \frac{\mathbb{P}[X_i = I|\boldsymbol{S}]}{\mathbb{P}[X_i = I]}.\end{aligned} \tag{28}$$

By the inverse function theorem, this completes the proof. ■

#### 4. Proof of lemma 6

We will now proof lemma 6 in order to finish the proof of theorem 5. Remember that the $i$th tensor factor of $\boldsymbol{E} \in \mathsf{P}_n$ is denoted $E_i$. Furthermore, the Pauli acting as $e \in P_1$ on qubit $i$ and as the identity everywhere else is denoted $\boldsymbol{e}^{(i)}$. Finally, for $\boldsymbol{E} \in \mathsf{P}_n$, we use $\boldsymbol{E}_{|i}$ as a shorthand for $(E_i)^{(i)}$. We define the weight of a Pauli error in the standard way.

*Definition 13 (weight).* The *weight* of a Pauli error $\boldsymbol{E} = E_1 \otimes E_2 \otimes \cdots \otimes E_n \in \mathsf{P}_n$ is defined as

$$\mathrm{wt}(\boldsymbol{E}) := |\{E_i \mid E_i \neq I, i \in \{1, \ldots, n\}\}|. \tag{29}$$

In the case of equal error rates $p$, the probability of an error is determined by its weight. Let us denote $\bar{p} := 1 - p$. We obtain a convenient expression for $\mathbb{P}[E_i = e, \boldsymbol{S}]$. For $e \neq I$, we have

$$\begin{aligned}&\mathbb{P}[E_i = e, \boldsymbol{S}] \\ &= \sum_{\substack{\boldsymbol{E} \in \mathsf{P}_n : \\ \boldsymbol{S}(\boldsymbol{E}) = \boldsymbol{S}, E_i = e}} \mathbb{P}[\boldsymbol{E}] \\ &= \sum_{\substack{\boldsymbol{E} \in \mathsf{P}_n : \\ \boldsymbol{S}(\boldsymbol{E}) = \boldsymbol{S}, E_i = e}} p^{\mathrm{wt}(\boldsymbol{E})}(\bar{p})^{n-\mathrm{wt}(\boldsymbol{E})} \\ &= p\sum_{\substack{\boldsymbol{E} \in \mathsf{P}_n : \\ \boldsymbol{S}(\boldsymbol{E}) = \boldsymbol{S}, E_i = e}} p^{\mathrm{wt}(\boldsymbol{E}_{-i})}(\bar{p})^{n-1-\mathrm{wt}(\boldsymbol{E}_{-i})}\end{aligned} \tag{30}$$

and, analogously, for $e = I$,

$$\begin{aligned}&\mathbb{P}[E_i = e, \boldsymbol{S}] \\ &= \bar{p}\sum_{\substack{\boldsymbol{E} \in \mathsf{P}_n : \\ \boldsymbol{S}(\boldsymbol{E}) = \boldsymbol{S}, E_i = e}} p^{\mathrm{wt}(\boldsymbol{E}_{-i})}(\bar{p})^{n-1-\mathrm{wt}(\boldsymbol{E}_{-i})}.\end{aligned} \tag{31}$$

By the definition of conditional probability, we obtain

$$\begin{aligned}&\mathbb{P}[\boldsymbol{S} \mid E_i = e] \\ &= \sum_{\substack{\boldsymbol{E} \in \mathsf{P}_n : \\ \boldsymbol{S}(\boldsymbol{E}) = \boldsymbol{S}, E_i = e}} p^{\mathrm{wt}(\boldsymbol{E}_{-i})}(\bar{p})^{n-1-\mathrm{wt}(\boldsymbol{E}_{-i})}.\end{aligned} \tag{32}$$

Lemma 6 is thus equivalent to the following lemma.

*Lemma 14.* Consider a perfect single error correcting code on $n$ qubits. Let $e, e' \in \mathsf{P}_1$. Then for any syndrome $\boldsymbol{S} \in \mathbb{F}_2^l \setminus \{\mathbf{0}\}$, error rate $p \in [0, 1]$ and qubit $i$ such that $\boldsymbol{S} \neq \boldsymbol{S}(\boldsymbol{e}^{(i)})$ and $\boldsymbol{S} \neq \boldsymbol{S}((\boldsymbol{e}')^{(i)})$ the following equality holds:

$$\begin{aligned}&\sum_{\substack{\boldsymbol{E} \in \mathsf{P}_n : \\ \boldsymbol{S}(\boldsymbol{E}) = \boldsymbol{S}, E_i = e}} p^{\mathrm{wt}(\boldsymbol{E}_{-i})}(\bar{p})^{n-1-\mathrm{wt}(\boldsymbol{E}_{-i})} \\ &= \sum_{\substack{\boldsymbol{E} \in \mathsf{P}_n : \\ \boldsymbol{S}(\boldsymbol{E}) = \boldsymbol{S}, E_i = e'}} p^{\mathrm{wt}(\boldsymbol{E}_{-i})}(\bar{p})^{n-1-\mathrm{wt}(\boldsymbol{E}_{-i})}.\end{aligned} \tag{33}$$

In other words, we have to show that the sums in the expression do not depend on $e$ except if $\boldsymbol{S} = S(\boldsymbol{e}^{(i)})$. This is the case if for all $w = 0, \ldots, n-1$ and $e, e' \in \mathsf{P}_1$ such that

$\boldsymbol{S} \neq \boldsymbol{S}(\mathrm{e}^{(i)}), \boldsymbol{S}((\mathrm{e}')^{(i)})$ we have

$$\begin{aligned} &|\{\boldsymbol{E} \mid E_i = e, \mathrm{wt}(E_{-i}) = w, \boldsymbol{S}(\boldsymbol{E}) = \boldsymbol{S}\}| \\ &\overset{!}{=} |\{\boldsymbol{E} \mid E_i = e', \mathrm{wt}(E_{-i}) = w, \boldsymbol{S}(\boldsymbol{E}) = \boldsymbol{S}\}|, \end{aligned} \tag{34}$$

since then the coefficients for each of the exponents appearing in the expressions will be equal. Therefore, in the following, we will derive an expression for the "modified" weight distribution given by

$$k_w(e, i, \boldsymbol{S}) \coloneqq |\underbrace{\{\boldsymbol{E} \mid E_i = e, \mathrm{wt}(\boldsymbol{E}_{-i}) = w, \boldsymbol{S}(\boldsymbol{E}) = \boldsymbol{S}\}}_{\coloneqq K_w(e,i,\boldsymbol{S})}|. \tag{35}$$

We will show that this distribution is independent of $e$ if $\boldsymbol{S} \neq \boldsymbol{S}(\boldsymbol{e}^{(i)})$. For the rest of this section, we fix a qubit $\hat{q} \in \{1, \ldots, n\}$, an error $\hat{e} \in \mathsf{P}_1$ which will act on $\hat{q}$ and some syndrome $\boldsymbol{0} \neq \boldsymbol{S}^* \in \mathbb{F}_2^l$, and we denote $k_w \coloneqq k_w(\hat{e}, \hat{q}, \boldsymbol{S}^*)$ and $K_w \coloneqq K_w(\hat{e}, \hat{q}, \boldsymbol{S}^*)$. For now, we do not assume that $\boldsymbol{S}^* \neq \boldsymbol{S}(\hat{e}^{(\hat{q})})$.

*Notation 15 (Perfect code property).* Since we consider a perfect single error correcting code, for each syndrome $\boldsymbol{S}$ there exists a unique single qubit error $\boldsymbol{e}^{(q)}$ with $\boldsymbol{S}(\boldsymbol{e}^{(q)}) = \boldsymbol{S}$. We denote this error by $\boldsymbol{S}^{-1}(\boldsymbol{S})$.

The core idea of the proof is to construct the sets $K_w$ iteratively. We can use the perfect code property to construct weight $w$ errors with syndrome $\boldsymbol{S}^*$ from weight $w-1$ errors with any syndrome $\boldsymbol{S}'$ by adding the unique single qubit error $\boldsymbol{S}^{-1}(\boldsymbol{S}' \oplus \boldsymbol{S}^*)$. We formalize this as follows.

*Definition 16 ($\boldsymbol{S}^*$ modification and $\hat{\boldsymbol{e}}^{(\hat{q})}$ extension).* Let $\boldsymbol{E} \in \mathsf{P}_n$. We say an error $\boldsymbol{E}^*$ is a $\boldsymbol{S}^*$ *modification* of $\boldsymbol{E}$ if $\boldsymbol{S}(\boldsymbol{E}^*) = \boldsymbol{S}^*$ and there exists a single qubit error $\boldsymbol{e}^{(q)}$ with $\boldsymbol{E}^* = \boldsymbol{E}\boldsymbol{e}^{(q)}$.

We say $\boldsymbol{E}^*$ is an $\hat{\boldsymbol{e}}^{(\hat{q})}$ *extension* of $\boldsymbol{E}$ if $\boldsymbol{E}^*$ is a $\boldsymbol{S}^*$ modification of $\boldsymbol{E}$ with $\mathrm{wt}(\boldsymbol{E}^*_{-\hat{q}}) = \mathrm{wt}(\boldsymbol{E}_{-\hat{q}}) + 1$ and $E^*_{\hat{q}} = \hat{e}$.

Note that this definition does depend on the choice of $\hat{\boldsymbol{e}}^{(q)}$ and $\boldsymbol{S}^*$, which is fixed for the rest of this section.

It is simple to construct a $\boldsymbol{S}^*$ modification for each error.

*Lemma 17.* Each error $\boldsymbol{E} \in \mathsf{P}_n$ has a **unique** $\boldsymbol{S}^*$ modification. We denote it $\boldsymbol{E}^*$.

*Proof.* Let $\boldsymbol{e}^{(q)} = \boldsymbol{S}^{-1}(\boldsymbol{S}^* \oplus \boldsymbol{S}(\boldsymbol{E}))$. Then $\boldsymbol{E}\boldsymbol{e}^{(q)}$ is a $\boldsymbol{S}^*$ modification of $\boldsymbol{E}$. Furthermore, for two possible $\boldsymbol{S}^*$ modifications $\boldsymbol{E}\boldsymbol{e}^{(q)}, \boldsymbol{E}(\boldsymbol{e}')^{(q')}$ with $\boldsymbol{S}(\boldsymbol{E}\boldsymbol{e}^{(q)}) = \boldsymbol{S}(\boldsymbol{E}(\boldsymbol{e}')^{(q')}) = \boldsymbol{S}^*$, we obtain $\boldsymbol{S}(\boldsymbol{e}^{(q)}) = \boldsymbol{S}((\boldsymbol{e}')^{(q')}) = \boldsymbol{S}^* \oplus \boldsymbol{S}(\boldsymbol{E})$. Because we consider a perfect code this implies $\boldsymbol{e}^{(q)} = (\boldsymbol{e}')^{(q')}$. Thus the $\boldsymbol{S}^*$ modification is unique. ∎

However, it is possible that an error $\boldsymbol{E}$ does not have a $\hat{\boldsymbol{e}}^{(\hat{q})}$ extension. This happens for example if the unique single qubit error that needs to be added to obtain the $\boldsymbol{S}^*$ modification is already in $\boldsymbol{E}$, or if it is on $\hat{q}$. We formalize this in the following corollary.

*Corollary 18.* Let $\boldsymbol{E} \in \mathsf{P}_n$ be an error with $E_{\hat{q}} = \hat{e}$ and $\mathrm{wt}(\boldsymbol{E}_{-\hat{q}}) = w$. $\boldsymbol{E}$ does **not** have an $\hat{\boldsymbol{e}}^{(\hat{q})}$ extension if and only if one of the following mutually exclusive conditions is true:

(i) $\boldsymbol{E}^* = \boldsymbol{E}$;
(ii) $\mathrm{wt}(\boldsymbol{E}^*_{-\hat{q}}) = w - 1 \wedge E^*_{\hat{q}} = \hat{e}$;
(iii) $\boldsymbol{E}^* \neq \boldsymbol{E} \wedge \mathrm{wt}(\boldsymbol{E}^*_{-\hat{q}}) = w \wedge E^*_{\hat{q}} = \hat{e}$;
(iv) $E^*_{\hat{q}} \neq \hat{e}$.

where as always $\boldsymbol{E}^*$ is the unique $\boldsymbol{S}^*$ modification of $\boldsymbol{E}$. If we write $\boldsymbol{E}^* = \boldsymbol{E}\boldsymbol{e}^{(q)}$, where $\boldsymbol{e}^{(q)}$ is a uniquely determined single qubit error acting on qubit $q$, these conditions are equivalent to

(i′) $e = I$;
(ii′) $E_q = e \wedge q \neq \hat{q} \wedge e \neq I$;
(iii′) $E_q \neq e \wedge E_q \neq I \wedge q \neq \hat{q} \wedge e \neq I$;
(iv′) $q = \hat{q} \wedge e \neq I$.

*Proof.* By definition $\boldsymbol{E}^* = \boldsymbol{E}\boldsymbol{e}^{(q)}$ is an $\hat{\boldsymbol{e}}^{(\hat{q})}$ extension of $\boldsymbol{E}$ if and only if

$$\begin{aligned} &E^*_{\hat{q}} = \hat{e} \ \wedge \ \mathrm{wt}(\boldsymbol{E}^*_{-\hat{q}}) = w + 1 \\ &\quad \Leftrightarrow q \neq \hat{q} \ \wedge \ E_q = I \ \wedge \ e \neq I, \end{aligned} \tag{36}$$

where we have used that $E_{\hat{q}} = \hat{e}$. Negating this statement and using that $\mathrm{wt}(\boldsymbol{E}^*_{-\hat{q}}) \in \{w-1, w, w+1\}$ leads to the conditions above. ∎

Since similar reasoning will be used repeatedly throughout this section, let us illustrate some of the cases in corollary 18 with an example. Consider the five-qubit perfect code with stabilizer generators $g_1 = X \otimes Z \otimes Z \otimes X \otimes I$, $g_2 = I \otimes X \otimes Z \otimes Z \otimes X$, $g_3 = X \otimes I \otimes X \otimes Z \otimes Z$, and $g_4 = Z \otimes X \otimes I \otimes X \otimes Z$. For this example, let $\hat{q} = 1, \hat{e} = X$, and $\boldsymbol{S}^* = (1, 0, 0, 1)$. The error $\boldsymbol{E} = X \otimes X \otimes X \otimes I \otimes I$ has the syndrome (0,1,0,1), and thus its $\boldsymbol{S}^*$ modification is obtained by applying $\boldsymbol{S}^{-1}((1, 1, 0, 0)) = \boldsymbol{X}^{(3)}$, resulting in $\boldsymbol{E}^* = X \otimes X \otimes I \otimes I \otimes I$. This is not a valid $\hat{\boldsymbol{e}}^{(\hat{q})}$ extension since the weight was reduced, corresponding to case 18 in corollary 18. The single qubit error we applied canceled with an existing error in $\boldsymbol{E}$. On the other hand, the error $\boldsymbol{E} = X \otimes I \otimes Z \otimes Z \otimes I$ has the syndrome $\boldsymbol{S}(\boldsymbol{E}) = (1, 0, 1, 0)$. Thus its $\boldsymbol{S}^*$ modification is obtained by adding $\boldsymbol{e} = \boldsymbol{S}^{-1}((0, 0, 1, 1)) = \boldsymbol{X}^{(5)}$, resulting in $\boldsymbol{E}^* = X \otimes I \otimes Z \otimes Z \otimes X$. This is a valid $\hat{\boldsymbol{e}}^{(\hat{q})}$ extension. Notice that the additional single qubit error was applied on qubit 5 where $\boldsymbol{E}$ acts trivially, or equivalently, $E^*_5 = e$.

In corollary 18, we categorized errors without a valid $\hat{\boldsymbol{e}}^{(\hat{q})}$ extension by their $\boldsymbol{S}^*$ modification. Now we characterize $k_w$ in terms of $\hat{\boldsymbol{e}}^{(\hat{q})}$ extensions.

*Lemma 19.* For any $w > 0$,

$$\begin{aligned} k_w = |\{&\boldsymbol{E} \in \mathsf{P}_n \mid \exists \boldsymbol{E}' \in \mathsf{P}_n : \boldsymbol{E} \text{ is a } \hat{\boldsymbol{e}}^{(\hat{q})} \text{ extension of } \boldsymbol{E}', \\ &E'_{\hat{q}} = \hat{e}, \mathrm{wt}(\boldsymbol{E}'_{-\hat{q}}) = w - 1\}|. \end{aligned}$$

*Proof.* By definition of $k_w$ (35), we have to show that

$$\begin{aligned} |\{\boldsymbol{E} \in \mathsf{P}_n \mid \quad &E_{\hat{q}} = \hat{e}, \mathrm{wt}(\boldsymbol{E}_{-\hat{q}}) = w, \boldsymbol{S}(\boldsymbol{E}) = \boldsymbol{S}^*\}| = \\ |\{\boldsymbol{E} \in \mathsf{P}_n \mid \quad &\exists \boldsymbol{E}' \in \mathsf{P}_n : \boldsymbol{E} \text{ is a } \hat{\boldsymbol{e}}^{(\hat{q})} \text{ extension of } \boldsymbol{E}', \\ &E'_{\hat{q}} = \hat{e}, \mathrm{wt}(\boldsymbol{E}'_{-\hat{q}}) = w - 1\}|. \end{aligned}$$

"$\supseteq$": By definition of $\hat{\boldsymbol{e}}^{(\hat{q})}$ extension.

"$\subseteq$": Let $\boldsymbol{E} \in \mathsf{P}_n$ be an error such that $E_{\hat{q}} = \hat{e}$, $\mathrm{wt}(\boldsymbol{E}_{-\hat{q}}) = w$ and $\boldsymbol{S}(\boldsymbol{E}) = \boldsymbol{S}^*$. Chose a qubit $q \neq \hat{q}$ such that $E_q \neq I$. Then $\boldsymbol{E}$ is an $\hat{\boldsymbol{e}}^{(\hat{q})}$ extension of $\boldsymbol{E}' \coloneqq \boldsymbol{E}\boldsymbol{E}_{|q}$. Furthermore $\mathrm{wt}(\boldsymbol{E}'_{-\hat{q}}) = \mathrm{wt}(\boldsymbol{E}_{-\hat{q}}) - 1$ and $E'_{\hat{q}} = \hat{e}$ by definition of $\boldsymbol{E}'$. ∎

While this establishes a connection between the weight distribution $k_w$ and the concept of $\hat{\boldsymbol{e}}^{(\hat{q})}$ extension, it is difficult to count all errors that are valid $\hat{\boldsymbol{e}}^{(\hat{q})}$ extensions. A number easier to characterize is

$$l_w \coloneqq |L_w| \tag{37}$$

with

$$L_w := \{\boldsymbol{E} \in \mathsf{P}_n \,|\boldsymbol{E} \text{ has a } \hat{\boldsymbol{e}}^{(\hat{q})} \text{ extension}, \\ E_{\hat{q}} = \hat{e}, \mathrm{wt}(\boldsymbol{E}_{-\hat{q}}) = w-1\}. \tag{38}$$

This is similar to the characterization in lemma 19, but $l_w > k_w$ because two different errors can have the same $\hat{\boldsymbol{e}}^{(\hat{q})}$ extension. We have to correct for this "double counting."

*Lemma 20.*

$$k_w = \frac{l_w}{w}. \tag{39}$$

*Proof.* By lemma 17 we have a well defined function $g : \mathsf{P}_n \mapsto \mathsf{P}_n$ that maps an error $\boldsymbol{E} \in \mathsf{P}_n$ to its $\boldsymbol{S}^*$ modification $\boldsymbol{E}^* \in P_n$. By lemma 19 and the definition of $L_w$, $g$ maps $L_w$ to $K_w$, and the restriction $g_{|L_w} : L_w \to K_w$ is surjective. Because the $\boldsymbol{S}^*$ modification is unique, the pre-images of two distinct elements of $K_w$ under $g$ are disjoint. Thus

$$|L_w| = \sum_{\boldsymbol{E}\in K_w} \left|g^{-1}_{|L_w}(\boldsymbol{E})\right|. \tag{40}$$

We want to determine the size of these pre-images. So let $\boldsymbol{E} \in K_w$. From the definition of $L_w$ and the definition of $\hat{\boldsymbol{e}}^{(\hat{q})}$ extension, it follows that $\boldsymbol{E}' \in g^{-1}_{|L_w}(\boldsymbol{E})$ if and only if there exists a qubit $q \neq \hat{q}$ such that $E_q \neq I$ and $\boldsymbol{E}' = \boldsymbol{E}\boldsymbol{E}_{|q}$. Thus, since by definition $\mathrm{wt}(\boldsymbol{E}_{-\hat{q}}) = w$, the preimage has $w$ elements. This concludes the proof. ∎

Thus we can characterize the weight distribution $k_w$ through the numbers $l_w$, for which we derive a recursive formula.

*Lemma 21.* There exists a recursive formula relating $k_w$ to $k_{w-1}$ and $k_{w-2}$.

*Proof.* We prove that

$$l_w = 3^{w-1}\binom{n-1}{w-1} - k_{w-1} - 3(n-w+1)k_{w-2} \\ - 2(w-1)k_{w-1} - \sum_{\substack{e'\in \mathsf{P}_1 \\ \hat{e}\neq e'}} k_{w-1}(e', \hat{q}, \boldsymbol{S}^* \tag{41}$$

for any $2 \leqslant w \leqslant n$. Lemma 20 then gives the corresponding equation for $k_w$.

The total number of errors $\boldsymbol{E} \in \mathsf{P}_n$ with $\mathrm{wt}(\boldsymbol{E}_{-\hat{q}}) = w-1$ and $E_{\hat{q}} = \hat{e}$ is $3^{w-1}\binom{n-1}{w-1}$ since there are $\binom{n-1}{w-1}$ ways to chose $w-1$ positions in $n-1$ positions, and three possible Paulis on each position. Next we count how many of them do **not** have an $\hat{\boldsymbol{e}}^{(\hat{q})}$ extension. The different conditions for this are given in corollary 18, where errors without an $\hat{\boldsymbol{e}}^{(\hat{q})}$ extension are categorized by their $\boldsymbol{S}^*$ modification. We count the number of errors $\boldsymbol{E} \in \mathsf{P}_n$ with $\mathrm{wt}(\boldsymbol{E}_{-\hat{q}}) = w-1$ and $E_{\hat{q}} = \hat{e}$ fulfilling each of these different conditions. We can group errors without a valid $\hat{\boldsymbol{e}}^{(\hat{q})}$ extension by their $\boldsymbol{S}^*$ modification, i.e.,

$$\{\boldsymbol{E} \in \mathsf{P}_n \,|\mathrm{wt}(\boldsymbol{E}_{-\hat{q}}) = w-1, \\ E_{\hat{q}} = \hat{e}, \boldsymbol{E} \text{ has no } \hat{\boldsymbol{e}}^{(\hat{q})} \text{ extension}\} \\ = \bigcup_{\substack{\boldsymbol{E}'\in \mathsf{P}_n: \\ \boldsymbol{E}' \text{ is not an } \hat{\boldsymbol{e}}^{(\hat{q})} \text{ extension}}} \{\boldsymbol{E} \in \mathsf{P}_n \mid \boldsymbol{E}^* = \boldsymbol{E}', \quad E_{\hat{q}} = \hat{e}, \mathrm{wt}(\boldsymbol{E}_{-\hat{q}}) = w-1\}, \tag{42}$$

where all the individual sets are disjoint because the $\boldsymbol{S}^*$ modification is unique. To do this, we have to consider the following cases, for each of which $\mathrm{wt}(\boldsymbol{E}_{-\hat{q}}) = w-1$ and $E_{\hat{q}} = \hat{e}$ hold.

Case (i): $\boldsymbol{E}^* = \boldsymbol{E}$. This condition is equivalent to $\boldsymbol{S}^* = \boldsymbol{S}(\boldsymbol{E})$. By definition there are $k_{w-1}$ such errors.

Case (ii): $\mathrm{wt}(\boldsymbol{E}^*_{-\hat{q}}) = w-2 \wedge E^*_{\hat{q}} = \hat{e}$. For each error $\boldsymbol{E}$ fulfilling this condition, we have that $\boldsymbol{E} = \boldsymbol{E}^*\boldsymbol{e}^{(q)}$ for a Pauli $e \in \mathsf{P}_1 \setminus \{I\}$ and a qubit $q \neq \hat{q}$ with $E^*_q = I$. For a given error $\boldsymbol{E}'$ with $\mathrm{wt}(\boldsymbol{E}'_{-\hat{q}}) = w-2$, $\boldsymbol{E}'_{\hat{q}} = \hat{e}$ and $\boldsymbol{S}(\boldsymbol{E}') = \boldsymbol{S}^*$, there are $n-1-(w-2) = n-w+1$ possibilities to chose a qubit $q \neq \hat{q}$ with $E'_q = I$. For each of these, there are three different Paulis one could add to this position. Each of these gives a distinct error $\boldsymbol{E}$ with $\boldsymbol{E}^* = \boldsymbol{E}'$. The total number of errors $\boldsymbol{E}'$ with $\mathrm{wt}(\boldsymbol{E}'_{-\hat{q}}) = w-2$, $\boldsymbol{E}'_{\hat{q}} = \hat{e}$ and $\boldsymbol{S}(\boldsymbol{E}') = \boldsymbol{S}^*$ is by definition $k_{w-2}$, and because the $\hat{\boldsymbol{e}}^{(\hat{q})}$ extension is unique they all give distinct contributions. Thus

$$|\{\boldsymbol{E} \in \mathsf{P}_n \mid \mathrm{wt}(\boldsymbol{E}^*_{-\hat{q}}) = w-2, E^*_{\hat{q}} = \hat{e}, \\ \mathrm{wt}(\boldsymbol{E}_{-\hat{q}}) = w-1, E_{\hat{q}} = \hat{e}\}| \\ = 3(n-w+1)\,|\{\boldsymbol{E}' \in \mathsf{P}_n \mid E'_{\hat{q}} = \hat{e}, \mathrm{wt}(\boldsymbol{E}'_{-\hat{q}}) = w-2, \\ \boldsymbol{S}(\boldsymbol{E}') = \boldsymbol{S}^*\}| \\ = 3(n-w+1)k_{w-2}.$$

Case (iii): $\boldsymbol{E}^* \neq \boldsymbol{E} \wedge \mathrm{wt}(\boldsymbol{E}^*_{-\hat{q}}) = w-1 \wedge E^*_{\hat{q}} = \hat{e}$. For each such error $\boldsymbol{E}$ it holds $\boldsymbol{E} = \boldsymbol{E}^*\boldsymbol{e}^{(q)}$ for a Pauli $e \in \mathsf{P}_1 \setminus \{I, E^*_q\}$ and a qubit $q \neq \hat{q}$ with $E^*_q \neq I$. For a given error $\boldsymbol{E}'$ with $\mathrm{wt}(\boldsymbol{E}'_{-\hat{q}}) = w-1$, there are $w-1$ choices for $q \neq \hat{q}$ such that $E'_q \neq I$, and for each choice of $q$ there are 2 possible choices of $e \in P_1 \setminus \{I, E'_q\}$. The total number of errors $\boldsymbol{E}'$ with $\mathrm{wt}(\boldsymbol{E}'_{-\hat{q}}) = w-1$, $E'_{\hat{q}} = \hat{e}$ and $\boldsymbol{S}(\boldsymbol{E}') = \boldsymbol{S}^*$ is by definition $k_{w-1}$, and again they give distinct contributions. Thus

$$|\{\boldsymbol{E} \in \mathsf{P}_n \,|\boldsymbol{E}^* \neq \boldsymbol{E}, \mathrm{wt}(\boldsymbol{E}^*_{-\hat{q}}) = w-1, \\ E^*_{\hat{q}} = \hat{e}, \mathrm{wt}(\boldsymbol{E}_{-\hat{q}}) = w-1, E_{\hat{q}} = \hat{e}\}| \\ = 2(w-1)|\{\boldsymbol{E}' \mid E'_{\hat{q}} = \hat{e}, \mathrm{wt}(\boldsymbol{E}'_{-\hat{q}}) = w-1, \boldsymbol{S}(\boldsymbol{E}') = \boldsymbol{S}^*\}| \\ = 2(w-1)k_{w-1}.$$

Case (iv): $E^*_{\hat{q}} \neq \hat{e}$. For each such error $\boldsymbol{E}$ there exists a corresponding $\boldsymbol{E}' = \boldsymbol{E}^*$ such that $\boldsymbol{E} = \boldsymbol{E}'\boldsymbol{e}^{(\hat{q})}$ for an appropriate Pauli $e \in \mathsf{P}_1 \setminus \{I\}$. Note that $\mathrm{wt}(\boldsymbol{E}'_{-\hat{q}}) = \mathrm{wt}(\boldsymbol{E}_{-\hat{q}}) = w-1$. The total number of errors $\boldsymbol{E}'$ with $\mathrm{wt}(\boldsymbol{E}'_{-\hat{q}}) = w-1$, $E'_{\hat{q}} \neq \hat{e}$ and $\boldsymbol{S}(\boldsymbol{E}') = \boldsymbol{S}^*$ is by definition $\sum_{e'\in\mathsf{P}_1|\hat{e}\neq e'} k_{w-1}(e', \hat{q}, \boldsymbol{S}^*)$, and because the $\boldsymbol{S}^*$ modification is unique the different $e'$ give different contributions.

There is no double counting because the union in (42) is disjoint. Finally we obtain the number of errors that have a valid $\hat{\boldsymbol{e}}^{(\hat{q})}$ extension by subtracting the number of errors without a valid $\hat{\boldsymbol{e}}^{(\hat{q})}$ extension from the total number of errors, which yields the recursion (41). ∎

With this recursive formula we can easily prove by induction that for a given qubit $q$, $k_w(e, q, \boldsymbol{S})$ is (almost) independent of $e$ and $\boldsymbol{S}$.

*Proof of lemma 14.* We consider again a fixed qubit $\hat{q}$ and syndrome $\boldsymbol{S}^* \in \mathbb{F}_2^l$, and prove that the numbers $k_w(e, \hat{q}, \boldsymbol{S}^*)$ are equal for any $e \in \mathsf{P}_1$ such that $\boldsymbol{S}(\boldsymbol{e}^{(\hat{q})}) \neq \boldsymbol{S}^*$. Let $\hat{e} \in \mathsf{P}_1$

with $\boldsymbol{S}((\hat{e})^{(\hat{q})}) \neq \boldsymbol{S}^*$. We consider two different cases, corresponding to different initial conditions for lemma 21. The two cases are given as

(1) $\boldsymbol{S}^* = \boldsymbol{S}((\boldsymbol{e}')^{(\hat{q})})$ for some $e' \neq \hat{e}$,

(2) $\boldsymbol{S}^* \neq \boldsymbol{S}(\boldsymbol{e}^{(\hat{q})})$ for any error $e$ acting on qubit $\hat{q}$.

Consider case 2 first. For $w = 0$, we have $k_0(\hat{e}, \hat{q}, \boldsymbol{S}^*) = 0$ independent of $(\hat{e}, \boldsymbol{S}^*)$ because $\boldsymbol{S}^* \neq \boldsymbol{S}(\boldsymbol{e}^{(\hat{q})}) \forall e \in P_1$. For $w = 1$, $k_1(\hat{e}, \hat{q}, \boldsymbol{S}^*) = 1$ is independent of $(\hat{e}, \boldsymbol{S}^*)$ because the only error $\boldsymbol{e}^{(q)}$ with $\boldsymbol{S}(\hat{e}^{(\hat{q})}\boldsymbol{e}^{(q)}) = \boldsymbol{S}^*$ is $\boldsymbol{S}^{-1}(\boldsymbol{S}(\hat{e}^{(\hat{q})}) \oplus \boldsymbol{S}^*))$ (and this error does not act on $\hat{q}$ because $\boldsymbol{S}^* \neq \boldsymbol{S}(\boldsymbol{e}^{(\hat{q})}) \forall e \in P_1$.). For $w > 1$, the claim follows by induction since the right hand side of the recursive equation in lemma 21 is now independent of $\hat{e}$ and $\boldsymbol{S}^*$. This concludes the proof for case 2. In case 1, the initial conditions are $k_0 = 0, k_1 = 0$. The rest of the proof is analogous. The only caveat is that the last term of (41) now also contains a term $k_{w-1}(e', \hat{q}, \boldsymbol{S}^*)$ for an error $e'$ with $\boldsymbol{S}(\boldsymbol{e}^{(\hat{q})}) = \boldsymbol{S}^*$. But this term can be computed using the same recursive equation, and does not depend on $e$. ■

This finally concludes the proof of lemma 14, and thus also the proof of theorem 5. As mentioned above, lemma 21 can also be used to calculate the numbers $k_w(e, \hat{q}, \boldsymbol{S}^*)$ for the remaining case $\boldsymbol{S}^* = \boldsymbol{S}(\boldsymbol{e}^{(\hat{q})})$. The correct initial conditions are $k_0 = 1, k_1 = 0$.

### C. Additional numerical results

Here, we provide data complementary to the results shown in Sec. III C. In particular, we consider the mean squared error (MSE) of the proposed estimator, and we show results with noisy measurements.

#### 1. MSE of the estimator

First, we demonstrate that the EM estimator achieves the Cramér-Rao bound (CRB). The MSE of the estimator $T$ of a parameter $\theta$ can expressed by the bias-variance decomposition

$$\mathrm{MSE} = \mathrm{bias}(T)^2 + \mathrm{var}(T). \tag{43}$$

Assume we want to estimate the error rates $\boldsymbol{\theta}$ of a code from $m$ independent syndrome observations. Then the covariance of any unbiased estimator $\boldsymbol{T}$ of $\boldsymbol{\theta}$ is bounded by the CRB

$$\mathrm{cov}_{\boldsymbol{\theta}}(\boldsymbol{T}) \geqslant \frac{I(\boldsymbol{\theta})^{-1}}{m}, \tag{44}$$

i.e., $\mathrm{cov}_{\boldsymbol{\theta}}(\boldsymbol{T}) - \frac{I(\boldsymbol{\theta})^{-1}}{m}$ is a positive semi-definite matrix; here, the Fisher information matrix $I$ is defined by

$$I_{i,j} = \mathbb{E}_S\left[\frac{\partial \ln(\mathbb{P}[\boldsymbol{S}|\boldsymbol{\theta}])}{\partial \theta_i} \frac{\partial \ln(\mathbb{P}[\boldsymbol{S}|\boldsymbol{\theta}])}{\partial \theta_j}\right]. \tag{45}$$

In particular, the variance in the estimate of a single parameter is bounded by the diagonal entries of the inverse of the Fisher information. The derivative $\frac{\partial \ln(\mathbb{P}[\boldsymbol{S}|\boldsymbol{\theta}]}{\partial \theta_e^i}$ of the log-likelihood with respect to a parameter $\theta_e^i$ was already computed in (28) as

$$\frac{\partial \ln(p(\boldsymbol{S}|\boldsymbol{\theta})}{\partial \theta_e^i} = \frac{\mathbb{P}[E_i = e \mid \boldsymbol{S}]}{\mathbb{P}[E_i = e]} - \frac{\mathbb{P}[E_i = I \mid \boldsymbol{S}]}{\mathbb{P}[E_i = I]}. \tag{46}$$

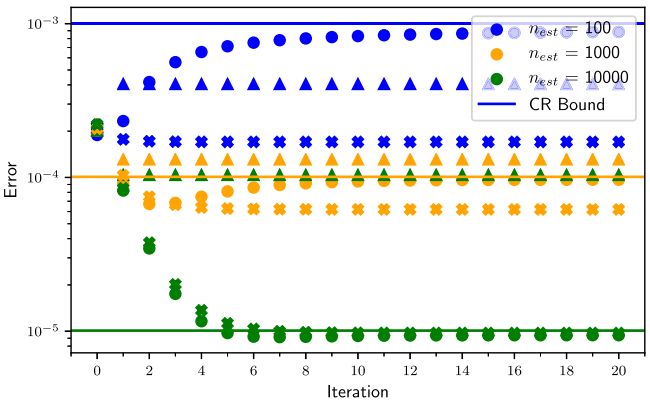



FIG. 6. Comparison of the MSE in $\theta_X^1$ between EM (circles), HEM (triangles) and regularized EM (crosses) for different amounts of estimation data $n_{\mathrm{est}}$ for a good initialization at the first concatenation level. The parameters were $p = 0.13$, $\alpha = 200$, and $n_{\mathrm{concat}} = 1$. $\beta = 200$ was used for the regularized version.

Since the probabilities $\mathbb{P}[e_i|\boldsymbol{S}]$ can be computed using BP, we can numerically evaluate this bound for concrete codes and noise models and compare our estimator to this bound. However, for concatenation levels beyond the first, it was necessary to approximate the expectation value over all syndromes by Monte Carlo sampling. We used $10^6$ samples to do this. As a side note, it is not sufficient to consider the CRB for direct observation of the errors (which is much easier to evaluate). It can be shown that the Fisher information always decreases when post-processing the data, and thus the bounds for syndrome observations must necessarily be higher than for direct measurements of the errors (in our cases the difference was about a factor 2). Finally, it should be noted that in our simulations we have access to the actual error rates which makes it possible to compute the MSE. In a real experiment, one could, for example, consider the variance instead. In our tests, the EM estimator exhibited a squared bias that was very small compared to the variance, such that the variance coincides with the MSE. However, the HEM estimator showed significant bias in some settings. In the following, we always consider the MSE in the estimation of $\theta_X^1$. However, plots for the other parameters look similarly. The MSE was always determined over $10^3$ simulations for each data point. We consider the MSE of the estimation at error rate $p = 0.13$.

For a relatively bad initialization results were already shown in Fig. 3. Here, we consider the situation where an accurate initialization is available, demonstrated by using $\alpha = 200$. An example comparing the MSE at the first concatenation level is shown in Fig. 6. For low data sizes, the initialization is more accurate than the estimate using the data set. In this case, HEM outperforms EM and even beats the CRB (remember that the CRB as it is used here only applies to unbiased estimators). The reason is that HEM has a strong bias towards the initial parameters, which did not decrease with the size of the data sets or the number of iterations in our simulations. At larger data sizes, this bias is detrimental, and it can be seen that EM outperforms HEM. Especially for low numbers of iterations, EM also exhibits some bias towards the

initialization. This can be desirable in case of a good initialization, since it explains why EM also slightly beats the CRB at low data sizes. In particular, we see that at $n_{\text{est}} = 100$ and $n_{\text{est}} = 1000$ EM performs better if a low number of around three iterations is used. Note that a small bias remains at higher iterations, which explains why EM also slightly beats the CRB. Especially interesting is the case $n_{\text{est}} = 1000$, where EM both improves over the initialization and clearly beats the CRB at low numbers of iterations. Since we do not know beforehand after how many iterations the procedure should be stopped, it is sensible to instead regularize the estimator in such a setting, such that it does not converge away from the improved value at low iterations. The regularization is done by introducing a Dirichlet prior

$$\mathbb{P}[\boldsymbol{\theta}^i] = \frac{1}{B(\alpha)} \prod_{e \in \mathsf{P}_1} \left(\theta_e^i\right)^{\beta_e^i} \tag{47}$$

over the initialization, representing information on its accuracy (see Ref. [20]). Here, $\beta_e^i = (1 - (\theta^{(0)})_e^i)\beta$ and the real hyper-parameter $\beta$ controls the strength of the regularization. The effect of this regularization, using $\beta = 200$, is also demonstrated in Fig. 6 (the cross-shaped markers). It can be seen that the regularized EM algorithm converges roughly to the minimum of the unregularized version, which was the desired effect. For large data sizes, the regularization introduces a minimal increase in the estimation error. We also tested regularizing the HEM version in the same manner, but no improvements were obtained. Similar results could be obtained for higher concatenation levels. The main difference is that HEM performs worse at higher levels.

#### 2. *Estimator with measurement noise*

We consider a phenomenological noise model, where Pauli errors occur independently between qubits and bit flips independently on each syndrome bit. The error rates can be different on each data qubit and syndrome bit. The maximum-likelihood decoder, described in Sec. III A, can be easily modified to include these measurement errors. This is done simply by including the measurement errors as additional nodes, connected to the factor corresponding to the syndrome bit that they flip. This does not destroy the tree structure, and thus decoding and determination of marginals can still be done via BP. Using this adapted maximum-likelihood decoder, we can estimate error rates in the same way as described in Sec. III B. It should be noted that we do not consider a fault-tolerant scheme with repeated measurements here, so identification of measurement errors is impossible on the first concatenation level. Similar to the experiments in the main text, we take the data qubits to be affected by a depolarizing channel with error rate $p$ each, and on each syndrome bit the outcome is flipped with probability $p_m$. In Fig. 7, some results are shown. As can be seen in Fig. 7(a), for a bad initialization HEM is unable to improve much over the initialization, while EM still reaches optimal error rates even in the presence of measurement errors, although the amount of iterations required is larger than in the case without measurement errors. The MSE of the estimation was again close to the CRB (not shown here). The case of a better initialization is shown in Fig. 7(b). In this setting, HEM clearly improves over

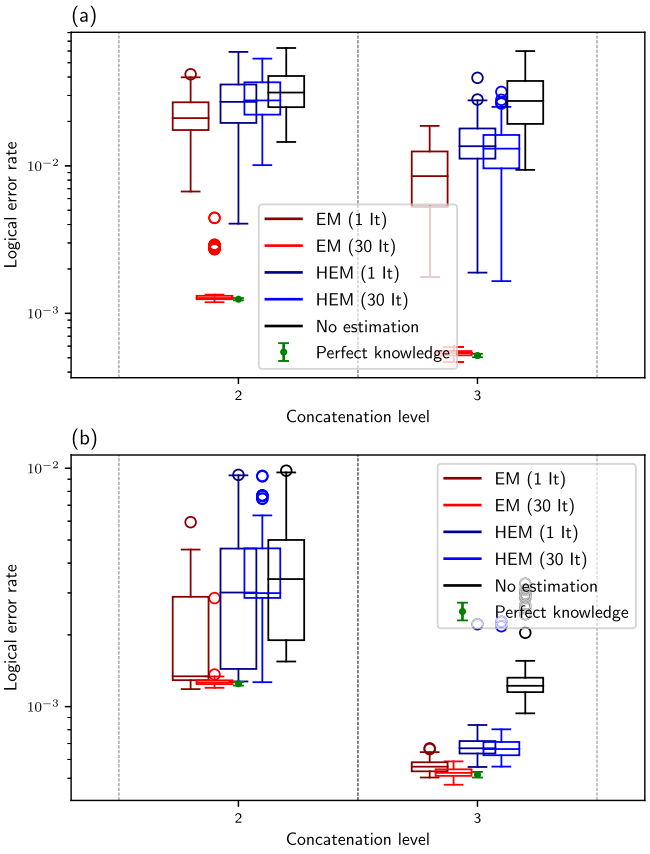



FIG. 7. Logical error rate of the maximum likelihood decoder with measurement errors for (a) $\alpha = 20$ and (b) 200. Also shown are error rates of the perfect knowledge decoder. The parameters were $p = 0.005$, $p_m = 0.005$, and $n_{\text{est}} = 10^4$.

the initialization, especially at higher concatenation levels. It is still outperformed by EM, and the difference is more significant at the second concatenation level. The amount of iterations before convergence of EM is only about 5, compared to about 30 for the bad initialization case.

## V. CONCLUSION

We investigated the estimation of stochastic error models from the syndrome statistics of a quantum error correction code, establishing both theoretical results on parameter identifiability as well as a practical estimation method. The results do not rely on the limit of low error rates, and our estimator outperforms other recently proposed methods [2–4]. Our work opens up a number of new research directions. On the theoretical side, it will be interesting to use our identifiability condition to prove results beyond perfect codes. It might also be possible to extend the result on perfect codes beyond the case of equal rates, since numerical results suggest that this assumption is not crucial. Furthermore, it would be interesting to consider the Cramér-Rao bound as a function of the code size, to estimate how the size of the data set must be scaled for large codes. The proposed estimator could be straightforwardly applied to quantum low density parity check codes, although the

problem arises that belief propagation is no longer exact in this scenario. One could also combine our estimator with methods from Refs. [3,26] to estimate time-dependent error rates and avoid the redecoding overhead, or consider its application to fault-tolerant circuits as was done for the hard assignment method in Ref. [2].

Our PYTHON implementation of the estimator is available on GitHub [27].


## ACKNOWLEDGMENTS

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - Cluster of Excellence Matter and Light for Quantum Computing (ML4Q) EXC 2004/1 - 390534769. Plots were created using the MATPLOTLIB [28] library. The simulations made use of the NUMPY [29] and NUMBA [30] PYTHON packages.



[1] Y. Fujiwara, Instantaneous quantum channel estimation during quantum information processing, arXiv:1405.6267 [quant-ph].

[2] A. G. Fowler, D. Sank, J. Kelly, R. Barends, and J. M. Martinis, Scalable extraction of error models from the output of error detection circuits, arXiv:1405.1454 [quant-ph].

[3] M.-X. Huo and Y. Li, Learning time-dependent noise to reduce logical errors: real time error rate estimation in quantum error correction, New J. Phys. **19**, 123032 (2017).

[4] J. R. Wootton, Benchmarking near-term devices with quantum error correction, Quantum Sci. Technol. **5**, 044004 (2020).

[5] J. Florjanczyk and T. A. Brun, In-situ adaptive encoding for asymmetric quantum error correcting codes, arXiv:1612.05823 [quant-ph].

[6] S. T. Spitz, B. Tarasinski, C. W. J. Beenakker, and T. E. O'Brien, Adaptive weight estimator for quantum error correction in a time-dependent environment, Adv. Quantum Technol. **1**, 1870015 (2018).

[7] J. Combes, C. Ferrie, C. Cesare, M. Tiersch, G. J. Milburn, H. J. Briegel, and C. M. Caves, In-situ characterization of quantum devices with error correction, arXiv:1405.5656 [quant-ph].

[8] J. Kelly, R. Barends, A. G. Fowler, A. Megrant, E. Jeffrey, T. C. White, D. Sank, J. Y. Mutus, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, E. Lucero, M. Neeley, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, A. Vainsencher, J. Wenner, and J. M. Martinis, Scalable in situ qubit calibration during repetitive error detection, Phys. Rev. A **94**, 032321 (2016).

[9] R. Laflamme, C. Miquel, J. P. Paz, and W. H. Zurek, Perfect Quantum Error Correcting Code, Phys. Rev. Lett. **77**, 198 (1996).

[10] A. Steane, Multiple-particle interference and quantum error correction, Proc. R. Soc: Lond. A. **452**, 2551 (1996).

[11] H. Bombin and M. A. Martin-Delgado, Topological Quantum Distillation, Phys. Rev. Lett. **97**, 180501 (2006).

[12] A. Zia, J. P. Reilly, and S. Shirani, Distributed parameter estimation with side information: A factor graph approach, in *Proceedings of the 2007 IEEE International Symposium on Information Theory, Nice, France* (IEEE, Piscataway, NJ, 2007), pp. 2556–2560.

[13] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, 10th ed. (Cambridge University Press, USA, 2011).

[14] F. Gaitan, *Quantum Error Correction and Fault Tolerant Quantum Computing* (Taylor & Francis, 2008).

[15] D. Gottesman, Class of quantum error-correcting codes saturating the quantum hamming bound, Phys. Rev. A **54**, 1862 (1996).

[16] J. Bierbrauer and Y. Edel, Quantum twisted codes, J. Comb. Des. **8**, 174 (2000).

[17] D. Poulin, Optimal and efficient decoding of concatenated quantum block codes, Phys. Rev. A **74**, 052333 (2006).

[18] A. P. Dempster, N. M. Laird, and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. Royal Stat. Soc. **39**, 1 (1977).

[19] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning* (The MIT Press, Cambridge, 2009).

[20] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)* (Springer-Verlag, Berlin, Heidelberg, 2006).

[21] B. Rahn, A. C. Doherty, and H. Mabuchi, Exact performance of concatenated quantum codes, Phys. Rev. A **66**, 032304 (2002).

[22] H. White, Maximum likelihood estimation of misspecified models, Econometrica **50**, 1 (1982).

[23] Y. Fujiwara, Ability of stabilizer quantum error correction to protect itself from its own imperfection, Phys. Rev. A **90**, 062304 (2014).

[24] A. Ashikhmin, C. Lai, and T. A. Brun, Quantum data-syndrome codes, IEEE J. Sel. Areas Commun. **38**, 449 (2020).

[25] N. Delfosse, B. W. Reichardt, and K. M. Svore, Beyond single-shot fault-tolerant quantum error correction, arXiv:2002.05180 [quant-ph].

[26] O. Cappé and E. Moulines, On-line expectation-maximization algorithm for latent data models, J. Royal Stat. Soc. **71**, 593 (2009).

[27] T. Wagner, D. Bruß, H. Kampermann, and M. Kliesch, Noise Estimation From Syndromes, Optimal noise estimation from syndrome statistics of quantum codes, GitHub repository, https://github.com/TWagner2/NoiseEstimationFromSyndromes (2020).

[28] J. D. Hunter, Matplotlib: A 2d graphics environment, Comput. Sci. Eng. **9**, 90 (2007).

[29] T. Oliphant, *NumPy: A guide to NumPy* (Trelgol Publishing, USA, 2006).

[30] S. K. Lam, A. Pitrou, and S. Seibert, Numba: A llvm-based python jit compiler, in *LLVM '15* (Association for Computing Machinery, New York, NY, USA, 2015).

# C

# Pauli Channels can be Estimated From Syndrome Measurements In Quantum Error Correction

| | |
|---|---|
| Title: | Pauli channels can be estimated from syndrome measurements in quantum error correction |
| Authors: | Thomas Wagner, Hermann Kampermann, Dagmar Bruß and Martin Kliesch |
| Journal: | Quantum |
| Publication status: | Published |
| Contribution by TW: | First author (input approx. 85%) |

This publication corresponds to reference [Wag+22a]. A summary of its contents is presented in chapter 6.

The research objective was devised by MK and me as a continuation of my previous work. In particular, MK suggested the focus on the question of identifiability. The project was regularly discussed by all authors. HK pointed me to useful literature about Schur-complements. All main results and their proofs were developed by me. During this, I regularly discussed my approaches with MK. The proofs were carefully checked and corrected by MK, and discussed with HK and DB. I wrote the initial draft of the manuscript, with significant contributions by MK. The manuscript was then proofread and improved by all my co-authors.

# Pauli channels can be estimated from syndrome measurements in quantum error correction

Thomas Wagner, Hermann Kampermann, Dagmar Bruß, and Martin Kliesch

Institut für Theoretische Physik, Heinrich-Heine-University Düsseldorf, Germany



The performance of quantum error correction can be significantly improved if detailed information about the noise is available, allowing to optimize both codes and decoders. It has been proposed to estimate error rates from the syndrome measurements done anyway during quantum error correction. While these measurements preserve the encoded quantum state, it is currently not clear how much information about the noise can be extracted in this way. So far, apart from the limit of vanishing error rates, rigorous results have only been established for some specific codes.

In this work, we rigorously resolve the question for arbitrary stabilizer codes. The main result is that a stabilizer code can be used to estimate Pauli channels with correlations across a number of qubits given by the pure distance. This result does not rely on the limit of vanishing error rates, and applies even if high weight errors occur frequently. Moreover, it also allows for measurement errors within the framework of quantum data-syndrome codes. Our proof combines Boolean Fourier analysis, combinatorics and elementary algebraic geometry. It is our hope that this work opens up interesting applications, such as the online adaptation of a decoder to time-varying noise.

## 1 Introduction

Quantum error correction is an essential part of most quantum computing schemes. It can be significantly improved if detailed knowledge about the noise affecting a device is available, as it is possible to optimize codes for specific noise [1, 2]. A prominent example is the XZZX-surface code [3]. Furthermore, common decoding algorithms, such as minimum weight matching [4–7] and belief propagation (see e.g. [8]), can incorporate information about error rates to return more accurate corrections. Other examples of decoders that can incorporate information about error rates are weighted union find [9] and tensor network decoders [10, 11]. The latter can also deal with correlated noise models, but are relatively slow. In the context of stabilizer codes, noise is usually modeled using Pauli channels, which are simple to understand and simulate. However, Pauli noise is more than a mere toy model, since randomized compiling can be used to project general noise onto a Pauli channel [12], which has also been demonstrated experimentally [13]. Furthermore, it is known that quantum error correction decoheres noise on the logical level [14]. Consequently, there has been much interest and progress in the estimation of Pauli channels [15–18]. Complementary to the standard benchmarking approaches, it has been suggested to perform (online) estimation of channels just from the syndromes of a quantum error correction code itself [2, 7, 19–24]. Such a scheme uses only measurements that do not destroy the logical information. It is thus suited for online adaptation of a decoder to varying noise, for example by adapting weights in a minimum

Thomas Wagner: thomas.wagner@uni-duesseldorf.de
Martin Kliesch: science@mkliesch.eu

weight matching decoder [7, 21]. It furthermore results in a noise model that can be directly used by the decoder [20]. Experimentally, online optimization of control parameters in a 9-qubit superconducting quantum processor has been demonstrated in an experiment by Google [25].

Since the state of the logical qubit is not measured, some assumptions are necessary in order for this estimation to be feasible. However, the precise nature of these assumptions is currently not well understood. In the general case, it is unclear for which combinations of noise models and codes the parameters can be identified using only the syndrome information. Apart from heuristics in the limit of very low error rates [20–22], only two special cases have been rigorously treated. For codes admitting a minimum weight matching decoder, such as the toric code, identifiability of a circuit noise model was shown by Spitz *et al.* [7]. In [24], identifiability results for the restricted class of perfect codes were proven.

It is not a priori clear that an estimation of the error rates just from the syndrome statistics should be possible at all for arbitrary stabilizer codes. There are several objections one could raise. For one, as mentioned above, the state of the logical qubit is not measured, so there is only limited information contained in the syndromes. Phrased another way, there are generally exponentially many errors with the same syndrome, which we therefore cannot distinguish. This also implies that the probability of a syndrome is an exponentially large sum of different error rates, and solving such a system for the error rates appears difficult at best. While it has been suggested to simplify the problem by only taking into account the lowest weight error compatible with each syndrome [20–22], this approximation strategy leads to demonstrably sub-optimal estimators [24].

## 2 Results

In this work, we show that the estimation task can be solved for arbitrary stabilizer codes. For any given quantum code, we describe a general class of Pauli channels whose parameters can be identified from the corresponding syndrome statistics. Our results also take into account measurement errors by using the framework of quantum data-syndrome codes [26–28]. We prove that a large amount of information can be extracted from the syndrome statistics. In the following theorem we make this statement more precise in terms of the pure distance, which is defined as the minimum weight of an undetectable error (see Section 2.3 for details). The pure distance measures up to which weight errors can necessarily be distinguished by their syndrome, and for most codes it coincides with the weight of the smallest stabilizer. Since there can be undetectable but logically trivial errors, the pure distance is usually much smaller than the distance. As an example, for the family of toric codes the pure distance is 4 independent of code size. More generally, the pure distance will be constant for any family of quantum low density parity check (LDPC) codes, since the stabilizer weights are constant.

We are interested in estimating a Pauli channel, i.e. a probability distribution over Pauli errors, which describes the new error occurring before each round of error correction (working in a phenomenological noise model). It is commonly assumed that errors on each qubit occur independently, in which case the channel is described by one error rate for each qubit. In this case, we say the noise is uncorrelated. In a more general setting, the Pauli channel could act on many qubits, such that errors on these qubits are not independent. If e.g. 2 qubit errors occur that are not a combination of independent single qubit errors, we say that the noise is correlated over 2 qubits. The corresponding 2 qubit error rates are then not simply a product of the single qubit error rates and must be additionally specified. For example, if $P(X_1X_2) \neq P(X_1)P(X_2)$, the errors on the first two qubits are correlated. This notion of correlations is made precise in Section 2.2.3 for classical and Section 2.3.2 for quantum codes. We now give an informal statement of our main results. A formal statement is given in Theorem 7, which also takes into account many detectable errors beyond the pure distance.

**Theorem 1** (Main result, informal)**.** *A stabilizer code with pure distance $d_p$ can be used to estimate Pauli noise with correlations across up to $\lfloor \frac{d_p-1}{2} \rfloor$ qubits.*

While this result is stated in a non-constructive fashion, its derivation suggests a concrete estimation protocol. We give a first heuristic discussion of the resulting estimators in Section 2.4. A detailed analysis of such estimators is ongoing work.

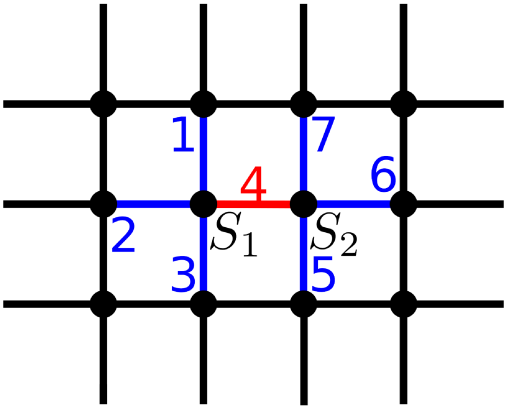



Figure 1: A qubit (edge) and two adjacent star operators (vertices) on the toric code.

The key idea behind Theorem 1 is that the error distribution is fully described by a set of moments. We will show that these moments can be estimated up to their sign by solving a polynomial system of equations. The appearance of multiple discrete solutions, differing in some signs, reflects the symmetries of the problem. For example in the case of single qubit noise, there is one symmetry for each logical operator of the code. However, under the additional mild assumption that all error rates are smaller than $\frac{1}{2}$, all moments must be positive, and thus the estimate is unique. We stress that our result does not rely on the limit of vanishing rates, and still applies in the presence of high weight errors. Perhaps surprisingly, even different errors with the same syndrome can occur frequently, as long as they arise as a combination of independent lower weight errors. Our result is arguably the strongest one can reasonably expect, since error rates cannot be identifiable if multiple independently occurring errors have the same syndrome.

The *adaptive weight estimator* by Spitz *et al.* [7] can be viewed as solving a special instance of the general equation system we present here, which is applicable only for codes that admit a minimum weight matching decoder. This connection is explained in Section 2.1.

Our arguments combine Boolean Fourier analysis, combinatorics and elementary algebraic geometry. Interestingly, a connection between Pauli channel learning and Boolean Fourier analysis was recently pointed out in an independent work by Flammia and O'Donnell [15].

We start our discussion with the motivating example of the toric code with independent Pauli-$X$ errors, where our results take a particularly simple form. We then discuss the general results. We first discuss the setting of classical codes, and later extend the results to quantum codes. This is for ease of exposition, since the presentation is considerably simplified in the classical setting and thus the underlying concepts become more apparent. We stress that, in contrast to the quantum setting, classically one can measure the individual bits of a codeword without destroying the encoded information. Thus, in the classical setting one does not have to rely solely on the syndrome information and can use easier techniques for error rates estimation. However, for a quantum code this is indeed the only information that can be measured without destroying the encoded state, and this is where our results are most relevant. Classically, our approach might still be useful in the setting of distributed source coding [29].

## 2.1 Example: the toric code

As a motivating example for the methods and proofs in the following sections, we derive an estimator for the simple setting of completely uncorrelated bit-flip noise on the toric code. That is, we assume that errors on different qubits occur independently, possibly with a different rate for each qubit, and that only bit-flip (Pauli-$X$) errors occur on each qubit. Thus, we consider the toric code essentially as a classical code. This constitutes a simple alternative derivation of the solution given by Spitz *et al.* [7].

We focus on a single qubit and two adjacent $Z$-stabilizers, as illustrated in fig. 1. Our task is to estimate the error rate $p_4$ of the marked qubit, or equivalently, the expectation value $E(Z_4) = 1 - 2p_4$. Since errors on each qubit are independent, the expectation of a stabilizer measurement

is simply the product of expectations of the adjacent bits. Therefore we obtain the following three equations,

$$\begin{aligned} E(S_1) &= E(Z_1)E(Z_2)E(Z_3)E(Z_4) \\ E(S_2) &= E(Z_4)E(Z_5)E(Z_6)E(Z_7) \\ E(S_1S_2) &= E(Z_1)E(Z_2)E(Z_3)E(Z_5)E(Z_6)E(Z_7)\,. \end{aligned}$$

This system admits a straightforward solution for the expectation of $Z_4$,

$$E(Z_4) = \pm\sqrt{\frac{E(S_1)E(S_2)}{E(S_1S_2)}}\,. \tag{1}$$

This coincides with the solution given by Spitz *et al.* [7, eq. (14)], as explained in Appendix E. Notice that there is a choice of sign, which corresponds to deciding whether $p_4 > \frac{1}{2}$ or $p_4 < \frac{1}{2}$. However, under the assumption $p_4 < \frac{1}{2}$ the solution is unique.

## 2.2 Identifiablity for classical codes

Let us now turn to the general setting, first for classical codes. We are interested in whether an error distribution can be estimated from repeated syndrome measurements alone. This is certainly not possible for completely arbitrary noise, since we do not measure the state of the logical qubit. However, we will show that for Pauli (and measurement) noise with limited correlations, the estimation is possible. As mentioned above, we will start with the setting of classical codes and perfect syndrome measurements. Then, we will show how to extend these results to quantum (data-syndrome) codes.

The key insight underlying our proof is that the estimation problem is best phrased in terms of moments instead of error rates. The proof then proceeds in the following steps. First, we notice that Fourier coefficients of the error distribution correspond to moments, and see that some of these moments can be estimated from the syndrome statistics. Then, we show that under certain independence assumptions, the full error distribution is characterized completely by a set of low weight transformed moments. We find that these transformed moments are related to the measured moments via a polynomial equation system. This system is described by a coefficient matrix $D$ whose rows essentially correspond to elements of the dual code. We then use local randomness properties of the dual code to find an explicit expression for the symmetric squared coefficient matrix $D^TD$, and finally show that $D^TD$ has full rank by using iterated Schur complements. This implies that the equation system has discrete solutions.

### 2.2.1 Notation

We use the short-hand notation $[n] \coloneqq \{1, \dots, n\}$ for the set of the first $n$ natural numbers. For any set $A$, we denote its powerset as $2^A \coloneqq \{B \mid B \subseteq A\}$. The field with two elements is denoted by $\mathbb{F}_2$. Often we will use $\mathbb{F}_2^n$ as a vector space over that field. That is, for $a, b \in \mathbb{F}_2^n$ the sum $a + b$ is understood to be taken component-wise modulo 2. All vectors are to be understood as column vectors. By $\mathrm{wt}(a) \coloneqq |\{i : a_i \neq 0\}|$ we denote the *weight* of $a$. For any logical statement $X$ we denote by $1[X]$ the *Iverson bracket* of $X$, which assumes the value 1 if $X$ is true and 0 otherwise. Naturally, trivial products, i.e. products over the empty set, are set to 1 as in $\prod_{x\in\emptyset} f(x) \coloneqq 1$. By $I_k$ we denote the $k \times k$ identity matrix and suppress $k$ when it can be inferred from the context.

### 2.2.2 Classical codes and Boolean Fourier analysis

Let us start with some basic elements of Boolean Fourier analysis. A detailed review of the topic is given in [30] (note that we use a different normalization convention here). We frequently identify a vector $s \in \mathbb{F}_2^n$ with its *indicator set* $\{i \in [n] \,:\, s_i \neq 0\}$. For example, for $s = (0, 1, 0, 1, 1, 0) \in \mathbb{F}_2^6$ we also write

$$s = \{2, 4, 5\} \subseteq [6]\,, \tag{2}$$

such that we can write $4 \in s$. For each $s \in \mathbb{F}_2^n$, we define the *parity function*

$$\chi_s : \mathbb{F}_2^n \to \{-1, +1\},$$
$$\chi_s(e) = (-1)^{s \cdot e} = (-1)^{\sum_{i \in s} e_i},$$

which is the group character of the abelian group $(\mathbb{F}_2^n, +)$. For a function $f : \mathbb{F}_2^n \to \mathbb{R}$, its *Boolean Fourier transform* $\widetilde{f}$ is the function

$$\widetilde{f} : 2^{[n]} \to \mathbb{R}, \quad \widetilde{f}(s) = \sum_{e \in \mathbb{F}_2^n} f(e)\chi_s(e). \tag{3}$$

The Boolean Fourier transform is also known under the name Walsh-Hadamard transformation. However, there are different conventions which lead to transforms with different orderings of bit strings and different notations for the characters $\chi_s$. This transformation is invertible, and the inverse is given by

$$f(e) = \frac{1}{2^n} \sum_{s \subseteq [n]} \widetilde{f}(s)\chi_s(e). \tag{4}$$

For two functions $f, g : \mathbb{F}_2^n \to \mathbb{R}$, their *Boolean convolution* $f * g$ is defined by

$$f * g(e) = \sum_{e' \in \mathbb{F}_2^n} f(e)g(e + e'). \tag{5}$$

As expected, convolutions become products under Boolean Fourier transform,

$$\widetilde{f * g} = \widetilde{f}\,\widetilde{g}. \tag{6}$$

A classical linear code $C$, encoding $k$ bits into $n$ bits, is a $k$-dimensional subspace of $\mathbb{F}_2^n$. It can be described by its parity check matrix $H \in \mathbb{F}_2^{(n-k)\times n}$, whose rows span the dual code $C^\perp = \{a \in \mathbb{F}_2^n \ : \ a \cdot c = 0 \ \forall c \in C\}$. $C^\perp$ can alternatively be interpreted as the set of parity functions which are 1 on all codewords. When an error $e \in \mathbb{F}_2^n$ occurs, the outcomes of the parity measurements $H$ can be summarized by the *syndrome*

$$\mathcal{O}(e) \coloneqq He. \tag{7}$$

From $\mathcal{O}(e)$, we can calculate the value of any parity measurement $s \in C^\perp$. We will assume that repeated rounds of syndrome measurements and corrections are performed. In each round, an error occurs according to the *error distribution* $P : \mathbb{F}_2^n \to [0, 1]$. We call the corresponding distribution of syndromes the *syndrome statistics*. The error correction capabilities of a classical linear code are indicated by its *distance*, which is defined as the smallest weight of an undetectable error, which is the same as the smallest weight of a codeword,

$$d \coloneqq \min\{\mathrm{wt}(e) \ : \ e \in C \setminus \{0\}\}. \tag{8}$$

### 2.2.3 Moments and noise model

We denote the Fourier transform of the error distribution $P : \mathbb{F}_2^n \to [0, 1]$ with $E$, i.e. for each $a \subset [n]$

$$E(a) \coloneqq \sum_{e \in \mathbb{F}_2^n} (-1)^{a \cdot e} P(e) = \widetilde{P}(a). \tag{9}$$

We interpret this as a *moment* of the distribution $P$, since $E(a)$ is exactly the expectation value of the parity measurement $a$ if one were to measure it repeatedly on errors distributed according to $P$. In particular, for $s \in C^\perp$, the corresponding moment $E(s)$ can be computed from the measured syndrome statistics, i.e. from the empirically measured frequency with which each syndrome occurs in repeated rounds of error correction. We will always assume $E(a) \neq 0$ for all $a \subseteq [n]$, which is, for example, fulfilled if $P(0) > 0.5$. Thus, our task is to find the distribution $P$, given some of its Fourier coefficients.

We will show that this task is feasible if there is some independence between errors on different subsets of bits, such that there are no correlations across a large number of qubits at once. To formalize this idea, we introduce a set of *channel supports* $\Gamma \subseteq 2^{[n]}$ (we will consider a concrete choice of $\Gamma$ later). For each $\gamma \in \Gamma$, there is an error channel that acts only on this subset, i.e. its error distribution $P_\gamma : \mathbb{F}_2^n \to [0,1]$ is only supported on errors $e \in \mathbb{F}_2^n$ that are 0 outside of $\gamma$. Equivalently, identifying vectors with their indicator set, we can view this as a distribution $P_\gamma : 2^{[n]} \to [0,1]$ which is only supported on $2^\gamma$. Since all the individual channels act independently, and the total error is the sum of the individual errors, the total distribution of errors is then given by

$$P = \mathop{\circledast}_{\gamma\in\Gamma} P_\gamma \,, \tag{10}$$

with the convolution (5). Physically, this expresses the action of several independent sources of errors, each only acting on a limited number of qubits. If $\Gamma = \{\{1\}, \dots, \{n\}\}$, this reduces to the commonly used model where each bit is affected by an independent channel. As a note, mathematically, this model corresponds to a convolutional factor graph, as introduced by Mao and Kschischang [31].

This representation in terms of individual channels is an over-parametrization of the total distribution if some supports in $\Gamma$ overlap. A first idea for a non-redundant distribution would be to use the set of all moments, but these are not all independent. Therefore, we define transformed moments $F(a)$ via a (generalized) inclusion-exclusion computation. Heuristically, we want to divide out all correlations on proper subsets of $a$ in order to capture only correlations across the full size of $a$. This bears some similarity to the canonical parametrization of Markov random fields, see [32]. For $b \subseteq a \subseteq [n]$, the *Möbius function* (on the partially ordered set $2^{[n]}$) is given by

$$\mu(a,b) = (-1)^{|a\setminus b|} = (-1)^{|a|+|b|} \,. \tag{11}$$

For each $a \subseteq [n]$, we define a transformed moment $F(a)$ via the *inclusion-exclusion transform*

$$F(a) \coloneqq \prod_{b\subseteq a} E(b)^{\mu(a,b)} \,. \tag{12}$$

It follows from the (multiplicative) generalized inclusion-exclusion principle (e.g. [33][Theorem 5.3], [34, Theorem A.2.4]) that the inverse of this transformation is given by

$$E(a) = \prod_{b\subseteq a} F(b) \,. \tag{13}$$

Now we will show that, depending on the choice of $\Gamma$, a small subset of transformed moments is already sufficient to uniquely specify the distribution. Remember $E_\gamma(a) = \widetilde{P}_\gamma(a)$ from (9). Since $P$ is given by the convolution (10) of all individual error distributions $P_\gamma$ we have,

$$E(a) = \prod_{\gamma\in\Gamma} E_\gamma(a\cap\gamma) \,, \tag{14}$$

where we used that $P_\gamma$ is only supported on $2^\gamma$, and thus $E_\gamma(a) = \tilde{P}_\gamma(a) = \tilde{P}_\gamma(a\cap\gamma) = E_\gamma(a\cap\gamma)$.

We can now express the transformed moments by the parameters of the individual channels. The proof is given in Appendix A.

**Lemma 2.** *For $\Gamma \subseteq 2^{[n]}$ let $E : 2^{[n]} \to \mathbb{R}$ satisfy (14). Then, the inclusion-exclusion transform (12) satisfies*

$$F(a) = \prod_{\gamma\in\Gamma : a\subseteq\gamma} \prod_{b\subseteq a} E_\gamma(b)^{\mu(a,b)} \tag{15}$$

*for all $a \subseteq [n]$. In particular, $F(a) = 1$ if there is no $\gamma \in \Gamma$ such that $a \subseteq \gamma$.*

Notice that trivially $F(\emptyset) = 1$. Thus, in conclusion, it suffices to determine the transformed moments $F(a)$ for $a$ in

$$\hat{\Gamma} \coloneqq \{\emptyset \neq a \subseteq \gamma \ : \ \gamma \in \Gamma\} \,. \tag{16}$$

The standard moments (13) are then determined by

$$E(a) = \prod_{b\in\hat{\Gamma}:b\subseteq a} F(b)\,. \tag{17}$$

Finally, the error distribution $P$ is determined from the standard moments by applying the inverse Fourier transform (4) to (9).

### 2.2.4 Identifiability and binomial systems

As discussed in the previous section, learning the transformed moments $(F(a))_{a\in\hat{\Gamma}}$ is sufficient to uniquely determine the error distribution. However, the only thing we can measure are the standard moments $(E(s))_{s\in C^\perp}$ corresponding to elements of the dual code, since these fully describe the syndrome statistics. Thus, learning the error distribution boils down to determining the transformed moments $(F(a))_{a\in\hat{\Gamma}}$ from this information. In particular, the error distribution is identifiable from the syndrome statistics if and only if the following system of polynomial equations admits a unique solution,

$$\begin{aligned} &\text{Given } (E(s))_{s\in C^\perp} \text{ find } (F(b))_{b\in\hat{\Gamma}} \text{ satisfying} \\ &E(s) = \prod_{b\in\hat{\Gamma}:b\subseteq s} F(b) \qquad \forall s\in C^\perp\setminus\{0\}\,, \end{aligned} \tag{18}$$

where we have omitted the trivial equation arising from the zero element of the dual code. This is a *binomial* system, i.e. each equation only has two terms, a constant on the left-hand side and a monomial on the right-hand side. In the notation of [35], the binomial system (18) can be expressed by the *coefficient matrix* $D$ whose rows are labeled by elements of $C^\perp\setminus\{0\}$ and whose columns are labeled by elements of $\hat{\Gamma}$. The entry $D[s,a]$ is the exponent of $F(a)$ in the monomial on the right-hand side of equation (18) for $E(s)$. We have transposed the notation comparing to [35]. Explicitly this means that the $\left|C^\perp\setminus\{0\}\right|\times\left|\hat{\Gamma}\right|$-matrix $D$ has the elements

$$D[s,a] = \begin{cases} 1 & \text{if } a\subseteq s \\ 0 & \text{otherwise} \end{cases}\,, \tag{19}$$

and the system is given by $\vec{E} = \vec{F}^D$ [35, Eq. (3)], where $\vec{E}$ is the vector containing the elements $E(s)$ and $\vec{F}$ the elements $F(a)$, as appearing in the system (18). In the special case of single bit noise, i.e. $\Gamma = \hat{\Gamma} = \{\{i\} : i\in[n]\}$, the rows of the coefficient matrix $D$ are exactly the dual codewords $s\in C^\perp$.

For now, let us assume that $D$ has full rank. It then follows from the theory of binomial system solving [35, Proposition 2] that the system (18) has a finite number of solutions, and these solutions only differ by multiplying some parameters with complex roots of unity. Since we are only interested in real solutions, this means we can determine the transformed moments $F(a)$, and thus also the standard moments $E(a)$, up to a sign. Thus, if we restrict all moments to be positive, the error distribution is uniquely determined by the syndromes. A simple condition for all moments to be positive is that the error probabilities of all channels are smaller than $\frac{1}{2}$, i.e. $P_\gamma(0) > \frac{1}{2}$ for all $\gamma\in\Gamma$. This implies $E_\gamma(a) > 0$ and thus $E(a) > 0$ for all $a\subseteq[n]$. In conclusion, the error distribution can be estimated uniquely from the syndromes, assuming that $D$ has full rank and that the error probability of each channel is smaller than $\frac{1}{2}$.

Let us stress that the appearance of multiple solutions, differing by the signs of some moments, reflects actual symmetries of the syndrome statistics as a function of the error rates. For example, consider again case of independent single bit errors, such that $\Gamma = \hat{\Gamma} = \{\{i\} : i\in[n]\}$. Then the transformed moments $F(\{i\})$ are equal to the standard moments $E(\{i\})$, and the rows of $D$ are simply the elements of the dual code. Thus, each row of $D$ has even overlap with every codeword $c\in C$. Thus, by (18), flipping the signs of all moments $(E(\{i\}))_{i\in c}$ on the support of the codeword $c$ does not change the measured moments $(E(s))_{s\in C^\perp}$. Since $E(\{i\}) = 1-2p_i$, where $p_i$ is the rate of errors on the $i$-th qubit, this simply means that flipping the error rates around $\frac{1}{2}$ on a codeword does not affect the syndrome statistics, i.e. we cannot distinguish these two sets of error rates. This observation shows that each codeword corresponds to a symmetry of the identifiability problem.

### 2.2.5 The rank of the coefficient matrix

We will now establish the most general set of error distributions for which unique identification of the error rates is possible. By the discussion in the previous section, this corresponds to finding the most general choice of $\Gamma$ for which the system (18) is solvable, i.e. for which the coefficient matrix $D$ is of full rank.

Denote the sets of bits that only support *detectable* errors as

$$\Gamma^{(D)} \coloneqq \{\gamma \subseteq [n] \ : \ \mathcal{O}(e) \neq 0 \ \forall e \subseteq \gamma\}\,, \tag{20}$$

where $\mathcal{O}(e)$ is the syndrome (7) of $e$. It is clear that, if one of the channels $\gamma \in \Gamma$ supports an undetectable error, one cannot estimate the corresponding error rate. Thus, identifiability can only hold if $\gamma \in \Gamma^{(D)}$. Similarly, if two different channels $\gamma_1, \gamma_2$ contain two errors with the same syndrome, the syndrome statistics only depends on the combined rate of those errors and thus the rates are not identifiable. Identifiability of the noise channel can only hold if $\gamma_1 \cup \gamma_2 \in \Gamma^{(D)}$ for all $\gamma_1, \gamma_2 \in \Gamma$. We will show that these are in fact the only restriction that must be fulfilled.

**Theorem 3** (Classical identifiability condition)**.** *Consider a classical code with $n$ bits subject to noise with an error distribution described by channel supports $\Gamma \subseteq 2^{[n]}$. Then the coefficient matrix $D$ defined in (19) is of full rank if and only if $\gamma_1 \cup \gamma_2 \in \Gamma^{(D)}$ for all $\gamma_1, \gamma_2 \in \Gamma$.*

The proof is given in Sections 2.2.6. Even a combination of two channels must not support an undetectable error. We can also give an equivalent condition in terms of $\hat{\Gamma}$, which represents the set of errors that can occur "independently". As explained in Appendix B, the assumption $\gamma_1 \cup \gamma_2 \in \Gamma^{(D)}$ for all $\gamma_1, \gamma_2 \in \Gamma$ is equivalent to $\mathcal{O}(e_1) \neq 0$ and $\mathcal{O}(e_1 + e_2) \neq 0$ for all $e_1, e_2 \in \hat{\Gamma}$ (viewed as binary vectors). In other words, the error distribution is identifiable if undetectable errors and errors that have the same syndrome only occur as a combination of independent errors. We stress that this is substantially weaker than the assumption that errors with the same syndrome never occur (which was made in some previous works such as [20, 21]). Indeed, in the total error distribution different errors with the same syndrome can occur frequently, and we are still able to identify the error rates. We only assume that such errors arise as a combination of independent errors.

The most important example is the following. Consider an error model where there is an independent channel on every subset of $t$ bits, i.e. choose $\Gamma$ as

$$\Gamma = \Gamma_t \coloneqq \{\gamma \subseteq [n] \ : \ |\gamma| = t\} \tag{21}$$

$$\hat{\Gamma} = \Gamma_{\leq t} \coloneqq \{e \subseteq [n] \ : \ 0 < |e| \leq t\}\,. \tag{22}$$

This means we only have correlations across at most $t$ bits. Then, Theorem 3 implies that the error distribution is identifiable as long as $d \geq 2t + 1$.

**Corollary 4** (Distance based identifiability condition)**.** *If the noise is described by channel supports $\Gamma = \Gamma_t$, the coefficient matrix $D$ is of full rank if the code has distance $d \geq 2t + 1$.*

*Proof.* For a code with distance $d$, $\Gamma^{(D)}$ contains every set of size at most $d - 1$. □

For $t = 1$, we see that error rates of the standard single bit noise model can be identified as long as $d \geq 3$. Informally, identification is possible if error correction is possible.

### 2.2.6 Orthogonal array properties of the coefficient matrix

This section is devoted to the proof of Theorem 3. Readers not interested in the proofs and mathematical techniques can skip to Section 2.3 for the results in the quantum setting. The first part of the proof is based on local randomness properties of the dual code. The elements of the dual code form a so-called *orthogonal array*. Since the entries of the coefficient matrix $D$ are related to the dual code, we can use this property to derive an explicit expression for the symmetric squared coefficient matrix $D^T D$. The proof is then finished by computing the rank of $D^T D$, which is possible with the help of combinatorial results derived in the next section.

By $T$ we denote the $|C^\perp| \times n$-matrix formed by all elements in the span of the rows of the parity check matrix $H$, i.e. the rows of $T$ are exactly the elements of $C^\perp$. It is known that the rows of $T$ look "locally uniformly random" on any subset of up to $d-1$ bits. One says that $T$ is an orthogonal array of strength $d-1$. This property is relatively easy to see for linear codes (e.g. [36]) and was shown for general (non-linear) classical codes by Delsarte [37]. We use a slightly extended version of the result for linear codes using the set $\Gamma^{(D)}$ from (20) instead of the distance. A proof is given in Appendix C.

**Lemma 5.** *Let $\gamma \in \Gamma^{(D)}$. In the restriction $T^{|\gamma}$ of $T$ to columns in $\gamma$, every bit-string appears equally often as a row.*

In other words, the rows of $T$ look locally uniformly random on any choice of bits that only supports detectable errors.

In Appendix D.2 we prove the following statement, as a corollary of Lemma 11.

**Lemma 6** (Positive-definiteness of intersection matrix)**.** *Let $M_t$ be the matrix whose rows and columns are labeled by the elements of $\Gamma_{\leq t}$ (from (22)) and which is defined entry-wise by*

$$M_t[a,b] = 2^{|a\cap b|}\,. \tag{23}$$

*Then $M_t$ is positive-definite.*

We call the matrix $M_t$ the *intersection matrix*. The dimensions of $M_t$ depend on $n$, but we do not make this dependence explicit in our notation. Using Lemma 6, we can finish the proof of Theorem 3.

*Proof of Theorem 3.* We will prove that the coefficient matrix $D$ from (19) has full rank (over $\mathbb{R}$) by proving that $D^T D$ has full rank. First, note that by the definition (20) of $\Gamma^{(D)}$, for any $\gamma \in \Gamma^{(D)}$, all subsets $a \subseteq \gamma$ are also elements of $\Gamma^{(D)}$. Thus, the assumption $\gamma_1 \cup \gamma_2 \in \Gamma^{(D)}$ from Theorem 3 implies $a \cup b \in \Gamma^{(D)}$ for all $a \cup b \in \hat{\Gamma}$, by the definition (16) of $\hat{\Gamma}$. The dot product of the columns of $D$ labeled by $a, b \in \hat{\Gamma}$ (as binary vectors) is the number of elements $s \in C^\perp \setminus \{0\}$ such that $a \cup b \subseteq s$. Thus, by Lemma 5, this number is $|C^\perp|\, 2^{-|a\cup b|}$, i.e. $D^T D[a,b] = |C^\perp|\, 2^{-|a\cup b|}$. We can write this as $D^T D[a,b] = |C^\perp|\, 2^{-|a|-|b|+|a\cap b|}$. Re-scaling the rows and columns leads to the modified matrix $(D^T D)'[a,b] = 2^{|a\cap b|}$. Let us denote $t = \max\{|a| \,:\, a \in \hat{\Gamma}\}$. Then $(D^T D)'$ is a principal sub-matrix of the intersection matrix $M_t$ defined in (23). By Lemma 6, $M_t$ is positive-definite. As a principal sub-matrix of a positive-definite matrix, $(D^T D)'$ is then also positive-definite and, in particular, has full rank. This implies that $D$ has full rank, which finishes the proof of Theorem 3. □

## 2.3 Extension to the quantum case

Now will consider the quantum setting, starting with a short overview of stabilizer codes. We also explain the concept of quantum data-syndrome codes [26–28], following Ashikhmin *et al.* [26], which allow for a unified treatment of data and measurement errors. Then, we state and explain our main result Theorem 7, which is the formal version of Theorem 1. Finally, we explain how to prove this theorem by extending our arguments from the classical to the quantum case.

### 2.3.1 Preliminaries

The Pauli group $\mathcal{P}^n$ on $n$ qubits is the group of *Pauli strings* generated by the Pauli operators $\{I, X, Y, Z, \}$ with phases,

$$\mathcal{P}^n = \left\{ \epsilon \bigotimes_{i=1}^{n} P_i \mid \epsilon \in \{\pm 1, \pm i\},\ P_i \in \{I, X, Y, Z\} \right\}.$$

The *weight* $\mathrm{wt}(P)$ of a Pauli string $P = \epsilon \bigotimes_{i=1}^n P_i$ is the number of non-identity components $P_i$. Modding out phases, one obtains the *effective Pauli group*

$$\mathsf{P}^n = \mathcal{P}^n / \{\pm 1, \pm i\}\,. \tag{24}$$

We define the *symplectic inner product* $\langle \cdot, \cdot \rangle_{\mathsf{P}}$ on $\mathsf{P}^n$ by

$$\langle e, e' \rangle_{\mathsf{P}} = \begin{cases} 1, & e \text{ and } e' \text{ anti-commute in } \mathcal{P}^n \\ 0, & e \text{ and } e' \text{ commute in } \mathcal{P}^n \end{cases} ; \tag{25}$$

note that this expression is well-defined since the commutation relation does not depend on the choice of representatives in $\mathcal{P}^n$. We identify $\mathsf{P}^1$ with $\mathbb{F}_2^2$ via the *phase space representation*,

$$X \mapsto (1,0),\ Z \mapsto (0,1),\ Y \mapsto (1,1), \tag{26}$$

which extends coordinate-wise to define a group isomorphism $\mathsf{P}^n \to \mathbb{F}_2^{2n}$. Thus, an element of $\mathsf{P}^n$ is represented by $n$ "$X$-bits" and $n$ "$Z$-bits". Explicitly, $X^{x_1}Z^{z_1} \otimes \cdots \otimes X^{x_n}Z^{z_n}$ is mapped to $(x_1, \ldots, x_n, z_1, \ldots, z_n)^T$. For example

$$X \otimes I \otimes Z \otimes Y \mapsto (1,0,0,1,\ 0,0,1,1)^T . \tag{27}$$

This identification will allow for the application of Boolean Fourier analysis to the Pauli group. We denote the operation of swapping the $X$-bits and $Z$-bits by a bar, $\overline{(x_1, \ldots, x_n, z_1, \ldots, z_n)} = (z_1, \ldots, z_n, x_1, \ldots, x_n)$. The symplectic inner product (25) then corresponds to

$$\langle e, e' \rangle_{\mathsf{P}} = \overline{e} \cdot e' , \tag{28}$$

where the dot product is evaluated in $\mathbb{F}_2^{2n}$, i.e. modulo 2. We define the *Pauli weight* $\mathrm{wt}_{\mathsf{P}}(e)$ of $e \in \mathsf{P}^n \cong \mathbb{F}_2^{2n}$ as the weight of the corresponding Pauli operator, i.e. $\mathrm{wt}_{\mathsf{P}}(e) = |\{i \in [n] \ : \ e_i \neq 0 \vee e_{i+n} \neq 0\}|$.

We describe errors using Pauli channels, i.e. quantum channels of the form

$$\rho \mapsto \sum_{e \in \mathsf{P}^n} P(e)\, e\rho e^\dagger , \tag{29}$$

where $P : \mathsf{P}^n \to [0,1]$ is a normalized probability distribution, the *error distribution* of the channel; note again that this expression is independent of the choices of representatives of $e$ in $\mathcal{P}^n$.

A stabilizer code is defined by a set of commuting Pauli operators $g^{(1)} \ldots, g^{(l)} \in \mathcal{P}^n$. They generate an abelian subgroup $\mathscr{S} \subseteq \mathcal{P}^n$, called stabilizer group, which must fulfill $-1 \notin \mathscr{S}$. This is the analogue of the classical dual code $C^\perp$. The codespace is defined as the simultaneous $+1$ eigenspace of the operators in $\mathscr{S}$. Standard error correction with a stabilizer code proceeds as follows. In each round all generators are measured. If an error $e \in \mathsf{P}^n$ occurred then the vector of all measurement outcomes can be represented by the syndrome $\mathcal{O}(e)$, defined as

$$\mathcal{O}(e)_i = \langle g^{(i)}, e \rangle_{\mathsf{P}} \quad \forall i = 1, \ldots, l . \tag{30}$$

Using the measured syndrome, and based on information about the error rates, one approximates the most likely logical error and applies it as a correction. The outcomes of all stabilizers are determined only by the measurements of the generators (in the case of perfect measurements) via linearity of $\langle \cdot, e \rangle_{\mathsf{P}}$. As in the classical case, we can define the distance of a stabilizer code. However, in contrast to the classical case, there are many errors that act trivially on the code space and do not affect the logical information. We define the *distance* as the smallest weight of an undetectable error that affects the state of the logical qubit, i.e.

$$d \coloneqq \min\{\mathrm{wt}_{\mathsf{P}}(e) \ : \ e \in \mathsf{P}^n \setminus \mathscr{S},\, \mathcal{O}(e) = 0 \} . \tag{31}$$

Moreover, we define the *pure distance* of a code as the smallest weight of any undetectable error,

$$d_p \coloneqq \min\{\mathrm{wt}_{\mathsf{P}}(e) \ : \ e \in \mathsf{P}^n \setminus \{I\},\, \mathcal{O}(e) = 0 \} . \tag{32}$$

The pure distance and the distance can differ significantly. For example, the distance of the toric code is equal to the lattice size. The pure distance on the other hand is 4 independent of the lattice size, since the weight of any star or plaquette stabilizer is 4.

In a practice, the stabilizer measurements themselves could also be faulty. In this case, one should measure additional elements of $\mathscr{S}$ to mitigate the effect of measurement errors. This is

captured by the framework of *quantum data-syndrome codes*, which allow for a unified treatment of data and measurement errors [26–28]. We now give the basic definitions, following [26]. In the context of data-syndrome codes, errors are described by a data and a measurement part, as $e = (e_d, e_m) \in \mathsf{P}^n \times \mathbb{F}_2^m \cong \mathbb{F}_2^{2n+m}$. The swapping of $X$- and $Z$-bits now only applies to the data bits, i.e. $\overline{e} = (\overline{e_d}, e_m)$ for $e \in \mathbb{F}_2^{2n+m}$. We extend the symplectic product to data-syndrome codes by

$$\langle (s_d, s_m), (e_d, e_m) \rangle_{\mathrm{DS}} = \langle s_d, e_d \rangle_{\mathsf{P}} + s_m \cdot e_m s_d \cdot \overline{e}_d + s_m \cdot e_m \,. \tag{33}$$

A quantum data-syndrome code is defined by a stabilizer code on $n$ physical qubits with generators $g^{(1)}, \ldots, g^{(l)} \in \mathsf{P}^n \cong \mathbb{F}_2^{2n}$ and a classical code that encodes $l$ bits into $m$ bits. We can always write the generator matrix of the classical code in the systematic form $G_C = \begin{bmatrix} I_l & A \end{bmatrix}$ with $A \in \mathbb{F}_2^{l \times (m-l)}$. Instead of just the generators $g^{(1)}, \ldots, g^{(l)}$, we measure the stabilizers $f^{(1)}, \ldots, f^{(m)}$ defined by

$$f^{(i)} = \sum_{j=1}^{l} G_C[j,i] g^{(j)} \,. \tag{34}$$

Note that $f^{(i)} = g^{(i)}$ for $i \leq l$. We collect the stabilizers $f^{(i)}$ into a matrix $F = \begin{bmatrix} f^{(1)\,T} \\ \vdots \\ f^{(m)\,T} \end{bmatrix}$. The measurements of the stabilizers can then be described by the parity check matrix

$$H_{\mathrm{DS}} = \begin{bmatrix} F & I_m \end{bmatrix} \,, \tag{35}$$

where the identity part describes the effect of measurement errors, as seen by the following discussion. If an error $e = (e_d, e_m) \in \mathsf{P}^n \times \mathbb{F}_2^m$ occurred, the syndrome $\mathcal{O}(e)$ is then described by

$$\mathcal{O}(e)_i = \langle h^{(i)}, e \rangle_{\mathrm{DS}} \,, \tag{36}$$

where $h^{(i)}$ is the $i$-th row of $H_{\mathrm{DS}}$. In phase space representation this simply means $\mathcal{O}(e) = H\overline{e} = F\overline{e}_d + e_m$, i.e. the syndrome is the sum of the ideal syndrome and the measurement errors. We denote the row span of $H_{\mathrm{DS}}$ as $\mathscr{S}_{\mathrm{DS}}$, since it is an analogue of the stabilizer group.

We define the Pauli weight of $e = (e_d, e_m) \in \mathsf{P}^n \times \mathbb{F}_2^m \cong \mathbb{F}_2^{2n+m}$ as $\mathrm{wt}_{\mathsf{P}}(e) = \mathrm{wt}_{\mathsf{P}}(e_d) + \mathrm{wt}(e_m)$. Analogous to stabilizer codes, we can define the *distance* of the data-syndrome code as

$$d \coloneqq \min\{\mathrm{wt}_{\mathsf{P}}(e) : e \in \mathsf{P}^n \times \mathbb{F}_2^m \setminus \{(h,0) : h \in \mathscr{S}\},\, \mathcal{O}(e) = 0\} \,,$$

and the *pure distance* as

$$d_p \coloneqq \min\{\mathrm{wt}_{\mathsf{P}}(e) : e \in \mathsf{P}^n \times \mathbb{F}_2^m \setminus \{0\},\, \mathcal{O}(e) = 0\} \,.$$

Naturally, the distance of a data-syndrome code cannot be larger than that of the underlying quantum code. Furthermore, it is not hard to see from the definitions that the pure distance of a data-syndrome code is the minimum of its distance and the pure distance of the underlying stabilizer code. Thus, the pure distance is limited primarily by the underlying quantum code.

All in all, measurement errors and data errors can be treated in a unified way, analogous to a standard stabilizer code.

### 2.3.2 Identifiability results for quantum codes

Now we extend the identifiability results from classical to quantum data-syndrome codes, by applying analogous arguments to the phase space representation. The explicit example of the toric code is discussed in Section 2.1.

We consider a quantum data-syndrome code on $n$ qubits and $m$ measurement bits, and set $N \coloneqq 2n + m$. Similar to the classical case, we consider an error model described by independent channels on some selections of qubits and measurement bits. We also allow that some of the channels contain only $X$- or $Z$-errors on some qubits. Thus, in the phase space representation, we consider a set of *channel supports* $\Gamma \subseteq 2^{[N]}$, and for each $\gamma \in \Gamma$ an error distribution $P_\gamma : 2^{[N]} \mapsto [0,1]$ that is only supported on $2^\gamma$, where we again identify binary vectors in $\mathbb{F}_2^N$ with subsets of

$[N]$. Since the total error is a sum of independently occurring errors, the total error distribution is again given by a Boolean convolution,

$$P = \circledast_{\gamma\in\Gamma} P_\gamma \,. \tag{37}$$

We denote the sets of bits in phase space representation that only support *detectable* errors as

$$\Gamma^{(D)} = \{\gamma \subseteq [N] \ : \ \mathcal{O}(e) \neq 0 \ \forall e \subseteq \gamma\} \tag{38}$$

and denote $\overline{\Gamma^{(D)}} = \{\overline{e} \ : \ e \in \Gamma^{(D)}\}$. Using these definitions, we can state our main result.

**Theorem 7** (General identifiability condition)**.** *Consider a quantum data-syndrome code with $n$ qubits and $m$ measurement bits subject to noise described by the channel supports $\Gamma \subseteq 2^{[N]}$, where $N \coloneqq 2n+m$. Assume that any union of two channel supports only supports detectable errors, i.e. for all $\gamma_1, \gamma_2 \in \Gamma$, $\gamma_1 \cup \gamma_2 \in \Gamma^{(D)}$. Furthermore assume $P_\gamma(0) > 0.5$ for all $\gamma \in \Gamma$. Then the total error distribution $P$ is identifiable from the syndrome statistics.*

As in the classical case, we can also consider the set of "independently occurring errors",

$$\hat{\Gamma} = \{\emptyset \neq e \subseteq [N] \ : \ \exists \gamma \in \Gamma \text{ such that } e \subseteq \gamma \ \}\,, \tag{39}$$

instead of the set of channel supports $\Gamma$. As explained Appendix B, an equivalent condition to $\gamma_1 \cup \gamma_2 \in \Gamma^{(D)}$ for all $\gamma_1, \gamma_2 \in \Gamma^{(D)}$ in terms of these independent errors is that $\mathcal{O}(e_1) \neq 0$ and $\mathcal{O}(e_1 + e_2) \neq 0$ for all $e_1, e_2 \in \hat{\Gamma}$. Thus, we require that independently occurring errors have different syndromes. As also discussed in the classical case, this does not preclude that different errors with the same syndrome occur frequently, but they must arise as a combination of independent errors.

The most important implication of Theorem 7 is the following.

**Corollary 8** (Pure distance and identifiability)**.** *Consider a quantum data-syndrome code of pure distance $d_p$ on $n$ qubits and $m$ measurement bits, subject to noise described by the channel supports $\Gamma = \{\gamma \subseteq [2n+m] \ : \ \mathrm{wt}_{\mathsf{P}}(\gamma) \leq t\}$. Furthermore assume $P_\gamma(0) > 0.5$ for all $\gamma \in \Gamma$. Then $P$ is identifiable from the syndrome statistics if $t \leq \lfloor \frac{d_p - 1}{2} \rfloor$.*

*Proof.* If the quantum data-syndrome code has pure distance $d_p$ then $\Gamma^{(D)}$ contains any $\gamma \subseteq [N]$ with Pauli weight at most $t$ (when viewed as a binary vector). □

In other words, with a code of pure distance $d_p$ we can estimate Pauli noise that is correlated across $\lfloor \frac{d_p - 1}{2} \rfloor$ combined qubits and measurement bits. This proves the informal Theorem 1. One could also make stronger independence assumptions, for example that data and measurement errors occur independently of each other. In this case one can consider the pure distance $d_Q$ of the underlying quantum code and the distance $d_C$ of the measurement code independently and can estimate correlations across at least $d_Q$ data qubits and $d_C$ measurement bits. Similarly, for CSS-codes, one can consider a separate $X$- and $Z$-distance if one assumes that $X$- and $Z$-errors occur independently.

Let us now discuss the proof of Theorem 7, which consists of carefully applying the arguments from the classical case to the phase space representation. In this framework, the main difference to the classical case is that the moments must be defined using the symplectic product instead of the normal dot product if we want them to match the measured expectation values. Thus, for any subset $a \subseteq [N]$, we define

$$E(a) \coloneqq \sum_{e\in\mathbb{F}_2^N} (-1)^{\langle a,e\rangle_{\mathrm{DS}}} P(e), \tag{40}$$

$$E_\gamma(a) \coloneqq \sum_{e\in\mathbb{F}_2^N} (-1)^{\langle a,e\rangle_{\mathrm{DS}}} P_\gamma(e)\,. \tag{41}$$

For $s \in \mathscr{S}_{\mathrm{DS}}$, $E(s)$ is again exactly the expectation value of the measurement of the stabilizer in repeated rounds of error correction. However, the relation between moments and Fourier coefficients now contains a "twist", i.e.

$$E(a) = \sum_{e\in\mathbb{F}_2^N} (-1)^{\langle a,e\rangle_{\mathrm{DS}}} P(e) = \sum_{e\in\mathbb{F}_2^N} (-1)^{\overline{a}\cdot e} P(e) = \widetilde{P}(\overline{a})\,.$$

A similar connection between measurements and Fourier coefficients has recently been pointed out by Flammia and O'Donnell [15].

Because $P_\gamma$ is only supported on $2^\gamma$, $\widetilde{P}_\gamma(a) = \widetilde{P}_\gamma(a \cap \gamma)$ and thus $E_\gamma(a) = E_\gamma(a \cap \overline{\gamma})$. It follows that

$$E(a) = \prod_{\gamma \in \Gamma} E_\gamma(a \cap \overline{\gamma})\,. \tag{42}$$

We can define the transformed moments $F(b)$ via the inclusion-exclusion transform (12) as in the classical case and obtain

$$E(a) = \prod_{b \subseteq a} F(b) \quad \forall a \subseteq [N]\,. \tag{43}$$

By replacing $\gamma$ with $\overline{\gamma}$ in Lemma 2, we obtain $F(a) = 1$ if there is no $\gamma \in \Gamma$ such that $a \subseteq \overline{\gamma}$. It thus suffices to know the transformed moments $F(b)$ for $b \in \overline{\hat{\Gamma}}$, where

$$\overline{\hat{\Gamma}} = \{\emptyset \neq A \subseteq [N] \;:\; \exists \gamma \in \Gamma \text{ such that } A \subseteq \overline{\gamma}\}\,.$$

The problem of learning the error rates has thus been reduced to the problem of learning the transformed moments from the measured expectation values. Explicitly, we need to solve the following equation system, which is analogous to the classical case.

$$\begin{aligned} &\text{Given } (E(s))_{s \in \mathscr{S}_{\mathrm{DS}}} \text{ find } (F(b))_{b \in \overline{\hat{\Gamma}}} \text{ satisfying} \\ &E(s) = \prod_{b \in \overline{\hat{\Gamma}} : b \subseteq s} F(b) \qquad \forall s \in \mathscr{S}_{\mathrm{DS}} \setminus \{0\}\,. \end{aligned} \tag{44}$$

This system can be described by the *coefficient matrix* $D$, whose rows are labeled by elements of $\mathscr{S}_{\mathrm{DS}}$ and whose columns are labeled by elements of $\overline{\hat{\Gamma}}$, with entries

$$D[s,a] = \begin{cases} 1 & \text{if } a \subseteq s \\ 0 & \text{otherwise} \end{cases}\,. \tag{45}$$

We are now in a position to finish the proof of Theorem 7.

*Proof of Theorem 7.* First we show that the coefficient matrix $D$ defined in (45) has full rank. Note that by the definition of $\Gamma^{(D)}$, if $\gamma \in \Gamma^{(D)}$, then also $a \in \Gamma^{(D)}$ for any $a \subseteq \gamma$. The assumption that $\gamma_1 \cup \gamma_2 \in \Gamma^{(D)}$ for all $\gamma_1, \gamma_2 \in \Gamma$ thus implies that $a \cup b \in \overline{\Gamma^{(D)}}$ for all $a, b \in \overline{\hat{\Gamma}}$. It is proven in Appendix C (Lemma 10) that the elements of $\mathscr{S}_{\mathrm{DS}}$ look locally uniformly random on any element of $\overline{\Gamma^{(D)}}$. The same arguments as in the proof of Theorem 3 thus imply that $D^T D$ can be re-scaled to be a principal sub-matrix of the matrix $M_t$ from (23) (for $t = \max\{|a| \;:\; a \in \overline{\hat{\Gamma}}\}$). Since $M_t$ is positive-definite by Lemma 6, we conclude that $D^T D$ and thus $D$ is of full rank. By [35][Proposition 2], the re-scaled moments $F(a)$, and thus also the standard moments $E(a)$, can be estimated up to their sign. If we assume all moments $E(a)$ to be positive, then the estimate is unique. A sufficient condition for all moments $E(a)$ to be positive is that $P_\gamma(0) > 0.5$ for all $\gamma \in \Gamma$. □

Similar to the classical case, the appearance of multiple solutions, differing by the signs of some moments, reflects the symmetries of the problem. This becomes especially apparent when considering the simple case of a stabilizer code with single qubit noise and no measurement errors. In this case, errors on each qubit are independent. Thus, for a stabilizer $S = S_1 \otimes \cdots \otimes S_n \in \mathcal{P}^n$, we have $E(S) = \prod_{i \in [n]} E(S_i)$, which is a simpler form of the equation system (44). Consider a representative $L \in \mathcal{P}^n$ of a logical operator (including stabilizers). Denote as

$$\mathcal{A}(L) \coloneqq \{A \in \mathcal{P}^n \;:\; \mathrm{wt}_{\mathsf{P}}(A) = 1,\, \langle A, L\rangle_{\mathsf{P}} = 1\} \tag{46}$$

the set of single qubit Pauli operators that anti-commute with $L$. Since $L$ commutes with all stabilizers $S$, each equation $E(S) = \prod_{i \in [n]} E(S_i)$ contains, on the right-hand side, an even number of terms $E(S_i)$ such that $S_i \in \mathcal{A}(L)$. Thus, flipping the sign of $E(S_i)$ for all $S_i \in \mathcal{A}(L)$ does not

change the measured syndrome statistics. We see that there is one symmetry for each representative of a logical operator.

In summary, the estimation problem is best phrased in terms of moments instead of error rates. From this perspective, estimating the error distribution boils down to solving a polynomial system (44). If the correlations in the error distribution are small enough, this system has a finite number of discrete solutions. These solutions are described by symmetries related to the logical operators of the code. Under the additional mild assumption that the error rates are smaller than $\frac{1}{2}$ the solution is unique.

## 2.4 Practical estimation

While the main result of this paper lies in establishing identifiability conditions in principle, our proofs also suggest a practical method to actually perform the estimation. Here, we briefly sketch this method and heuristically comment on its sample complexity. A detailed analysis is ongoing work.

We suggest a method of moments estimator, i.e. to use empirical expectation values $\hat{E}(s)$ for $E(s)$ in (44). The basic steps (for the quantum case, but the classical case is exactly analogous) are the following:

1. Perform $m$ syndrome measurements, and use them to compute the empirical expectation value $\hat{E}(s)$ for each stabilizer $s$.

2. Insert these empirical expectation values into (44) and solve the resulting binomial system to obtain an estimate $\left(\hat{F}(b)\right)_{b\in\bar{\bar{\Gamma}}}$ of the transformed moments.

3. Insert this estimate into (43) to obtain an estimate of the moments.

4. Perform the inverse Fourier transform (4) to obtain an estimate $\hat{P}$ of the Pauli error rates.

The relevant binomial system can be solved e.g. using the methods described by Chen and LiTien-Yien [35]. It will generally be over-determined. In principle, one can select a subset of equations such that the system is exactly determined, and then solve it analytically, resulting in a closed form expression for the estimate of the error rates in terms of the empirically measured expectation values. This is illustrated by the example of the toric code (Section 2.1), for which our method reproduces the estimator suggested by Spitz *et al.* [7]. This estimator has also been applied byVarbanov *et al.* [38]. However, since the empirical expectations contain a certain error, this solution might not always yield a proper probability distribution, if the number of samples used is low. Furthermore, for an over-determined system, selecting a subset of equations removes some of the measured information, which can reduce the accuracy of the estimate. In such cases it might be preferable to instead use a least-squares solver for the over-determined system.

Since our algorithm is designed for an on-line setting where only syndrome information is available, the sample complexity must necessarily be worse than that of algorithms using arbitrary measurements, such as [17] and [15]. Each syndrome measurement contains a measurement of each stabilizer generator, therefore the expectation values of all stabilizers can be estimated simultaneously and no large measurement overhead is needed. However, the estimation error has to be propagated through the binomial system, the inclusion-exclusion transform and the inverse Fourier transform. Heuristically, we expect a sample complexity similar to that of factor graph learning [39], which was applied to Pauli channel estimation by Flammia and Wallman [17, Result 3]. However, there will be a complicating factor accounting for the conditioning of the binomial system described by the coefficient matrix $D$. Compared to other algorithms designed for the on-line setting [2, 7, 19–24], we expect that our method compares favorably, since these algorithms are either designed for a limited class of codes and Pauli noise models, deal with likelihood functions based on syndrome probabilities, not moments, which quickly become intractable, or only work in the limit of vanishing error rates.

In practice, it will not be necessary to estimate the expectation values of all stabilizers, but only a limited subset will be sufficient, depending on how many equation of the binomial system are retained. We expect that it is sufficient to only consider a selection of neighboring stabilizers

for each qubit. For example, in the toric code example Section 2.1, only three expectation values are needed per qubit, independent of the system size.

Finally, in case that the Pauli noise model is not strictly identifiable, we expect that the estimate will combine the rates of indistinguishable errors, but still give a good estimate of these total error rates.

# 3 Discussion

In this work, we considered the estimation of Pauli channels just from the measurements done anyway during the decoding of a quantum code. We established a general condition for the feasibility of this estimation and explained the relation to the pure distance of the code. Essentially, the estimation is possible as long as the noise has no correlations which exceed the detection capabilities of the code. This result does not rely on the limit of vanishing error rates, and applies even if high weight errors occur frequently, as long as these high weight errors arise as a combination of independent lower weight errors. Our results cover general stabilizer codes and quantum data-syndrome codes [26–28] and also take into account measurement errors. The previously proposed *adaptive weight estimator* [7] can be seen as a special case of our general results, since it solves a specific instance of the system (18) for single qubit noise and a limited class of codes, those that admit a minimum weight matching decoder.

An interesting new problem is an analogue of our results on the logical level. Since we cannot identify the rates of undetectable errors, there is no full identifiability for correlations larger than the pure distance. However, we expect that the logical channel can still be identified, as long as there are no correlations larger than the actual distance of the code. This would practically allow for an analogue of Theorem 1, using the distance instead of the pure distance. Furthermore, our proof naturally suggests a method of moments estimator for the Pauli error rates by inserting the empirical moments in (18). A detailed analysis of this method, including rigorous performance guarantees as well as a better assessment of its sample complexity, is ongoing work.


## Acknowledgments

We thank Daniel Miller for discussions on algebraic geometry. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – Cluster of Excellence Matter and Light for Quantum Computing (ML4Q) EXC 2004/1 – 390534769. The work of M.K. is also supported by the DFG via the Emmy Noether program (grant number 441423094) and by the German Federal Ministry of Education and Research (BMBF) within the funding program "quantum technologies – from basic research to market" in the joint project MIQRO (grant number 13N15522).


## Competing interests

The authors declare no competing interest.

## Appendix

In this appendix, we provide some auxiliary statements in order to keep this work largely self-contained. We also provide the proof of Lemma 6 which was omitted in the main text. Finally, we explain how the toric code example in Section 2.1 relates to the results derived by Spitz *et al.* [7].

## A Proof of Lemma 2

Inserting equation (14) into equation (12) yields

$$F(a) = \prod_{c\subseteq a}(\prod_{\gamma\in\Gamma} E_\gamma(c\cap\gamma))^{\mu(a,c)} = \prod_{\gamma\in\Gamma}\prod_{c\subseteq a} E_\gamma(c\cap\gamma)^{\mu(a,c)}$$
$$= \prod_{\gamma\in\Gamma}\prod_{b\subseteq\gamma}\prod_{c\subseteq a: c\cap\gamma=b} E_\gamma(b)^{\mu(a,c)} = \prod_{\gamma\in\Gamma}\prod_{b\subseteq\gamma} E_\gamma(b)^{\sum_{c\subseteq a: c\cap\gamma=b}\mu(a,c)}. \tag{47}$$

Writing $c = b\cup r$ with $r\cap\gamma=\emptyset$, and using the fact that the sum is empty if $b\not\subseteq a$, we obtain

$$\sum_{c\subseteq a: c\cap\gamma=b}\mu(a,c) = (-1)^{|a|}\sum_{c\subseteq a: c\cap\gamma=b}(-1)^{|c|}$$
$$= (-1)^{|a|}(-1)^{|b|}\,1[b\subseteq a]\sum_{c\subseteq a\setminus\gamma}(-1)^{|c|} = \mu(a,b)\,1[b\subseteq a]\sum_{0\le a\le|a\setminus\gamma|}\binom{|a\setminus\gamma|}{a}(-1)^a$$
$$= \mu(a,b)\,1[b\subseteq a]\,1[a\setminus\gamma=\emptyset] = \mu(a,b)\,1[b\subseteq a]\,1[a\subseteq\gamma]\,.$$

Substituting this identity into (47) yields the desired expression.

## B Equivalent condition on channel supports

Below Theorems 3 and 7, we stated two equivalent characterizations of detectable errors in terms of channel supports. The following lemma formalizes this statement. The proof is the same in both the classical and the quantum case.

**Lemma 9.** *Let $\Gamma^{(D)}$ be as defined in equation (20), $\Gamma\subseteq 2^{[n]}$ and let $\hat{\Gamma} = \{e\subseteq\gamma : \gamma\in\Gamma\}$. Then $\gamma_1\cup\gamma_2\in\Gamma^{(D)}$ for all $\gamma_1,\gamma_2\in\Gamma$ if and only if $\mathcal{O}(e_1)\neq 0$ and $\mathcal{O}(e_1+e_2)\neq 0$ for all $e_1,e_2\in\hat{\Gamma}$.*

*Proof.* Note that by the definition of $\Gamma^{(D)}$, if $\gamma\in\Gamma^{(D)}$ then $e\in\Gamma^{(D)}$ for all $e\subseteq\gamma$. Thus, $\gamma_1\cup\gamma_2\in\Gamma^{(D)}$ implies that $e_1,e_2\in\Gamma^{(D)}$ for all $e_1\subseteq\gamma_1$ and $e_2\subseteq\gamma_2$ and, in particular, $\mathcal{O}(e_1)\neq 0$. Furthermore, for any such $e_1,e_2$, we have that $e_1+e_2\subseteq\gamma_1\cup\gamma_2$, since the addition as binary vectors corresponds to the symmetric difference of the indicator sets. This implies, in particular, $\mathcal{O}(e_1+e_2)\neq 0$.

The other direction of the equivalence is proven similarly by noting that any subset of $e\subseteq\gamma_1\cup\gamma_2$ can be written as $e_1+e_2$ for appropriate choices of $e_1\subseteq\gamma_1$ and $e_2\subseteq\gamma_2$, where it could be that either $e_1$ or $e_2$ is the empty set. □

## C Orthogonal array properties

In this section, we prove local randomness properties of stabilizer codes. Consider a quantum data-syndrome code with stabilizers $\mathscr{S}_{\mathrm{DS}}$ on $n$ qubits and $m$ measurement bits. Set $N := 2n+m$. By $T$ we denote the $|\mathscr{S}_{\mathrm{DS}}|\times N$-matrix formed by all elements in the span of the rows of the parity check matrix $H_{\mathrm{DS}}$, i.e. the rows of $T$ are exactly the elements of $\mathscr{S}_{\mathrm{DS}}$. Then, the following local randomness property holds.

**Lemma 10.** *Let $\gamma\in\Gamma^{(D)}$. In the restriction $T^{|\overline{\gamma}}$ of $T$ to columns in $\gamma$, every bit-string appears equally often as a row.*

We adapt the proof of [36, Theorem 3.29] to our situation.

*Proof.* We denote the restriction of a matrix $M$ to the columns indexed by $\gamma\subseteq[n]$ by $M^{|\gamma}$. By the definition (36) of a syndrome and the definition ( equation (38)) of $\Gamma^{(D)}$ it follows that for any row $h$ of $H_{\mathrm{DS}}$ and any $e\subseteq\gamma$, $\langle h_i,e\rangle_{\mathrm{DS}}=0$ if and only if $e=0$ (as a binary vector). Since $\langle h_i,e\rangle_{\mathrm{DS}} = \overline{h}_i\cdot e$, we conclude that the columns of $H_{\mathrm{DS}}^{|\overline{\gamma}}$ are linearly independent (over $\mathbb{F}_2$), i.e. $H_{\mathrm{DS}}^{|\overline{\gamma}}$ is of rank $|\overline{\gamma}|$. The rows of $T^{|\overline{\gamma}}$ are, by definition of $T$, the vectors $\zeta^T H_{\mathrm{DS}}^{|\overline{\gamma}}$ for $\zeta\in\mathbb{F}_2^l$, where $l$ is the number of rows of $H$. The number of times a bit-string $z$ appears as a row in $T$ is thus equal to the number of vectors $\zeta\in\mathbb{F}_2^l$ with $\zeta^T H^{|\overline{\gamma}} = z$. Since $H^{|\overline{\gamma}}$ has rank $|\overline{\gamma}|$, this number is $2^{l-|\overline{\gamma}|}$ independent of $z$. □

Lemma 5 for classical codes is a direct corollary, which follows by only considering the classical (measurement) part of the data-syndrome code.

# D Analysis of the intersection matrix

In this appendix, we first summarize the definition and the most important properties of the Schur complement (see e.g. the book by Horn and Johnson [40]). On this basis we then prove Lemma 6 on the intersection matrix.

## D.1 Iterated Schur complements

For a block matrix

$$M_2 = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}, \tag{48}$$

with $M_{11}$ being invertible, the *Schur complement* of $M_{11}$ in $M_2$ is defined as

$$M_2/M_{11} = M_{22} - M_{21}M_{11}^{-1}M_{12}\,. \tag{49}$$

If $M_{11}$ is invertible, then $M_2$ has full rank if and only if $M_2/M_{11}$ has full rank. Furthermore, then $M_2$ is positive definite if and only if $M_{11}$ and $M_2/M_{11}$ are positive definite. For the extended block matrix

$$M_3 = \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{pmatrix} \tag{50}$$

the upper left block of $M_3/M_{11}$ is given by $M_2/M_{11}$ and the following *quotient property of Schur complements* holds:

$$M_3/M_2 = (M_3/M_{11})/(M_2/M_{11})\,. \tag{51}$$

## D.2 The intersection matrix and proof of Lemma 6

In the Section 2.2.6, we have shown that the coefficient matrix $D$ has full rank, which implies identifiability of the error distribution. The only thing missing is the proof of Lemma 6, which was omitted. We will now supply this proof, i.e. we show that the intersection matrix $M_t$ defined in (23) is positive-definite for any $t \in \mathbb{N}$.

Ordering the indexing sets $a, b \in \Gamma_{\leq t}$ by their size, for $r \leq t$, we can view $M_r$ as a sub-matrix in the upper left corner of $M_t$. $M_t$ is an $(\alpha_t - 1) \times (\alpha_t - 1)$ matrix, where we define $\alpha_t = \sum_{k=0}^{t} \binom{n}{k}$. Note that in the case $t = 1$, $M$ takes a very simple form. Since two single-element sets are either equal or disjoint, $M_1$ has entries 2 on the diagonal and 1 elsewhere, i.e.

$$M_1 = I + \mathbf{1}_1\mathbf{1}_1^T\,, \tag{52}$$

where $\mathbf{1}_t$ denotes the all ones vector of dimension $\binom{n}{t}$. This matrix is clearly positive-definite. In general, $M_t$ does not admit such a simple expression. However, we will show that certain *Schur complements* of $M_t$ can always be expressed in this simple form (see Appendix D.1 for the required background on Schur complements).

**Lemma 11.** *Let $M_t$ be as in (23). Then the Schur complement $M_t/M_{t-1}$ of $M_{t-1}$ in $M_t$ is well-defined and given by*

$$M_t/M_{t-1} = I + \frac{\mathbf{1}_t\mathbf{1}_t^T}{\alpha_{t-1}}\,, \tag{53}$$

*where $\mathbf{1}_t$ is the all ones vector of dimension $\binom{n}{t}$ and $\alpha_t = \sum_{k=0}^{t} \binom{n}{k}$. For $t = 1$ we set $M_1/M_0 := M_1$.*

Lemma 6 is a simple corollary of this statement:

*Proof of Lemma 6.* Clearly, $M_t/M_{t-1}$ is positive-definite. Since a matrix is positive-definite if the upper left block and its Schur complement are both positive-definite, it follows that $M_t$ is positive-definite. □

For the proof of Lemma 11, some further preliminaries and notation are needed. We can view $M_t$ as a block matrix consisting of blocks $(M_{w,w'} : w, w' \in \{1, \ldots, t\})$, where $M_{w,w'}$ labels the block whose rows are indexed by sets of size $w$ and whose columns are indexed by sets of size $w'$. By repeatedly applying the quotient property of Schur complements ((51) in the Appendix), we can express $M_t/M_{t-1}$ as a chain of Schur complements:

$$\begin{aligned} M_t/M_{t-1} &= (M_t/M_{t-2})/(M_{t-1}/M_{t-2}) \\ &= [(M_t/M_{t-3})/(M_{t-2}/M_{t-3})]/(M_{t-1}/M_{t-2}) \\ &= [[[(M_t/M_1)/(M_2/M_1)]/(M_3/M_1)]\ldots]/(M_{t-1}/M_{t-2})\,. \end{aligned}$$

This chain is described by the recursive definitions

$$M_t^{(0)} = M_t\,, \tag{54}$$

$$M_t^{(1)} = (M_t/M_1)\,, \tag{55}$$

$$M_t^{(i)} = M_t^{(i-1)}/(M_i/M_{i-1})\,. \tag{56}$$

We use the same labeling of blocks for $M_t^{(i)}$ as for $M_t$, i.e. $M_{w,w'}^{(i)}$ is the block of $M_t^{(i)}$ whose rows are labeled by sets of weight $w$ and whose columns are labeled by sets of weight $w'$, i.e.

$$M_{w,w'}^{(i)} = [M_t^{(i-1)}]_{w,w'}\,. \tag{57}$$

Of course, this is only well-defined if such a block actually exists in $M_t^{(i)}$, i.e. if $w, w' > i$. In this case neither the entries nor the dimensions of $M_{w,w'}^{(i)}$ depend on $t$, so the index t is indeed not needed.

The following useful observations follow directly from the properties of the Schur complement:

**Lemma 12.** *The matrix $M_t$ defined in (23) and the matrices $M_t^{(i)}$ defined in (56) have the following properties:*

1. $M_t^{(i)} = M_t/M_i$
2. $M_i/M_{i-1} = M_{i,i}^{(i-1)}$
3. $M_{w,i}^{(i-1)} = (M_{i,w}^{(i-1)})^T$
4. *For any $w, w' > i$:*
   $M_{w,w'}^{(i)} = M_{w,w'}^{(i-1)} - M_{w,i}^{(i-1)}(M_i/M_{i-1})^{-1}M_{i,w'}^{(i-1)}$

*Proof of Lemma 12.*

1. For $i = 0, 1$ the statement holds by definition (interpreting $M_t^{(i)}/M_0 = M_t^{(i)}$). The statement then follows by induction, using the quotient property (51).

2. We denote with $M_{>i-1}$ the sub-matrix of $M_t$ whose rows and columns are labeled by sets of size $> i-1$. Furthermore, $M_{>i-1,i-1}$ denotes the sub-matrix whose rows are labeled by sets of size $> i-1$ and whose columns are labeled with sets of size $i-1$. The matrix $M_{i-1,>i-1}$ is defined analogously and for any matrix we use a subscript $_{i,i}$ to denote the block whose rows and columns are labeled by sets of weight $i$. Remember the notation $M_{i,i}^{(i-1)} \coloneqq [M_t^{(i-1)}]_{i,i}$. Then,

$$\begin{aligned} M_{i,i}^{(i-1)} &= [M_t/M_{i-1}]_{i,i} \\ &= [M_{>i-1} - M_{>i-1,i-1}M_{i-1}^{-1}M_{i-1,>i-1}]_{i,i} \\ &= M_{i,i} - [M_{>i-1,i-1}M_{i-1}^{-1}M_{i-1,>i-1}]_{i,i} \\ &= M_{i,i} - M_{i,i-1}M_{i-1}^{-1}M_{i-1,i} \\ &= M_i/M_{i-1}\,. \end{aligned}$$

3. This statement holds because $M_t$ is symmetric.

4. The calculation is analogous to the one above. Using property 1 and the definition of Schur complements (49), we obtain

$$\begin{aligned} M^{(i)}_{w,w'} &= [M^{(i-1)}_{>i} - M^{(i-1)}_{>i,i}(M^{(i-1)}_{i,i})^{-1}M^{(i-1)}_{i,>i}]_{w,w'} \\ &= M^{(i)}_{w,w'} - M^{(i-1)}_{w,i}(M^{(i-1)}_{i,i})^{-1}M^{(i-1)}_{i,w'} \,. \end{aligned}$$

We substitute $M^{(i-1)}_{i,i}$ with $M_i/M_{i-1}$, since we use this form later.

□

Lemma 11 is now a direct consequence of the following more explicit statement.

**Lemma 13.** *$M^{(i)}_t$ is given entry-wise by*

$$M^{(i)}_t[a,b] = f^{(i+1)}(|a \cap b|) + u^{(i)}(|a|,|b|)\,, \tag{58}$$

*where*

$$f^{(i)}(x) = \sum_{k=i}^{x} \binom{x}{k} = 2^x - \sum_{k=0}^{i-1} \binom{x}{k} \tag{59}$$

$$u^{(i)}(w,w') = \frac{\binom{w-1}{i}\binom{w'-1}{i}}{\alpha_i} \tag{60}$$

$$\alpha_i = \sum_{k=0}^{i} \binom{n}{k} \tag{61}$$

To recover Lemma 11, notice that $f^{(i)}(x) = 1[x = i]$ if $x \leq i$ and $u^{(i-1)}(i,i) = \frac{1}{\alpha_{i-1}}$. Thus we obtain

$$\begin{aligned} (M_i/M_{i-1})[a,b] &= M^{(i-1)}_{i,i}[a,b] \\ &= 1[|a \cap b| = i] + u^{(i-1)}(i,i) \\ &= 1[a = b] + \frac{1}{\alpha_{i-1}}\,, \end{aligned} \tag{62}$$

where we used that $|a| = |b| = i$ because we are considering the block indexed by $i,i$. Choosing $i = t$ yields Lemma 11.

*Proof of Lemma 13.* We compute $M^{(i)}_t$ block-wise. Remember that the rows of $M^{(i-1)}_{w,w'}$ are indexed by subsets of $[n]$ of weight $w$ and the columns by subsets of $[n]$ of size $w'$, and that $M^{(i-1)}_t = M_t/M_{i-1}$ only contains blocks $M^{(i-1)}_{w,w'}$ with $w, w' \geq i$.

We prove the lemma by induction. The base case $i = 0$ is follows directly as

$$M^{(0)}_{w,w'}[a,b] = M_{w,w'}[a,b] = 2^{|a\cap b|} = 2^{|a\cap b|} - 1 + 1 = f^{(1)}(|a \cap b|) + u^{(0)}(w,w')\,. \tag{63}$$

For the induction step we assume that (58) holds for $i-1$ and we will prove it for $i$. Using the induction hypothesis, we apply (62) to obtain

$$M_i/M_{i-1} = M^{(i-1)}_{i,i} = I + \frac{\mathbf{1}_i\mathbf{1}_i^T}{\alpha_{i-1}}\,, \tag{64}$$

where $\mathbf{1}_i$ is again the all ones vector of length $\binom{n}{i}$. The inverse of this matrix can be computed via the Sherman-Morrison formula to be

$$(M_i/M_{i-1})^{-1} = I - \frac{\mathbf{1}_i\mathbf{1}_i^T}{\alpha_{i-1} + \mathbf{1}_i^T\mathbf{1}_i} = I - \frac{\mathbf{1}_i\mathbf{1}_i^T}{\alpha_{i-1} + \binom{n}{i}} = I - \frac{\mathbf{1}_i\mathbf{1}_i^T}{\alpha_i}\,. \tag{65}$$

Combining this expression with property 4 in Lemma 12, we obtain

$$M^{(i)}_{w,w'} = M^{(i-1)}_{w,w'} - M^{(i-1)}_{w,i} M^{(i-1)}_{i,w'} + \frac{M^{(i-1)}_{w,i} \mathbf{1}_i \mathbf{1}_i^T M^{(i-1)}_{i,w'}}{\alpha_i} \,. \tag{66}$$

Calculating this term by term, starting with the last term, we obtain

$$\begin{aligned}(\mathbf{1}_i^T M^{(i-1)}_{i,w'})[a] &= \sum_{|c|=i} M^{(i-1)}_{i,w'}[c,a] = \sum_{|c|=i} (f^{(i)}(|a\cap c|) + u^{(i-1)}(i,w')) \\ &= \sum_{|c|=i} (1[c\subseteq a] + u^{(i-1)}(i,w')) = \binom{w'}{i} + \binom{n}{i} u^{(i-1)}(i,w') \,,\end{aligned}$$

where $\sum_{|c|=i}$ indicates a sum over all subsets of $[n]$ of size $i$. The third equality used that

$$f^{(i)}(|a\cap c|) = 1[|a\cap c| = i] = 1[c\subseteq a] \tag{67}$$

since $|c| = i$ and the last equality used $|a| = w'$. Defining

$$S^{(i)}(w') = \binom{w'}{i} + \binom{n}{i} u^{(i-1)}(i,w') \tag{68}$$

and using the symmetry of $M^{(i-1)}_{i,w}$ we obtain

$$M^{(i-1)}_{w,i} \mathbf{1}_i \mathbf{1}_i^T M^{(i-1)}_{i,w'} = S^{(i)}(w) S^{(i)}(w') \mathbf{1}_i \mathbf{1}_i^T \,. \tag{69}$$

Next, we consider the second term of (66),

$$\begin{aligned}(M^{(i-1)}_{w,i} M^{(i-1)}_{i,w'})[a,b] &= \sum_{|c|=i} M^{(i-1)}_{w,i}[a,c] M^{(i-1)}_{i,w'}[c,b] \\ &= \sum_{|c|=i} [f^{(i)}(|a\cap c|) + u^{(i-1)}(i,w)][f^{(i)}(|c\cap b|) + u^{(i-1)}(i,w')] \\ &= \sum_{|c|=i} 1[c\subseteq a] 1[c\subseteq b] + u^{(i-1)}(i,w) \sum_{|c|=i} 1[c\subseteq b] \\ &\quad + u^{(i-1)}(i,w') \sum_{|c|=i} 1[c\subseteq a] + \sum_{|c|=i} u^{(i-1)}(i,w) u^{(i-1)}(i,w') \\ &= \binom{|a\cap b|}{i} + u^{(i-1)}(i,w)\binom{w'}{i} + u^{(i-1)}(i,w')\binom{w}{i} + \binom{n}{i} u^{(i-1)}(i,w) u^{(i-1)}(i,w') \\ &= \binom{|a\cap b|}{i} + T^{(i)}(w,w') \,,\end{aligned} \tag{70}$$

where we defined

$$T^{(i)}(w,w') = u^{(i-1)}(i,w)\binom{w'}{i} + u^{(i-1)}(i,w')\binom{w}{i} + \binom{n}{i} u^{(i-1)}(i,w) u^{(i-1)}(i,w') \,. \tag{71}$$

Thus, combining (69), (70), and the induction assumption (58) with (66) we obtain

$$\begin{aligned}M^{(i)}_{w,w'}[a,b] &= f^{(i)}(|a\cap b|) + u^{(i-1)}(w,w') - \binom{|a\cap b|}{i} - T^{(i)}(w,w') + \frac{S^{(i)}(w) S^{(i)}(w')}{\alpha_i} \\ &= f^{(i+1)}(|a\cap b|) + \frac{S^{(i)}(w) S^{(i)}(w') + \alpha_i (u^{(i-1)}(w,w') - T^{(i)}(w,w'))}{\alpha_i} \,,\end{aligned} \tag{72}$$

where we used $f^{(i+1)}(|a\cap b|) = f^{(i)}(|a\cap b|) - \binom{|a\cap b|}{i}$. Finally, we calculate the last term. By direct calculation from the definitions (71), (68) we obtain

$$S^{(i)}(w) S^{(i)}(w') - \binom{n}{i} T^{(i)}(w,w') = \binom{w}{i}\binom{w'}{i} \,, \tag{73}$$

and thus,

$$S^{(i)}(w)S^{(i)}(w') + \alpha_i(u^{(i-1)}(w,w') - T^{(i)}(w,w')) \\ = S^{(i)}(w)S^{(i)}(w') - \binom{n}{i}T^{(i)}(w,w') + \alpha_{i-1}u^{(i-1)}(w,w') + \binom{n}{i}u^{(i-1)}(w,w') - \alpha_{i-1}T^{(i)}(w,w'), \tag{74}$$

where we have used that $\alpha_i = \binom{n}{i} + \alpha_{i-1}$. Inserting the definitions of $u^{(i)}$ and $T^{(i)}$ from (60) and (71), we obtain $\alpha_{i-1}u^{(i)}(w,w') = \binom{w-1}{i-1}\binom{w'-1}{i-1}$ and $\binom{n}{i}u^{(i-1)}(w,w') - \alpha_{i-1}T^{(i)}(w,w') = -\binom{w-1}{i-1}\binom{w'}{i} - \binom{w'-1}{i-1}\binom{w}{i}$. Together with (73) this yields

$$\begin{aligned}(74) &= \binom{w}{i}\binom{w'}{i} + \binom{w-1}{i-1}\binom{w'-1}{i-1} - \binom{w-1}{i-1}\binom{w'}{i} - \binom{w'-1}{i-1}\binom{w}{i} \\ &= \left[\binom{w}{i} - \binom{w-1}{i-1}\right]\left[\binom{w'}{i} - \binom{w'-1}{i-1}\right] \\ &= \binom{w-1}{i}\binom{w'-1}{i},\end{aligned}$$

where the last step is based on Pascals identity. Substituting this result back into (72) we obtain

$$M^{(i)}_{w,w'}[a,b] = f^{(i+1)}(|a\cap b|) + \frac{\binom{w-1}{i}\binom{w'-1}{i}}{\alpha_i}, \tag{75}$$

which finishes the proof. □

# E Connection to Adaptive Weight Estimator

Here, we show that the solution for the toric code derived in Section 2.1 coincides with the solution given by Spitz *et al.* [7].

The solution derived in Section 2.1 is

$$E(Z_4) = \pm\sqrt{\frac{E(S_1)E(S_2)}{E(S_1S_2)}}. \tag{76}$$

On the other hand, the solution given by Spitz *et al.* [7, eq. (14)] is

$$p_4 = \frac{1}{2} \mp \sqrt{\frac{1}{4} - \frac{P(S_1 = S_2 = 1) - P(S_1 = 1)P(S_2 = 1)}{1 - 2P(S_1 \neq S_2)}}. \tag{77}$$

Here, $P$ is used to denote the probabilities of events under random sampling of the errors. Note that in [7] the result is phrased in terms of expectation values of the stabilizer outcomes and errors viewed as taking values 0 or 1, which directly translates into the probabilities above.

To see that these two solutions coincide, first notice that $E(Z_4) = 1 - 2p_4$. Thus, (76) can be rewritten as

$$p_4 = \frac{1}{2} \mp \sqrt{\frac{E(S_1)E(S_2)}{4E(S_1S_2)}}. \tag{78}$$

Similarly, we have $E(S_i) = 1-2P(S_i = 1)$ and $E(S_1S_2) = 1-2P(S_1 \neq S_2)$. Using these equations, equality of the solutions can be shown as follows:

$$\begin{aligned}\frac{E(S_1)E(S_2)}{4E(S_1S_2)} &= \frac{(1-2P(S_1=1))(1-2P(S_2=1))}{4(1-2P(S_1\neq S_2))} \\ &= \frac{1-2P(S_1=1)-2P(S_2=1)+4P(S_1=1)P(S_2=1)+1-2P(S_1\neq S_2)-1+2P(S_1\neq S_2)}{4(1-2P(S_1\neq S_2))} \\ &= \frac{1}{4} - \frac{2P(S_1=1)+2P(S_2=1)-4P(S_1=1)P(S_2=1)-2P(S_1\neq S_2)}{4(1-2P(S_1\neq S_2))} \\ &= \frac{1}{4} - \frac{P(S_1=S_2=1)-P(S_1=1)P(S_2=1)}{1-2P(S_1\neq S_2)},\end{aligned}$$

where in the last equality we used that

$$\begin{aligned}
&P(S_1 = 1) + P(S_2 = 1) - P(S_1 \neq S_2) \\
&= P(S_1 = 1, S_2 = 1) + P(S_1 = 1, S_2 = 0) \\
&\quad + P(S_1 = 1, S_2 = 1) + P(S_1 = 0, S_2 = 1) \\
&\quad - P(S_1 = 1, S_2 = 0) - P(S_1 = 0, S_2 = 1) \\
&= 2P(S_1 = S_2 = 1)\,.
\end{aligned}$$

This shows that the two solutions (76) and (77) are indeed equivalent.

## References


[1] A. Robertson, C. Granade, S. D. Bartlett, and S. T. Flammia, *Tailored codes for small quantum memories*, Phys. Rev. Applied **8**, 064004 (2017).

[2] J. Florjanczyk and T. A. Brun, *In-situ adaptive encoding for asymmetric quantum error correcting codes* (2016).

[3] J. P. Bonilla Ataides, D. K. Tuckett, S. D. Bartlett, S. T. Flammia, and B. J. Brown, *The XZZX surface code*, Nat. Commun. **12**, 2172 (2021).

[4] O. Higgott, *Pymatching: A python package for decoding quantum codes with minimum-weight perfect matching* (2021).

[5] E. Dennis, A. Kitaev, A. Landahl, and J. Preskill, *Topological quantum memory*, J. Math. Phys. **43**, 4452 (2002), arXiv:quant-ph/0110143 [quant-ph].

[6] N. H. Nickerson and B. J. Brown, *Analysing correlated noise on the surface code using adaptive decoding algorithms*, Quantum **3**, 131 (2019).

[7] S. T. Spitz, B. Tarasinski, C. W. J. Beenakker, and T. E. O'Brien, *Adaptive weight estimator for quantum error correction in a time-dependent environment*, Advanced Quantum Technologies **1**, 1870015 (2018).

[8] Z. Babar, P. Botsinis, D. Alanis, S. X. Ng, and L. Hanzo, *Fifteen years of quantum LDPC coding and improved decoding strategies*, IEEE Access **3**, 2492 (2015).

[9] S. Huang, M. Newman, and K. R. Brown, *Fault-tolerant weighted union-find decoding on the toric code*, Physical Review A **102**, 10.1103/physreva.102.012419 (2020).

[10] C. T. Chubb, *General tensor network decoding of 2d pauli codes* (2021).

[11] A. S. Darmawan and D. Poulin, *Linear-time general decoding algorithm for the surface code*, Physical Review E **97**, 10.1103/physreve.97.051302 (2018).

[12] J. J. Wallman and J. Emerson, *Noise tailoring for scalable quantum computation via randomized compiling*, Phys. Rev. A **94**, 052325 (2016).

[13] M. Ware, G. Ribeill, D. Ristè, C. A. Ryan, B. Johnson, and M. P. da Silva, *Experimental Pauli-frame randomization on a superconducting qubit*, Phys. Rev. A **103**, 042604 (2021).

[14] S. J. Beale, J. J. Wallman, M. Gutiérrez, K. R. Brown, and R. Laflamme, *Quantum error correction decoheres noise*, Phys. Rev. Lett. **121**, 190501 (2018).

[15] S. T. Flammia and R. O'Donnell, *Pauli error estimation via population recovery*, Quantum **5**, 549 (2021).

[16] R. Harper, W. Yu, and S. T. Flammia, *Fast estimation of sparse quantum noise*, PRX Quantum **2**, 010322 (2021).

[17] S. T. Flammia and J. J. Wallman, *Efficient estimation of Pauli channels*, ACM Transactions on Quantum Computing **1**, 10.1145/3408039 (2020).

[18] R. Harper, S. T. Flammia, and J. J. Wallman, *Efficient learning of quantum noise*, Nat. Phys. **16**, 1184 (2020).

[19] Y. Fujiwara, *Instantaneous quantum channel estimation during quantum information processing* (2014).

[20] A. G. Fowler, D. Sank, J. Kelly, R. Barends, and J. M. Martinis, *Scalable extraction of error models from the output of error detection circuits* (2014).

[21] M.-X. Huo and Y. Li, *Learning time-dependent noise to reduce logical errors: real time error rate estimation in quantum error correction*, New J. Phys. **19**, 123032 (2017).

[22] J. R. Wootton, *Benchmarking near-term devices with quantum error correction*, Quantum Science and Technology **5**, 044004 (2020).
[23] J. Combes, C. Ferrie, C. Cesare, M. Tiersch, G. J. Milburn, H. J. Briegel, and C. M. Caves, *In-situ characterization of quantum devices with error correction* (2014).
[24] T. Wagner, H. Kampermann, D. Bruß, and M. Kliesch, *Optimal noise estimation from syndrome statistics of quantum codes*, Phys. Rev. Research **3**, 013292 (2021).
[25] J. Kelly, R. Barends, A. G. Fowler, A. Megrant, E. Jeffrey, T. C. White, D. Sank, J. Y. Mutus, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, E. Lucero, M. Neeley, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, A. Vainsencher, J. Wenner, and J. M. Martinis, *Scalable in situ qubit calibration during repetitive error detection*, Phys. Rev. A **94**, 032321 (2016).
[26] A. Ashikhmin, C.-Y. Lai, and T. A. Brun, *Quantum data-syndrome codes*, IEEE Journal on Selected Areas in Communications **38**, 449 (2020).
[27] Y. Fujiwara, *Ability of stabilizer quantum error correction to protect itself from its own imperfection*, Phys. Rev. A **90**, 062304 (2014), arXiv:1409.2559 [quant-ph].
[28] N. Delfosse, B. W. Reichardt, and K. M. Svore, *Beyond single-shot fault-tolerant quantum error correction*, IEEE Transactions on Information Theory **68**, 287 (2022).
[29] A. Zia, J. P. Reilly, and S. Shirani, *Distributed parameter estimation with side information: A factor graph approach*, in *2007 IEEE International Symposium on Information Theory* (2007) pp. 2556–2560.
[30] R. O'Donnell, *Analysis of Boolean Functions* (Cambridge University Press, 2014).
[31] Y. Mao and F. Kschischang, *On factor graphs and the fourier transform*, IEEE Trans. Inf. Theory **51**, 1635 (2005).
[32] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning* (The MIT Press, 2009).
[33] M. Aigner, *A Course in Enumeration*, Vol. 238 (Springer-Verlag Berlin Heidelberg, 2007).
[34] S. Roman, *Field Theory* (Springer, New York, 2006).
[35] T. Chen and LiTien-Yien, *Solutions to systems of binomial equations*, Annales Mathematicae Silesianae **28**, 7 (2014).
[36] A. S. Hedayat, N. J. A. Sloane, and J. Stufken, *Orthogonal arrays: theory and applications* (Springer New York, NY, 1999).
[37] P. Delsarte, *Four fundamental parameters of a code and their combinatorial significance*, Information and Control **23**, 407 (1973).
[38] B. M. Varbanov, F. Battistel, B. M. Tarasinski, V. P. Ostroukh, T. E. O'Brien, L. DiCarlo, and B. M. Terhal, *Leakage detection for a transmon-based surface code*, NPJ Quantum Inf. **6**, 10.1038/s41534-020-00330-w (2020).
[39] P. Abbeel, D. Koller, and A. Y. Ng, *Learning factor graphs in polynomial time & sample complexity* (2012).
[40] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. (Cambridge University Press, 2012).

# D

# Learning Logical Quantum Noise in Quantum Error Correction

| | |
|---|---|
| Title: | Learning logical quantum noise in quantum error correction |
| Authors: | Thomas Wagner, Hermann Kampermann, Dagmar Bruß and Martin Kliesch |
| Journal: | Physical Review Letters |
| Publication status: | Submitted |
| Contribution by TW: | First author (input approx. 90%) |

This publication corresponds to reference [Wag+22b]. A summary of its contents is presented in chapter 6.

The main result of this work was already conjectured by me in [Wag+22a], but initially not pursued further. The project was initiated by a discussion with MK about various approaches to the question of logical identifiability. The main results and their proofs were then developed by me, and subsequently discussed with all co-authors. The proofs were carefully checked and corrected by MK, and discussed with HK and DB. I wrote the initial draft of the manuscript, with significant contributions by MK. The manuscript was then proofread and improved by all my co-authors.

# Learning logical quantum noise in quantum error correction

Thomas Wagner,* Hermann Kampermann, Dagmar Bruß, and Martin Kliesch†
*Institute for Theoretical Physics, Heinrich-Heine-University Düsseldorf, Germany*

The characterization of quantum devices is crucial for their practical implementation but can be costly in experimental effort and classical post-processing. Therefore, it is desirable to measure only the information that is relevant for specific applications and develop protocols that require little additional effort. In this work, we focus on the characterization of quantum computers in the context of stabilizer quantum error correction. Our main result is that the logical error channel induced by Pauli noise can be estimated from syndrome data under minimal conditions. More precisely, we show that the estimation is possible as long as the code can correct the noise.



For any quantum device, it is desirable to characterize both its individual components as well as their interplay [1, 2]. For the characterization of single quantum gates, protocols such as quantum process tomography (e.g. Ref. [3]) or gate set tomography [4–6] can be used. To characterize the interplay of multiple components, randomized benchmarking [7, 8], as well as crosstalk detection [9] and estimation [10, 11] protocols are available. The general goals are

(i) to build trust in the correct functioning of the device,
(ii) to be able to reduce the errors on the hardware level and improve the software calibration, and
(iii) to compare different devices and platforms in a fair way.

However, such characterization protocols can be quite resource-intensive, requiring many experimental runs of the device and such protocols' output can be challenging to interpret. Therefore, it has become a pressing issue to obtain easy-to-use information, such as Pauli error rates directly [11–15], ideally using only data that is easy to obtain. The estimation of Pauli noise is also practically interesting because randomized compiling can be used to project the actual noise onto Pauli noise [16, 17].

In the context of quantum error correction (QEC), it has been suggested to reduce the experimental effort of characterization by extracting information from the syndrome data, which is usually collected during error correction anyway [18–26]. This approach has the additional advantage of benchmarking all components in the context of the targeted application and making it easier to detect crosstalk. Indeed, syndrome data has been used to calibrate decoders and observe signatures of crosstalk in experiments on the [4,1,2]-code [27] and the repetition code [28].

For general stabilizer codes, however, the theoretical foundation of such schemes is currently lacking. Since the syndrome measurements must preserve the encoded state, it is not a priori clear that they should even contain sufficient information about the noise to be useful for QEC. For example, as shown in our previous work [26], a complete Pauli channel can only be estimated from syndrome data if it is known that the Pauli errors are not correlated across too many qubits, quantified by the *pure* distance. This limit on correlations can be quite strict, as can be seen for the toric code, which has a pure distance of 4 independent of system size. Hence, this assumption is violated by natural noise processes such as error propagation in the stabilizer measurements, which can introduce data errors on all participating qubits.

In this work, we show that the estimation of error rates is possible under much more practical conditions if one focuses only on information which is actually relevant for QEC. It is not necessary to distinguish between logically equivalent errors. Thus, it suffices to estimate the logical noise channel instead of the physical one. At least for phenomenological Pauli noise models, we prove that the situation is as good as one could reasonably hope: as long as the noise affecting a stabilizer code can be corrected by it, one can also estimate the logical noise channel from the corresponding syndrome measurements.

The proof is based on our general framework [26], but extended to consider the logical instead of the physical channel. Similar to randomized benchmarking, we consider the problem in Fourier space [12]. This representation corresponds to a description of the logical channel in terms of moments instead of probabilities. Exploiting a weak assumption of limited correlations, we can further simplify the description by switching from regular moments to a set of canonical moments. Both the logical channel and the syndrome measurements can be represented by linear equations on a small set of canonical moments. By considering the ranks of these two linear systems, we then show that the syndrome measurements determine the logical channel. Computing the ranks boils down to counting a specific subset of logical operators of the code, which we solve by employing a recent generalization of the cleaning lemma [29] of QEC.

## I. STABILIZER CODES

Let us quickly recap the most important features of stabilizer codes for our purposes. A more thorough introduction can e.g. be found in the books [30, 31]. A stabilizer code is described by a commuting subgroup $\mathscr{S} \subseteq \mathcal{P}^n$

* thomas.wagner@uni-duesseldorf.de
† science@mkliesch.eu

of the $n$-qubit Pauli group, called *stabilizer group*. It must fulfill $-I \notin \mathscr{S}$. The *codespace* is then the simultaneous $+1$-eigenspace of all the stabilizers. As is usual in the context of QEC, we disregard phases and view $\mathscr{S}$ as a subgroup of the *effective Pauli group* $\mathsf{P}^n := \mathcal{P}^n/\{\pm 1, \pm i\}$. This is an Abelian group, but the relevant commutation relations of $\mathcal{P}^n$ can be encoded in the *bicharacter* $\langle \cdot, \cdot \rangle$ on $\mathsf{P}^n$, given by

$$\langle a, e \rangle := \begin{cases} +1, & a \text{ and } e \text{ commmute in } \mathcal{P}^n \\ -1, & a \text{ and } e \text{ anti-commmute in } \mathcal{P}^n \end{cases} . \quad (1)$$

By definition, all elements of $\mathscr{S}$ act trivially on the encoded states. We can also consider Pauli operators that map the code space to itself, but do not necessarily act as the identity. These form the set $\mathscr{L} \subseteq \mathsf{P}^n$ of *logical operators*. It can be shown that $\mathscr{L}$ is exactly the set of Pauli operators that commute with all stabilizers. Formally, we can express this as the *annihilator* $\mathscr{S}^\perp$ of $\mathscr{S}$ in $\mathsf{P}^n$ under the above bicharacter, i.e.

$$\mathscr{L} := \mathscr{S}^\perp := \{l \in \mathsf{P}^n \,:\, \langle s, l \rangle = +1 \, \forall s \in \mathscr{S}\} . \quad (2)$$

In particular, we have $\mathscr{S} \subseteq \mathscr{L}$ since each stabilizer is itself a logical operator that implements the logical identity. If a logical operator (other than a stabilizer) occurs as an error, this cannot be detected and the encoded state is corrupted. The *distance* $d$ of a code is defined as the minimal weight of an element of $\mathscr{L} \setminus \mathscr{S}$. This measures the error correction capabilities of the code. We call a set of qubits $R \subseteq \{1, \ldots, n\}$ *correctable* if it only supports trivial logical operators. In particular, if $|R| < d$, then $R$ is correctable. This is however generally not an equivalence, and there can be many correctable regions of size much larger than $d$. For example, any rectangular region of side length at most $d-1$ on the $d \times d$ toric code is correctable, but contains more than $d$ qubits.

We will focus on phenomenological Pauli noise models, and do not take into account the details of error propagation inside the measurement circuits. We can then consider rounds of error correction, and between two rounds a new Pauli error occurs. These Pauli errors are described by a channel $P$, which is given by a probability distribution over Pauli errors,

$$P : \mathsf{P}^n \mapsto [0, 1] . \quad (3)$$

Later we will also impose some locality assumptions on this channel.

Standard error correction using a stabilizer code proceeds as follows: In each round, a set of generators $g_1, \cdots, g_m \in \mathscr{S}$ is measured. Ideally, the state lies in the codespace and thus all measurements return $+1$. However, if an error $e \in \mathsf{P}^n$ occurred beforehand, the outcome of the measurement of $g_i$ is $\langle g_i, e \rangle = \pm 1$. The collection of measurement outcomes of all generators is called the *syndrome*. Based on the syndrome, a decoder tries to guess the error that occurred, and applies it as a correction $r$. Since errors that only differ by stabilizers are logically equivalent, the ideal decoding strategy is to return a maximum likelihood estimate of the form

$$r = \arg\max_{e \in \mathsf{P}^n} \sum_{s \in \mathscr{S}} P(es) . \quad (4)$$

Thus, full knowledge of the physical channel $P$ is not necessary for optimal decoding. Instead, it is sufficient to know the *logical channel* $P_L$, which we define by averaging $P$ over cosets of $\mathscr{S}$

$$\begin{aligned} &P_L : \mathsf{P}^n \to [0, 1], \\ &P_L(e) = \frac{1}{|\mathscr{S}|} \sum_{s \in \mathscr{S}} P(es) . \end{aligned} \quad (5)$$

In standard error correction, it is assumed that the logical channel is known, and the task is to find a good decoding for each syndrome. Here however, we will consider a "reverse" problem: Given (an estimate of) the syndrome statistics, can we (uniquely) obtain the logical channel $P_L$? Perhaps surprisingly, we will show that this is possible as long as the noise affecting the code is correctable in a certain sense.

## II. MOMENTS

To tackle this estimation problem, we will first switch our description of $P$ via a Fourier transform. The Fourier transform $\mathscr{F}[f]$ of a function $f : \mathsf{P}^n \to \mathbb{R}$ is defined as

$$\begin{aligned} &\mathscr{F}[f] : \mathsf{P}^n \to \mathbb{R}, \\ &\mathscr{F}[f](a) = \sum_{e \in \mathsf{P}^n} \langle a, e \rangle f(e) . \end{aligned} \quad (6)$$

This is also sometimes called Walsh-Hadmard transform [12]. From the definition, we see that for any stabilizer $s \in \mathscr{S}$, $\mathscr{F}[P](s)$ is exactly the expectation of $s$ in repeated rounds of error correction. It can thus be computed from the measured syndrome statistics. In analogy, we denote $E = \mathscr{F}[P]$ and call this the set of moments, i.e. there is one moment $E(a)$ for each $a \in \mathsf{P}$. One should however keep in mind that only the moments corresponding to stabilizers can be measured without destroying the encoded information. Since the Fourier transform is an invertible transformation, with inverse given by

$$\mathscr{F}^{-1}[f](e) = \frac{1}{|\mathsf{P}^n|} \sum_{a \in \mathsf{P}^n} \langle a, e \rangle f(a) , \quad (7)$$

knowing all moments $E$ is equivalent to knowing the complete error distribution $P$.

Since we are only interested in learning the logical channel, only a subset of all moments needs to be estimated. These are exactly the moments corresponding to logical operators. To see why this is the case, let us first introduce the convolution on $\mathsf{P}^n$. For two functions $f, g : \mathsf{P}^n \to \mathbb{R}$, their convolution is defined by

$$(f * g)(e) = \sum_{e' \in \mathsf{P}} f(e') g(ee') . \quad (8)$$

As expected, it can be shown that convolutions transform into products under Fourier transform:

$$\mathscr{F}[f * g] = \mathscr{F}[f] \cdot \mathscr{F}[g] \,. \tag{9}$$

The logical channel $P_L$, defined in (5), can be written as the convolution of the physical channel $P$ with the uniform probability distribution over stabilizers $U_{\mathscr{S}}$,

$$P_L = P * U_{\mathscr{S}} \,. \tag{10}$$

It is well known that $\mathscr{F}[U_{\mathscr{S}}] = \Phi_{\mathscr{S}^\perp} = \Phi_{\mathscr{L}}$, where $\Phi_{\mathscr{L}}$ is the indicator function of $\mathscr{L}$ [32]. Therefore the logical channel can be characterized in Fourier space by the moments

$$E_{\mathscr{L}} \coloneqq E \cdot \Phi_{\mathscr{L}} \,. \tag{11}$$

This is a special instance of the averaging/subsampling-duality explained in [32]. To summarize the above discussion, the logical channel is fully characterized by the moments corresponding to logical operators. The estimation problem can then be phrased as follows: Given the moments $E_{\mathscr{S}}$ of all stabilizers, compute the moments $E_{\mathscr{L}}$ of all logical operators.

## III. CORRECTABLE NOISE

The above estimation problem cannot be solved for arbitrary channels $P$, since in general the moments are independent of each other. Here, our assumption of limited correlations becomes important.

To formalize this assumption, consider a set of *supports* $\Gamma \subseteq 2^{\{1,\dots,n\}}$, where $2^{\{1,\dots,n\}}$ denotes the power-set of $\{1,\dots,n\}$. These supports are allowed to overlap with each other. We assume that on each support $\gamma \in \Gamma$, there acts an independent Pauli channel $P_\gamma : \mathsf{P}^\gamma \to [0,1]$. Thus, the noise is correlated across each support, but not between different supports. If the supports are small, any high weight error must arise as a combination of independent lower weight errors. This is the scenario where error correction has a chance to improve the fidelity. On the other hand, if the supports are too large, error correction usually fails. Thus, we assume that the noise is *correctable* in the following sense.

**Definition 1** (Correctable noise)**.** *A Pauli channel $P$ described by a set of supports $\Gamma \subseteq 2^{\{1,\dots,n\}}$ is called* correctable *if the following two conditions are fulfilled:*

- *For all $\gamma_1, \gamma_2 \in \Gamma$, the union $\gamma_1 \cup \gamma_2$ is a correctable region.*
- *$P(I) > \frac{1}{2}$.*

We see from the definition of distance that the first condition is fulfilled in particular if $|\gamma| \leq \lfloor \frac{d-1}{2} \rfloor$ for all $\gamma \in \Gamma$. The second condition simply states that the total error rate should not be too large. It guarantees that all moments are positive, i.e. $E(a) > 0$ for all $a \in \mathsf{P}^n$. Note that it is fulfilled in particular if all the independent channels have sufficiently low error rates, namely $P_\gamma(I) > \frac{1}{2}$ for all $\gamma \in \Gamma$.

Since the multiplication of independent Pauli random variables corresponds to a convolution of their probability distributions, the full channel $P$ can be written as a convolution of the independent local channels,

$$P = \ast_{\gamma \in \Gamma} P_\gamma. \tag{12}$$

In this notation, we set $P_\gamma(e) = 0$ if $\mathrm{supp}(e) \not\subseteq \gamma$. In order to better capture this structure in Fourier space, we can introduce a set of *canonical moments* $F$ (which we called "transformed moments" before [26]). For $a, b \in \mathsf{P}^n$, let us write $b \leq a$ if $b$ is a substring of $a$. Then we define the canonical moments as

$$\begin{aligned} &F : \mathsf{P}^n \to \mathbb{R} \,, \\ &F(a) = \prod_{b \in \mathsf{P}^n : b \leq a} E(b)^{\mu(b,a)} \,, \end{aligned} \tag{13}$$

where $\mu$ is the Möbius function defined by

$$\mu(b,a) = \begin{cases} (-1)^{|a|-|b|}, & b \leq a \\ 0, & \text{otherwise} \end{cases} \,, \tag{14}$$

which is well known in combinatorics [33]. The Möbius function is defined in such a way that in eq. (13), we divide out that part of the moment $E(a)$ that is already described by substrings $b \leq a$, without "double counting" any substring. Essentially, while the regular moments $E$ also capture correlations across all subsets of their support, the canonical moments only capture correlations across their full support. The advantage is that a small set of canonical moments is sufficient to fully describe the channel. In particular, the following two facts about canonical moments are shown in lemma 13 in the appendix. First of all, we only need to consider the canonical moments that lie completely inside a channel support $\gamma$, since

$$F(a) = 1 \text{ if } \mathrm{supp}(a) \not\subseteq \gamma \; \forall \gamma \in \Gamma \,. \tag{15}$$

The set of such canonical moments is $F_{\Gamma'} = (F(a))_{a \in \Gamma'}$, where

$$\Gamma' = \{a \in \mathsf{P} \,:\, \exists \gamma \in \Gamma \text{ such that } \mathrm{supp}(a) \subseteq \gamma\} \,. \tag{16}$$

Furthermore, the regular moments $E$ are obtained from the canonical moments $F$ by

$$E(a) = \prod_{b \leq a} F(b) \,. \tag{17}$$

## IV. IDENTIFIABILITY

Since the moments $E_{\mathscr{S}}$ can be obtained from the syndrome measurements, and the channel is fully described

by the canonical moments $F_{\Gamma'}$, estimation of the physical channel boils down to solving the system of equations

$$E(s) = \prod_{a\in\Gamma',a\subseteq s} F(a)\,. \tag{18}$$

For correctable noise, all moments are positive. Then, eq. (18) can be transformed into a system of linear equations by taking logarithms. This system can be expressed by the coefficient matrix $D_{\mathscr{S}}$, whose rows are labeled by stabilizers and whose columns are labeled by elements of $\Gamma'$, with entries

$$D_{\mathscr{S}}[s,a] = \begin{cases} 1, & a\subseteq s \\ 0, & \text{otherwise} \end{cases}\,. \tag{19}$$

As we have proven before [26], a unique solution exists if the range of correlations of the error channel $P$ is smaller than the pure distance of the code. Correctable noise generally does not fulfill this strict condition. Thus, the system is underdetermined and the physical channel $P$ cannot be estimated just from the syndrome measurements.

We are, however, only interested in estimating the logical channel (5), which contains less information. As derived in section II, it suffices to consider the moments $E_{\mathscr{L}}$. The question is now whether the moments $E_{\mathscr{L}}$ can be computed from the measured moments $E_{\mathscr{S}}$, i.e. whether the corresponding equations of the form eq. (18) are linearly dependent after taking logarithms. In other words, the logical channel can be uniquely estimated from the syndrome measurements if

$$\operatorname{rank}(D_{\mathscr{S}}) = \operatorname{rank}(D_{\mathscr{L}})\,. \tag{20}$$

This condition is equivalent to $\operatorname{rank}(D_{\mathscr{S}}^T D_{\mathscr{S}}) = \operatorname{rank}(D_{\mathscr{L}}^T D_{\mathscr{L}})$. We will prove this by showing the even stronger statement

$$D_{\mathscr{S}}^T D_{\mathscr{S}} \propto D_{\mathscr{L}}^T D_{\mathscr{L}}\,. \tag{21}$$

First, note that $D_{\mathscr{S}}^T D_{\mathscr{S}}$ can be easily computed from its definition,

$$D_{\mathscr{S}}^T D_{\mathscr{S}}[a,b] = |\{s\in\mathscr{S} \,:\, a\le s \text{ and } b\le s\}|\,. \tag{22}$$

The analogous statement holds for $D_{\mathscr{L}}$. By rewriting eq. (21) in terms of individual entries, we see that the logical channel can be uniquely estimated from the syndrome statistics if for all $a,b\in\Gamma'$,

$$\begin{aligned} &|\{s\in\mathscr{S} \,:\, a\le s \text{ and } b\le s\}| \\ &= c\,|\{l\in\mathscr{S}^\perp \,:\, a\le l \text{ and } b\le l\}|\,, \end{aligned} \tag{23}$$

where $c$ is a constant independent of $a,b$. This is a counting problem that depends only on global properties of the stabilizers and logical operators, but not on their specific form. To solve this counting problem, we will employ the well known cleaning lemma, which was first stated by Bravyi and Terhal [34]. Informally, this lemma states that any correctable region can be cleaned from logical operators.

**Lemma 2** (Cleaning Lemma)**.** *Let $R$ be a correctable region. Then any coset $[l]\in\mathscr{L}/\mathscr{S}$ of logical operators has a representative $l\in\mathscr{L}$ that has no support on $R$, i.e.* $\operatorname{supp}(l)\cap R = \emptyset$.

Using this lemma, we can prove eq. (23). For all $a,b\in\Gamma'$ we have,

$$\begin{aligned} &|\{l\in\mathscr{L} \,:\, a\le l \text{ and } b\le l\}| \\ &= \sum_{l\in\mathscr{L}} [a\le l \text{ and } b\le l] \\ &= \sum_{[l]\in(\mathscr{L}/\mathscr{S})} \sum_{s\in\mathscr{S}} [a\le ls \text{ and } b\le ls] \\ &= \sum_{[l]\in(\mathscr{L}/\mathscr{S})} \sum_{s\in\mathscr{S}} [a\le s \text{ and } b\le s] \\ &= |\mathscr{L}/\mathscr{S}|\cdot|\{s\in\mathscr{S} \,:\, a\le s \text{ and } b\le s\}|\,. \end{aligned}$$

In the second equality, we split the total sum into smaller sums over logically equivalent subsets of logical operators. Then, the third equality follows from the cleaning lemma: since $a$ and $b$ correspond to canonical moments, they must be fully contained in some supports $\gamma_a,\gamma_b\in\Gamma$. For correctable noise, $\gamma_a\cup\gamma_b$ is a correctable region. Thus, the union of the supports of $a$ and $b$ is fully contained in $\gamma_a\cup\gamma_b$, it must also be a correctable region. By the cleaning lemma, we can choose the representative $l$ of the coset $[l]$ such that it acts trivially on that region. Then, $a$ is a substring of $ls$ iff it is a substring of $s$, and the same holds for $b$. This finishes the proof of eq. (23).

We can summarize the discussion of the main text in the following theorem:

**Theorem 3.** *A Pauli channel $P$ can be estimated up to logical equivalence from the syndrome measurements of a stabilizer code if $P$ is correctable in the sense of definition 1.*

Note that while we focused on stabilizer codes with perfect measurements for simplicity, several generalizations of this result are possible. Measurement errors can be incorporated using the framework of quantum data-syndrome codes [35]. Furthermore, we can also consider subsystem codes [36], which generalize stabilizer codes by allowing for some non-commuting measurements. A full account of these generalizations, including all proofs that are omitted in the main text, is given in the appendix. The main theorem presented there might also be interesting in contexts other than QEC.

## V. CONCLUSION

We have shown that the measurements performed during QEC contain enough information to estimate a large class of phenomenological Pauli noise models up to logical equivalence. Informally, as long as the code can correct the noise, it can also be estimated from the syndrome

measurements. This result opens up new characterization possibilities since the previous results have focused only on estimating physical channels. Our result applies to data-syndrome codes and general subsystem codes, which encompass most codes in the literature.

The focus of this work is on phenomenological noise models. For quantum communication or storage, this might be a reasonable assumption. In the context of fault-tolerant quantum computing, however, full circuit level noise models are more realistic than phenomenological ones, which introduces additional complications already for decoding in the first place. A common approach to this problem is to consider approximate noise models. For example, a minimum-weight perfect matching decoder maps the actual noise to a simplified graph with weighted edges [23, 37]. Here, our results apply directly, and the edge weights can be estimated up to logical equivalence by solving our equation system (18).

The situation is less clear if one is interested in more details than such an effective noise model provides. In this case, one might attempt to transfer our results using a cut-off for late errors, following Delfosse *et al.* [38], or using a mapping from circuit noise to subsystem codes, as given in Refs. [39–41]. We think that our work can serve as a basis for many possible research questions on characterization in the context of QEC.

## ACKNOWLEDGMENTS

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – Cluster of Excellence Matter and Light for Quantum Computing (ML4Q) EXC 2004/1 – 390534769. The work of D.B. and H.K. is also supported by the German Federal Ministry of Education and Research (BMBF) within the funding program "Quantum technologies – from basic research to market" in the joint project QSolid (grant number 13N16163). The work of M.K. is also supported by the DFG via the Emmy Noether program (grant number 441423094) and by the German Federal Ministry of Education and Research (BMBF) within the funding program "Quantum technologies – from basic research to market" in the joint project MIQRO (grant number 13N15522).

## APPENDIX

In this appendix, we give a self contained account of our results in a generalized setting. The core arguments are similar to the main text. The main difference is that we distinguish between the set of accessible measurements and the set of stabilizers (or gauge group), which describes logical equivalence. Consequently, a more general version of the cleaning lemma is needed. All proofs omitted in the main text are also provided in this generalized setting. Finally, we apply the result to the classes of subsystem codes, which encompasses most quantum-error correction codes that have been constructed, and quantum data-syndrome codes, which allows for a treatment of measurement errors.

### Notation

We denote as $[n] \coloneqq \{1, \dots, n\}$ the set of the first $n$ positive integers. The field with two elements is denoted $\mathbb{F}_2$. For a statement $Q$, we denote with $[Q]$ the Iverson bracket of $Q$, which takes the value 1 if $Q$ is true and 0 if $Q$ is false. The powerset of a set $A$ is the set of all subsets of $A$, including the empty set, and it is denoted as $2^A$. We denote the four Pauli matrices as $I = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, $X = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, $Y = \left(\begin{smallmatrix} 0 & -i \\ i & 0 \end{smallmatrix}\right)$ and $Z = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. We also use $I$ for the generic identity matrix, or a generic identity element of a group.

### 1. Mathematical background

For the discussions of stabilizer quantum-error correction, some background on the Pauli group will be useful. The *n-qubit Pauli group* $\mathcal{P}^n$ is the group generated by tensor products of Pauli operators and the imaginary unit, i.e.,

$$\mathcal{P} \coloneqq \{\alpha \bigotimes_{i=1}^{n} e_i \,:\, e_i = \{I, X, Z, Y\}, \alpha \in \{\pm 1, \pm \mathrm{i}\}\}\,. \quad (24)$$

Since phases can often be disregarded, we also work with the *effective Pauli group*

$$\mathsf{P}^n \coloneqq \mathcal{P}/\{\pm 1, \pm i\}\,. \quad (25)$$

This in an Abelian group.

In QEC, errors and stabilizer measurements are often described via an isomorphism $\mathsf{P}^n \to \mathbb{F}_2^{2n}$ and a scalar product on $\mathbb{F}_2^{2n}$. We will not make use of this identification, and instead express these concepts using group characters of finite Abelian groups. We give a short introduction here and collect the most important facts for our purposes. A more thorough description can be found for example in Refs. [32, 42].

A *group character* of a finite Abelian group $A$ is a group homomorphism

$$\chi : A \to S^1\,, \quad (26)$$

where $S^1 \coloneqq \{c \in \mathbb{C} : |c| = 1\}$ is the unit circle. Group characters themselves form a group $\widehat{A}$ under pointwise multiplication, called the dual group of $A$. Pontryagain duality guarantees that for any locally compact and hence for any finite Abelian group, $\widehat{A}$ is isomorphic to $A$. Thus, we can express group characters by elements of the original group. This notion can be expressed by a *bicharacter*. In the context of QEC, bicharacters express measurement outcomes of stabilizer measurements.

**Definition 4.** *A* bicharacter *of a finite Abelian group $A$ is a map*

$$\langle \cdot, \cdot \rangle : A \times A \to S^1 , \tag{27}$$

*such that the map $a \mapsto \langle a, \cdot \rangle$ is an isomorphism of $A$ and $\widehat{A}$.*

This is similar to a scalar product, although we often have $\langle a, a \rangle = +1$. Thus, we also have a notion of "orthogonal complement", which is the annihilator. For a subgroup $B \subseteq A$, we define the *annihilator* of $A$ as

$$B^\perp = \{a \in A \,:\, \langle a, b \rangle = +1 \,\forall b \in B\} . \tag{28}$$

We always have $(B^\perp)^\perp = B$. Furthermore, taking the annihilator reverses the order of inclusions. That is, for any two subgroups $B, C \subseteq A$, if $B \subseteq C$, then $C^\perp \subseteq B^\perp$. In contrast to a scalar product, it is possible that $B \subseteq B^\perp$.

Using the bicharacter $\langle \cdot, \cdot \rangle$, we can define the *Fourier transform* of a map $f : A \to \mathbb{C}$ as

$$\begin{aligned} &\mathscr{F}[f] : A \to \mathbb{C} \\ &\mathscr{F}[f](a) = \sum_{b \in A} \langle a, b \rangle f(b) . \end{aligned} \tag{29}$$

This is an invertible transformation with inverse

$$\begin{aligned} &\mathscr{F}^{-1}[f] : A \to \mathbb{C} \\ &\mathscr{F}^{-1}[f](a) = \frac{1}{|A|} \sum_{b \in A} \langle b, a^{-1} \rangle f(b) . \end{aligned} \tag{30}$$

Furthermore, we will use the convolution of two maps $f, g : A \to \mathbb{C}$, which is defined as

$$(f * g)(a) = \sum_{b \in A} f(b) g(ab^{-1}) . \tag{31}$$

As expected, convolutions are mapped to products by the Fourier transform, i.e.

$$\mathscr{F}[f * g] = \mathscr{F}[f] \cdot \mathscr{F}[g] . \tag{32}$$

For any subgroup $B \subseteq A$, we denote with $\Phi_B$ the indicator function of $B$, i.e. $\Phi_B(a) = 1$ if $a \in B$ and $\Phi_B(a) = 0$ otherwise. Furthermore, we denote the scaled indicator function as $U_B \coloneqq \frac{1}{|B|} \Phi_B$, which is the uniform probability distribution on $B$. It can be shown that the following duality holds.

**Lemma 5** (Ref. [32])**.** *For any subgroup $B \subseteq A$ of an Abelian group $A$:*

$$\mathscr{F}[U_B] = \Phi_{B^\perp} \tag{33}$$

$$\mathscr{F}[\Phi_B] = |A| U_{B^\perp} . \tag{34}$$

All important groups considered in this work have a direct product structure, i.e.

$$A = \prod_{i=1}^{n} A_i . \tag{35}$$

We will then always use the *product bicharacter* on $A$, which is given by the product of bicharacters on the $A_i$,

$$\langle a, b \rangle = \prod_{i=1}^{n} \langle a_i, b_i \rangle \tag{36}$$

for any $a = (a_1, a_2, \ldots, a_n) \in A$ and similar $b$. The *support* of an element $a \in A$ is

$$\operatorname{supp}(a) = \{i \in [n] \,:\, a_i \neq I\} . \tag{37}$$

We will say that $a$ is *supported* on a region $R \subseteq [n]$ if $\operatorname{supp}(a) \subseteq R$. The corresponding subgroup to a region $R$ is denoted as $A_R \coloneqq \prod_{i \in R} A_i$. This is naturally embedded as a subgroup in $A$. The *complement* of $R \subseteq [n]$ is denoted as $R^c = [n] \backslash R$. If we use the product bicharacter on $A$, we have that

$$A_R^\perp = A_{R^c} . \tag{38}$$

Given an element $a \in A$, we denote with $a_R$ its restriction to $R$, i.e. $a_R = a$ on $R$ and $a_R = 1$ on $R^c$.

Finally, we will be interested in functions with local support. Given a function $f_R : A_R \to \mathbb{C}$, there are two important ways to extend it to a function $f : A \to \mathbb{C}$. The first is to set $f(a) = 0$ if $a \notin A_R$. This is called the *impulsive* extension. The second is the to set $f(a) = f_R(a_R)$, which is called *periodic* extension. These two possibilities transform into each other under Fourier transform.

**Lemma 6** (Ref. [32])**.** *Let $f_R : A_R \to \mathbb{C}$ and $g_R : A_R \to \mathbb{C}$ be its Fourier transform (on $A_R$). Let $f$ be the impulsive extension of $f_R$ and $g$ be the periodic extension of $g_R$. Then $\mathscr{F}[f] = g$.*

We will mainly work with three groups. These are the effective Pauli group $\mathsf{P}^n$, the group of bit-strings $\mathbb{F}_2^m$, and their direct product $\mathsf{G}^{n,m} \coloneqq \mathsf{P}^n \times \mathbb{F}_2^m$. Since all elements of these groups have order two, bicharacters of these groups will only take values $\pm 1$.

For $\mathsf{P}^n$, the bicharacter encodes commutation relations, and is also called scalar commutator,

$$\langle a, e \rangle = \begin{cases} +1, & a \text{ and } e \text{ commmute in } \mathcal{P}^n \\ -1, & a \text{ and } e \text{ anti-commmute in } \mathcal{P}^n \end{cases} . \tag{39}$$

Note that this is the product bicharacter when we view $\mathsf{P}^n = \prod_{i=1}^n \mathsf{P}^1$. On $\mathbb{F}_2^m$, we use the bicharacter that is related to the usual scalar product,

$$\langle e, f\rangle = (-1)^{\sum_i e_i f_i}\,, \tag{40}$$

and again this coincides with the product bicharacter. Finally, on $\mathsf{G}^{n,m}$, we directly use the product bicharacter

$$\langle (a,e),(b,f)\rangle = \langle a,b\rangle \cdot \langle e,f\rangle\,. \tag{41}$$

### 2. Setting and main result

Now, we state and prove our main result in an abstract setting first. For ease of exposition, we still stick to terminology close to that of QEC.

We consider the group $A = \prod_{i=1}^n A_i$, where each $A_i$ is either $\mathsf{P}$ or $\mathbb{F}_2$. It comes equipped with the product bicharacter. Both errors and measurements are described as element of $A$.

We are interested in estimating an error channel described by a probability distribution $P : A \to [0,1]$. For this purpose, we assume that we have access to a group of *measurements* $\mathscr{M} \subseteq A$. The assumption that the measurements from a group is relatively weak. In the context of QEC, we measure a set of generators and all other outcomes are defined by products of the generator outcomes. We will perform multiple rounds of measurements, and assume that before each round an independent error $e \in A$ occurs. The outcome of measurement $s \in \mathscr{M}$ is described by $\langle s, e\rangle$. We will refer to this as a *phenomenological* noise model, since errors are independent and identically distributed between rounds and no new errors arise during the round of measurements. Errors that give a $+1$ outcome for every measurement $s \in \mathscr{M}$ are called *undetectable*. The set of undetectable errors is exactly $\mathscr{U} = \mathscr{M}^\perp$. Furthermore, we will only be interested in estimating the channel up to some logical equivalence, described by a subgroup $\mathscr{G} \subseteq A$ which we will call *gauge group*. An overview of these groups and their relations is give in fig. 1. Errors differing only by an element $s \in \mathscr{G}$ are considered logically equivalent. More precisely, we are interested in estimating the *logical channel* $P_L$ which is obtained by averaging over cosets of $\mathscr{G}$, resulting in

$$\begin{aligned} &P_L : A \to [0,1]\,,\\ &P_L(e) = \frac{1}{|\mathscr{G}|}\sum_{s\in\mathscr{G}} P(es)\,. \end{aligned} \tag{42}$$

In the setting of stabilizer codes, $P_L$ describes the action of the noise on the encoded information. The logical channel can be conveniently expressed as a the convolution

$$P_L = P * U_{\mathscr{G}} \tag{43}$$

of $P$ with the uniform distribution $U_{\mathscr{G}}$. Complementary to the gauge group, we define $\mathscr{L} \coloneqq \mathscr{G}^\perp$ and call this the set of *logical operators*. Finally, we will require that these sets are related by the dual inclusion relations $\mathscr{G} \subseteq \mathscr{U}$ and $\mathscr{M} \subseteq \mathscr{L}$ (one implies the other by taking annihilators). The condition $\mathscr{G} \subseteq \mathscr{U}$ means that logically equivalent errors must have the same measurement outcomes.

$$\begin{array}{ccc} \mathscr{G} & \subseteq & \mathscr{U} \\ {\scriptstyle\perp}\uparrow\downarrow & & \uparrow\downarrow{\scriptstyle\perp} \\ \mathscr{L} & \supseteq & \mathscr{M} \end{array}$$

Figure 1. An overview over the abstract setting, described by four groups. The main ingredients are a group of measurements $\mathscr{M}$ and a gauge group $\mathscr{G}$. The gauge group describes which errors are considered logically trivial. The annihilator of $\mathscr{M}$ is the group of undetectable errors $\mathscr{U} = \mathscr{M}^\perp$, and the annihilator of $\mathscr{G}$ is the group of logical operators $\mathscr{L} = \mathscr{G}^\perp$. The groups fulfill the dual inclusions $\mathscr{G} \subseteq \mathscr{U}$ and $\mathscr{M} \subseteq \mathscr{L}$.

In this setting, our main result can be stated as follows:

**Theorem 7.** *Let $\mathscr{G}, \mathscr{M} \subseteq A$ be a gauge group and measurement group fulfilling $\mathscr{G} \subseteq \mathscr{M}^\perp$. If the error channel $P$ is correctable in the sense of definition 9, then the logical channel $P_L$ defined by the gauge group $\mathscr{G}$ can be uniquely estimated from the expectations of the measurements $\mathscr{M}$.*

To recover the result for stabilizer codes, as treated in the main text, we set $A = \mathsf{P}^n$. The set of measurements $\mathscr{M}$ and the gauge group $\mathscr{G}$ are then identical, both equal to the stabilizer group of the code. Consequently, the undetectable errors $\mathscr{U}$ and the logical operators $\mathscr{L}$ also coincide, and are both given by the logical operators of the code. Later, we will also explain how to specialize our setting to the more general classes of subsystem codes (appendix 9) and quantum data-syndrome codes (appendix 10) with phenomenological noise.

### 3. Moments

We start the proof of theorem 7 by describing the estimation problem in Fourier space. We define the moments

$$E = \mathscr{F}[P]\,. \tag{44}$$

Since the Fourier transform is invertible, the set of all moments $(E(a))_{a\in A}$ fully characterizes the channel $P$. Furthermore, for an element $s \in \mathscr{M}$, $E(s)$ is the expectation of the measurement of $s$ in repeated rounds. Thus, the moments corresponding to $\mathscr{M}$ can be obtained from our measurements.

The logical channel can also be fully characterized by a subset of moments. Applying a Fourier transform to eq. (43) and using lemma 5 results in

$$\mathscr{F}[P_L] = E \cdot \Phi_{\mathscr{L}}\,. \tag{45}$$

Thus, to obtain the logical channel, we have to compute all the moments corresponding to $\mathscr{L}$, while we can only

measure moments corresponding to $\mathscr{M}$. We will see that this is indeed possible, assuming that the error channel is not correlated over too large regions. These assumptions on the noise are formalized in the next section.

## 4. Local noise

To formalize the assumption of limited correlations, we assume the total error in each round is a product of many local errors that occur independently. The noise then factorizes into a set of local channels, characterized by the corresponding set of local supports $\Gamma \subseteq 2^{[n]}$. For each $\gamma \in \Gamma$, there is a local channel $P_\gamma : A_\gamma \mapsto [0,1]$. We extend the local channels impulsively to $A$, i.e. we set $P_\gamma(e) = 0$ if $e \notin A_\gamma$. The total error distribution is then given by

$$P = \ast_{\gamma\in\Gamma} P_\gamma \,. \tag{46}$$

Denoting $E_\gamma = \mathscr{F}[P_\gamma]$, we obtain

$$E = \prod_{\gamma\in\Gamma} E_\gamma \,. \tag{47}$$

Here, each $E_\gamma$ must be extended periodically from $A_\gamma$ to $A$, due to lemma 6. More explicitly, we have $E_\gamma(e) = E_\gamma(e_\gamma)$.

We assume that the individual regions are small enough to not support logically non-trivial undetectable errors. To formalize this, we define a notion of correctable region, which is inspired by the setting of topological codes [34].

**Definition 8.** *A region $R \subseteq [n]$ is called* correctable *if every undetectable error $e \in \mathscr{U}$ supported on $R$ is logically trivial, i.e. $e \in \mathscr{G}$.*

Using this definition, we can state our assumptions on the noise.

**Definition 9.** *A channel of the form (46) is called* correctable *if the following two conditions are fulfilled:*

1. *For all $\gamma_1, \gamma_2 \in \Gamma$, $\gamma_1 \cup \gamma_2$ is a correctable region.*
2. *All moments are positive, i.e. $E(a) > 0$ for all $a \in A$.*

The first condition states that correlations can not be so large that uncorrectable errors occur frequently. Instead uncorrectable errors should only be allowed to occur as a combination of many smaller independent errors. We will later relate this condition to the distance of a code. The second condition essentially states that the total error rate is not too large. A sufficient conditions for this is $P(I) > \frac{1}{2}$, which is guaranteed to be fulfilled if $P_\gamma(I) > \frac{1}{2}$ for all $\gamma \in \Gamma$.

## 5. Canonical moments

The factorization (47) can be used to find a more compact characterization of the moments $E$. Intuitively, we note that the moment $E(a)$ captures correlations across all substrings of $a$. In particular, $E(a)$ can be non-trivial even if $a$ is not contained in a support $\gamma \in \Gamma$ of our noise model. We will find an alternative set of moments $F : A \to \mathbb{R}$, called *canonical moments*, that only capture correlations across their whole support. In particular, the canonical moments fulfill $F(a) = 1$ if $a$ is not contained in a support. Thus, a small set of low weight canonical moments is sufficient to fully characterize the channel.

Formally, we define the canonical moments by a Möbius inversion. Möbius inversion is a generalization of the inclusion-exclusion principle of combinatorics [33]. Essentially, we divide out correlations on substrings of $a$ from the moment $E(a)$, while being careful not to double count any substrings. This leaves only correlations across the full support. In order to do this, we consider $A$ as a partially ordered set (poset), where the ordering is the substring relation.

**Definition 10** (substring ordering)**.** *We say $a \in A$ is $a$* substring *of $b \in A$ if for all $i \in [n]$ either $a_i = I$ or $a_i = b_i$. In this case we write $a \leq b$.*

We will need the *Möbius function* of this poset. For our purposes, the Möbius function is defined to be the function fulfilling the following inversion theorem, which can be found e.g. in Ref. [33, Theorem 5.5] or [43] in case of the multiplicative version.

**Definition 11** (Möbius function and Möbius inversion)**.** *Let $S$ be a partially ordered set. The Möbius function $\mu$ of $S$ is the function $\mu : S \times S \mapsto \mathbb{R}$ such that for any two functions $f, g : S \mapsto \mathbb{R}$,*

$$f(t) = \prod_{s\leq t} g(s) \,, \tag{48}$$

*if and only if*

$$g(t) = \prod_{s\leq t} f(a)^{\mu(s,t)} \,. \tag{49}$$

In our setting, we obtain the following.

**Lemma 12.** *The Möbius function of $(A, \leq)$ is given by*

$$\mu : A \times A \to \mathbb{R} \,,$$
$$\mu(b,a) = \begin{cases} (-1)^{|a|-|b|} & \text{if } b \leq a \,, \\ 0 & \text{otherwise.} \end{cases} \tag{50}$$

*Proof.* For any given $a \in A$, the poset $\{b \in A : b \leq a\}$ is isomorphic to the poset $\{s \subseteq \mathrm{supp}(a)\}$ ordered by set inclusion. The Möbius function of this is well known to be $\mu(s,t) = (-1)^{|t|-|s|}$. Alternatively, one can use [33, Proposition 5.4] and induction. □

Now, we define the *canonical moments* as

$$F(a) \coloneqq \prod_{b\leq a} E(b)^{\mu(b,a)}\,. \tag{51}$$

This definition essentially corresponds to the canonical factorization of a factor graph describing the moments, compare Refs. [44, 45]. The canonical moments have two important properties.

**Lemma 13** (Properties of canonical moments)**.**

1. *The moments can be expressed by the canonical moments as*

$$E(a) = \prod_{b\leq a} F(b)\,. \tag{52}$$

2. *For any $a \in A$ such that $a$ is not contained in any support, i.e.* $\operatorname{supp}(a) \not\subseteq \gamma$ *for all* $\gamma \in \Gamma$*, we have* $F(a) = 1$.

*Proof.* The first statement is given by the definition 11 of the Möbius function.

Now, we prove the second statement. First, from the definition (51) of the canonical moments $F$ and the decomposition (47) we obtain

$$F(a) = \prod_{b\leq a}\prod_{\gamma\in\Gamma} E_\gamma(b)^{\mu(b,a)} = \prod_{\gamma\in\Gamma}\prod_{b\leq a} E_\gamma(b)^{\mu(b,a)}\,. \tag{53}$$

We can evaluate the second product by splitting it into products over $B = \{b \in A_\gamma : b \leq a\}$ and $B^c = \{c \in A_{\gamma^c} : c \leq a\}$ as

$$\begin{aligned}\prod_{b\leq a} E_\gamma(b)^{\mu(b,a)} &= \prod_{b'\in B}\prod_{c'\in B^c} E_\gamma(b')^{\mu((b',c'),a)}\\ &= \prod_{b'\in B} E_\gamma(b')^{\sum_{c'\in B^c}\mu((b',c'),a)}\,,\end{aligned}$$

where we have used the periodicity $E_\gamma(b) = E_\gamma(b_\gamma)$ and have denoted $b = (b', c') \in A = A_\gamma \times A_{\gamma^c}$. From the explicit expression (50) for the Möbius function we obtain

$$\begin{aligned}&\sum_{c'\in B^c}\mu((b',c'),a) = \mu(b',a)\sum_{c'\in B^c}(-1)^{|c'|}\\ &= \mu(b',a)\sum_{0\leq w\leq|\gamma^c\cap\operatorname{supp}(a)|}(-1)^{|c|}\binom{|\gamma^c\cap\operatorname{supp}(a)|}{w}\\ &= \mu(b',a)\,[|\gamma^c\cap\operatorname{supp}(a)| = 0] = \mu(b',a)\,[\operatorname{supp}(a)\subseteq\gamma]\,,\end{aligned}$$

where in the second equality we sorted the elements of $B^c$ by their weight, and the third equality follows from the fact that $\sum_{i=0}^n(-1)^i\binom{n}{i} = [n = 0]$.

Putting everything together proves lemma 13. □

We conclude that $E$ and hence the physical channel $P$ is characterized by the set of low weight canonical moments corresponding to

$$\Gamma' \coloneqq \{a \in A : \exists\gamma \in \Gamma \text{ such that } \operatorname{supp}(a) \subseteq \gamma\}\,. \tag{54}$$

Indeed, we can express the regular moments by the equation system

$$E(a) = \prod_{b\in\Gamma':b\leq a} F(b)\,. \tag{55}$$

Since we can measure $E(s)$ for $s \in \mathscr{M}$, the estimation problem can now be phrased in terms of these equations.

## 6. Rank of the coefficient matrix

Since we assumed that all moments are positive, eq. (55) corresponds to a linear system after taking logarithms. This linear system can be compactly expressed by a coefficient Matrix $D$, whose rows are labeled regular moments and whose columns are labeled by canonical moments. For $a \in A$ and $b \in \Gamma'$, the corresponding entry of $D$ is

$$D[a,b] = \begin{cases}1 & b\leq a\\ 0 & \text{otherwise}\end{cases}\,. \tag{56}$$

In particular, the measurements we can perform are characterized by the submatrix $D_{\mathscr{M}}$, whose rows are labeled by element of $\mathscr{M}$, and the logical channel is similarly characterized by the submatrix $D_{\mathscr{L}}$. Note that since $\mathscr{M} \subseteq \mathscr{L}$, $D_{\mathscr{M}}$ is a submatrix of $D_{\mathscr{L}}$. The estimation problem can be solved if the rows of $D_{\mathscr{L}}$ are linearly dependent on the rows of $D_{\mathscr{M}}$.

Based on these considerations, to proof theorem 7, we need to show

$$\operatorname{rank}(D_{\mathscr{M}}) = \operatorname{rank}(D_{\mathscr{L}}) \tag{57}$$

or equivalently

$$\operatorname{rank}(D_{\mathscr{M}}^T D_{\mathscr{M}}) = \operatorname{rank}(D_{\mathscr{L}}^T D_{\mathscr{L}})\,. \tag{58}$$

To show eq. (58), we first reduce it to a counting problem. From the definition (56) of $D$, we obtain that

$$D_{\mathscr{M}}^T D_{\mathscr{M}}[a,b] = |\{s \in \mathscr{M} : a \leq s \text{ and } b \leq s\}|\,, \tag{59}$$

and analogously for $D_{\mathscr{L}}$. The advantage of this formulation is that we only need to consider global properties of groups $\mathscr{M}$ and $\mathscr{L}$, but not the specific form of their elements. We will prove that $D_{\mathscr{M}}^T D_{\mathscr{M}} \propto D_{\mathscr{L}}^T D_{\mathscr{L}}$, i.e. that

$$\begin{aligned}&|\{s \in \mathscr{M} : a \leq s \text{ and } b \leq s\}|\\ &= \alpha|\{l \in \mathscr{L} : a \leq l \text{ and } b \leq l\}|\,,\end{aligned} \tag{60}$$

for a constant $\alpha$ independent of $a, b$.

## 7. A general cleaning lemma

Our proof of eq. (60) relies on an abstract variant of the cleaning lemma which was recently derived by Kalachev

and Sadov [29]. The original cleaning lemma was formulated for stabilizer codes in [34]. It states that any correctable region can be "cleaned" of logical operators, i.e. any logical operator has a representative that it is supported outside of this correctable region. Since all logical operators arise from their representatives by multiplication with stabilizers, this means that on each correctable region, the group of logical operators and the stabilizer group look essentially identical. This then solves the counting problem (60).

We will derive a similar result in our more general setting, however the specifics are a bit different. For standard stabilizer codes, the gauge group $\mathscr{G}$ and the group of measurements are both given by the stabilizer group $\mathscr{S}$. Then $\mathscr{S} \subseteq \mathscr{S}^\perp = \mathscr{L}$, and the cleaning lemma is then a statement about the quotient group $\mathscr{L}/\mathscr{S}$. Since in general, the gauge group and the measurements do not coincide, the situation is more complicated. Instead of the inclusion $\mathscr{S} \subseteq \mathscr{L}$, we will consider the dual inclusions $\mathscr{G} \subseteq \mathscr{U}$ and $\mathscr{M} \subseteq \mathscr{L}$. The cleaning lemma will then be a statement about $\mathscr{L}/\mathscr{M}$. One can also find a similar statement about $\mathscr{U}/\mathscr{G}$.

We start with the abstract version of the cleaning lemma mentioned above. Translated to our setting, it states the following.

**Lemma 14** (abstract cleaning lemma [29])**.** *For any three subgroups $\eta, \xi$ and $\alpha$ of an Abelian group $A$ such that $\xi \subseteq \eta^\perp$, we have*

$$|(\eta^\perp \cap \alpha)/(\xi \cap \alpha)| \cdot |(\xi^\perp \cap \alpha^\perp)/(\eta \cap \alpha^\perp)| = |\xi^\perp/\eta| \,. \quad (61)$$

*Proof.* This follows from Theorem 3.10 of [29], using the following translation. The lattice $L$ is the lattice of subgroups of $A$, where the join of two subgroups is the subgroup generated by their union, and the meet of two subgroups is their intersection. The grading is the $Q^+$-grading given by the size of a group. The quasi-complementation $\dagger$ is the annihilator $\perp$. This is similar to the setting of [29][Section 5.3]. □

As a corollary, we obtain the following "concrete" cleaning lemma.

**Lemma 15** (cleaning lemma)**.** *Let $R \subseteq [n]$ be a correctable region. Then every coset $[l] \in \mathscr{L}/\mathscr{M}$ has a representative $l$ that has no support on $R$, i.e.* $\mathrm{supp}(l) \cap R = \emptyset$.

In order to prove this statement, we make use of a simple technical lemma.

**Lemma 16.** *Let $\eta$ and $\alpha$ be subgroups of an Abelian group $A$ and $\xi \subseteq \eta$. Then there is a canonical embedding*

$$(\eta \cap \alpha)/(\xi \cap \alpha) \to \eta/\xi \quad (62)$$

*Proof.* The embedding is defined by mapping the equivalence class $[a] \in (\eta \cap \alpha)/(\xi \cap \alpha)$ to $[a] \in \eta/\xi$. This map is well-defined: If $[a] = [b]$ in $(\eta \cap \alpha)/(\xi \cap \alpha)$, then $a = bc$ with $c \in (\xi \cap \alpha) \subseteq \xi$, and thus $[a] = [b]$ in $\eta/\xi$. Now we show injectivity. If $[a] = 1$ in $\eta/\xi$, then $a \in \xi$. Since we also have $a \in \alpha$ by definition, it follows $a \in \xi \cap \alpha$, and thus $[a] = 1$ in $(\eta \cap \alpha)/(\xi \cap \alpha)$. □

*Proof of lemma 15.* Since $\mathscr{G} \subseteq \mathscr{U} = \mathscr{M}^\perp$, we can apply the abstract cleaning lemma 14 with $\eta = \mathscr{M}$, $\xi = \mathscr{G}$ and $\alpha = A_R$. We obtain

$$|(\mathscr{U} \cap A_R)/(\mathscr{G} \cap A_R)| \cdot |(\mathscr{L} \cap (A_R)^\perp)/(\mathscr{M} \cap (A_R)^\perp)| = |\mathscr{L}/\mathscr{M}| \,.$$

Since $R$ is correctable, the first term is 1. Thus,

$$|(\mathscr{L} \cap (A_R)^\perp)/(\mathscr{M} \cap (A_R)^\perp)| = |\mathscr{L}/\mathscr{M}| \,. \quad (63)$$

By lemma 16, the group on the left-hand side is embedded in the group on the right hand side. Thus this equation implies that they are actually equal. Since we use the product bicharacter on $A$, $(A_R)^\perp = A_{R^c}$. Thus any element $[l] \in \mathscr{L}/\mathscr{M}$ has a representative $l \in \mathscr{L} \cap (A_{R^c})$, i.e. a representative that has no support on $R$. □

## 8. Cleaning up

Using the cleaning lemma 15, we can now finish the proof of eq. (60), and thus of theorem 7, using similar arguments to the stabilizer code case.

*Proof of theorem 7.* Let $a, b \in \Gamma'$, i.e. $a$ and $b$ correspond to non-trivial canonical moments. Then we have,

$$\begin{aligned} &|\{l \in \mathscr{L} \,:\, a \leq l \text{ and } b \leq l\}| \\ &= \sum_{l \in \mathscr{L}} [a \leq l \text{ and } b \leq l] \\ &= \sum_{[l] \in (\mathscr{L}/\mathscr{M})} \sum_{s \in \mathscr{M}} [a \leq ls \text{ and } b \leq ls] \\ &= \sum_{[l] \in (\mathscr{L}/\mathscr{M})} \sum_{s \in \mathscr{M}} [a \leq s \text{ and } b \leq s] \\ &= |\mathscr{L}/\mathscr{M}| \cdot |\{s \in \mathscr{M} \,:\, a \leq s \text{ and } b \leq s\}| \,, \end{aligned} \quad (64)$$

where we have used the following steps. In the second equality, we split the sum into a sum over cosets $[l]$ of $\mathscr{M}$, where each coset is described by a representative $l \in \mathscr{L}$. The third equality used the cleaning lemma 15 in the following way: By the properties of $\Gamma'$ from eq. (16), the support of $a$ and $b$ must be contained in supports $\gamma_a, \gamma_b \in \Gamma$. Then, by the assumption that the noise is correctable (definition 9), $\mathrm{supp}(a) \cup \mathrm{supp}(b)$ must be a correctable region. Thus, by lemma 15, we can always choose the representative $l$ such that it has no support on $\mathrm{supp}(a) \cup \mathrm{supp}(b)$. Then the substring relations $a \leq ls$ and $b \leq ls$ are only determined by $s$. This finishes the proof of theorem 7. □

We will now discuss some specializations of this theorem for different classes of QEC codes. The case of stabilizer codes was already treated in detail in the main text. We can in fact treat even more general classes of codes.

### 9. Subsystem codes

Subsystem codes [31, 36] are an important generalization of stabilizer codes. We will explain the basic principles following [46]. A subsystem code can be viewed as a stabilizer code where some of the logical qubits are not used to encode information. The corresponding logical operators can be measured without destroying the encoded information. The primary advantage is that this can often lead to stabilizer measurements of lower weights. Furthermore, some fault-tolerant schemes are naturally described in the language of subsystem codes [46]. Finally, the effect of circuit noise can also be expressed in the language of subsystem codes [39–41]. Thus, there is some hope that the following results can also be used to treat circuit noise models instead of phenomenological noise models for stabilizer codes.

A subsystem code can be described by a gauge group $\mathscr{G} \subseteq \mathsf{P}^n$, whose elements act trivially on the encoded information. The gauge group contains the stabilizers as well as the logical operators that only act on the unused logical qubits. Unless the code is a standard stabilizer code, $\mathscr{G}$ is not Abelian (when viewed as a subgroup of $\mathcal{P}^n$). In the effective Pauli group, this can be expressed as $\mathscr{G} \not\subseteq \mathscr{G}^\perp$. The stabilizer group is then the center of the gauge group in $\mathcal{P}^n$. Expressed in the effective Pauli group:

$$\mathscr{S} := \mathscr{G}^\perp \cap \mathscr{G} \,. \tag{65}$$

As for stabilizer codes, error detection is performed by measuring the stabilizer group in each round, resulting in a set of $\pm 1$ outcomes called the syndrome. This can be done by either measuring a set of generators of $\mathscr{S}$, or by splitting the stabilizers into products of possibly non-commuting gauge operators and measuring these gauge operators. Since the gauge operators do not affect the encoded information, the fact that these measurements do not commute does not affect the encoded information, and it might allow for measurements with lower weight than the stabilizer generators.

Operators that affect the encoded information without being detected are called *(dressed) logical operators*. The set of such operators is given by $\mathscr{L}_d = \mathscr{S}^\perp$. If the operators act only on the actual logical qubits, but not on the discarded gauge logical qubits, then they are called *bare logical operators*. The group of bare logical operators is $\mathscr{L}_b = \mathscr{G}^\perp$. As usual, the distance of the code is defined as the smallest weight of an undetectable error that non-trivially affects the logical information, i.e. as the minimal weight of an element of $\mathscr{L}_d \setminus \mathscr{G}$.

We consider a subsystem code subject to phenomenological noise, where before each error correction round an error $e \in \mathsf{P}^n$ occurs according to some distribution $P(e)$. Here, we assume perfect measurements. In each round, the syndrome of the data error that was accumulated over all previous rounds is measured, and thus the errors in different rounds are not independent. However, if we consider the syndromes relative to the syndrome of the previous round, then we only detect the new errors. The same effect is achieved by tracking the Pauli frame, or by applying a correction between rounds that returns the state to the code space. Thus, in each round we can obtain the measurement outcomes $\langle s, e\rangle$ for each $s \in \mathscr{S}$. This means that set of available measurements $\mathscr{M}$ is exactly the stabilizer group $\mathscr{S}$.

To summarize, we can apply theorem 7 to subsystem codes by setting $\mathscr{G} = \mathscr{G}$, $\mathscr{M} = \mathscr{S}$, $\mathscr{U} = \mathscr{L}_d$ and $\mathscr{L} = \mathscr{L}_b$. Note that eq. (65) implies the inclusion $\mathscr{M} \subseteq \mathscr{G}^\perp$, as required by theorem 7. Furthermore, the definition of distance implies that any region of size at most $d-1$ is correctable in the sense of definition 8. Thus, if the error distribution factorizes into independent channels $P_\gamma$, such that each support $\gamma$ contains no more than $\lfloor \frac{d-1}{2} \rfloor$ qubits, the first part of definition 9 is also fulfilled. If furthermore $P_\gamma(I) > \frac{1}{2}$ for all $\gamma$, then $E_\gamma > 0$ for all $\gamma$ and thus $E = \prod_\gamma E_\gamma > 0$. Then, the channel $P$ is correctable in the sense of definition 9. We obtain the following corollary.

**Corollary 17.** *Phenomenological data noise with error rates smaller than $\frac{1}{2}$ can be estimated up to logical equivalence from the measurements of a subsystem code if the noise is not correlated over more than half the distance of the code.*

### 10. Quantum data-syndrome codes

So far, we have only treated data errors and assumed perfect measurements. Now, we will consider measurement errors in a phenomenological noise model. A simple framework for this is provided by quantum data-syndrome codes [35, 47], which allow for a unified treatment of data and measurement errors. It should however be noted that, while quantum data-syndrome codes capture a large class of fault-tolerant measurement schemes, some adaptive schemes such as flag fault-tolerance [48] are not easily described in this language. Since this section is only concerned with phenomenological noise models, we do not take into account errors that happen during the execution of the measurement circuits and error propagation in these circuits.

To define a quantum data-syndrome code, we first pick an underlying stabilizer code with stabilizer group $\mathscr{S} \subseteq \mathsf{P}^n$. In each round, instead of just a set of generators, a larger set of redundant stabilizers $g_1, \cdots, g_m \in \mathscr{S}$ is measured. The simplest and most common case is to simply repeat the measurements of the generators. More generally, the redundant stabilizers can be chosen according to a classical code, as described in [35]. An error can then be described by a data error $e_d \in \mathsf{P}^n$ and a measurement error $e_m \in \mathbb{F}_2^m$, i.e. $e_m[i] = 1$ if the measurement of $g_i$ returned the wrong outcome and $e_m[i] = 0$ otherwise. The measurements of the generators, including measurement errors, can be described by the extended

parity check matrix

$$H = \begin{bmatrix} G & I_m \end{bmatrix} , \tag{66}$$

where the rows of $G$ are the original stabilizers. That is, each generator $g_i \in \mathsf{P}$ is extended to an element $f_i = (g_i, \hat{i}) \in \mathsf{G}^{n,m}$, where $\hat{i}$ is the $i$-th standard basis vector. Then, the outcome of the measurement of $f_i$ if an error $e = (e_d, e_m) \in \mathsf{G}^{n,m} = \mathsf{P}^n \times \mathbb{F}_2^m$ occurred is exactly given by $\langle f_i, e\rangle$, using the bicharacter of $\mathsf{G}^{n,m} \coloneqq \mathsf{P}^n \times \mathbb{F}_2^m$ (eq. (41)). The set of measurements we have access to is thus the group generated by the extended stabilizers $f_i$,

$$\mathscr{M} \coloneqq \langle f_1, \dots, f_m \rangle . \tag{67}$$

As always, the collection of measurement outcomes for all $s \in \mathscr{M}$ is called the *syndrome*, and it can be obtained by measuring the generators $f_i$ of $\mathscr{M}$. The undetectable errors $\mathscr{U} = \mathscr{M}^\perp$ are exactly those that result in a trivial syndrome.

Similarly to subsystem codes, the concepts of measurements and stabilizers do not coincide. In fact, the measurements do not corresponds to undetectable errors, $\mathscr{M} \not\subseteq \mathscr{U}$. Since errors differing by elements of $\mathscr{M}$ do not have the same syndrome it follows that, in particular, they cannot be considered logically equivalent. Instead, logical equivalence is still described by the stabilizer group $\mathscr{S} \subseteq \mathsf{P}^n$ of the underlying code, which we view as a subgroup of $\mathsf{G}^{n,m}$. The *logical operators* $\mathscr{L}$ are those operators that map the codespace of the underlying quantum code to itself, i.e. $\mathscr{L} \coloneqq \mathscr{S}^\perp$, where the annihilator is in $\mathsf{G}^{n,m}$, not just in $\mathsf{P}^n$. These groups then fulfill the dual inclusion relations $\mathscr{S} \subseteq \mathscr{U}$ and $\mathscr{M} \subseteq \mathscr{L}$.

Motivated by the discussion above, the *distance* of a data-syndrome code is defined as the minimal weight of an element of $\mathscr{U} \setminus \mathscr{S}$ [35]. Remember that a region $R \subseteq [n+m]$ is correctable if there is no element of $\mathscr{U} \setminus \mathscr{S}$ that is supported on $R$ (definition 8). Thus, as expected, if $|R| < d$, $R$ is correctable.

Consider a phenomenological noise model where in each round a new error $e = (e_d, e_m)$ occurs according to a distribution $P$. If we were to always reset to the ground state of our code between two rounds of measurements, we could now directly apply theorem 7, setting $\mathscr{G} = \mathscr{S}$. However, in a more realistic setting we want to preserve the information between rounds and thus our measurements will act on the accumulated data error in each round and not just on the new error. Similar to the previous section, we can remedy this by considering the syndrome relative to the previous one. However, this will effectively propagate measurement errors between rounds. If $a_d$ is the accumulated data error in a given round, $a_m$ the measurement errors in that round, and $e = (e_d, e_m)$ is the new error occurring in the next round, then the product of outcomes for the measurement $s \in \mathscr{M}$ is given by

$$\langle s, (a_d, a_m)\rangle \langle s, (a_d, 0)(e_d, e_m)\rangle = \langle s, (e_d, e_m a_m)\rangle . \tag{68}$$

Thus, effectively we measure the new data error $e_d$ and the combined measurement error $a_m e_m$ from both rounds. Since the measurement errors are assumed to be independent between rounds, the distribution $\tilde{P}$ of $(e_d, a_m e_m)$ factorizes in the same way as $P$, but the strength of measurement errors is increased. Here, it is important that we divide the measurements into disjoint pairs of consecutive rounds such that the measurement errors are also independent between each pair. By theorem 7, as long as the original noise is correctable, we can then estimate the adjusted distribution from the syndrome measurements, up to logical equivalence.

Depending on how exactly error correction is performed, $\tilde{P}$ might be the most relevant distribution. If we are instead interested in the original error distribution $P$, we can also obtain this by post-processing as follows. Denote as $P_m$ the marginal distribution of $P$ on the measurement errors, i.e.

$$P_m(e_m) = \sum_{e_d \in \mathsf{P}^n} P(e_d, e_m) . \tag{69}$$

We can write this as $P_m = (P * \Phi_{\mathsf{P}^n}) \cdot \Phi_{\mathbb{F}_2^m}$, where we view $\mathsf{P}^n$ and $\mathbb{F}_2^m$ as subgroups of $\mathsf{G}^{n,m}$. Since $(\mathsf{P}^n)^\perp = \mathbb{F}_2^m$, we have by lemma 5,

$$\begin{aligned} E_m &\coloneqq \mathscr{F}[P_m] = \mathscr{F}[P * \Phi_{\mathsf{P}^n}] * \mathscr{F}[\Phi_{\mathbb{F}_2^m}] \\ &= \left(E \cdot |\mathsf{G}^{n,m}| U_{\mathbb{F}_2^m}\right) * |\mathsf{G}^{n,m}| U_{\mathsf{P}^n} \\ &= \frac{|\mathsf{G}^{n,m}|^2}{|\mathsf{P}^n||\mathbb{F}_2^m|} (E \cdot \Phi_{\mathbb{F}_2^m}) * \Phi_{\mathsf{P}^n} = |\mathsf{G}^{n,m}| (E \cdot \Phi_{\mathbb{F}_2^m}) * \Phi_{\mathsf{P}^n} . \end{aligned}$$

Explicitly, this means that $E_m(e_d, e_m) = |\mathsf{G}^{n,m}|\, E(0, e_m)$. Since the measurement errors are independent between rounds, the adjusted distribution is given by $\tilde{P} = P * P_m$. Thus, the moments of the adjusted distribution are $\tilde{E} \coloneqq \mathscr{F}[\tilde{P}] = E \cdot E_m$. We can obtain these up to logical equivalence, i.e. we obtain $\tilde{E} \cdot \Phi_{\mathscr{L}}$, and are interested in the moments $E \cdot \Phi_{\mathscr{L}}$ of the original logical channel. Since $\mathscr{G} = \mathscr{S} \subseteq \mathsf{P}^n$, we have $\mathbb{F}_2^m \subseteq \mathscr{L} = \mathscr{G}^\perp$. Thus, in particular we have access to the moment $\tilde{E}(0, e_m)$ for any $e_m \in \mathbb{F}_2^m$, and by the above discussion we have $\tilde{E}(0, e_m) = E(0, e_m) E_m(0, e_m) = \frac{1}{|\mathsf{G}^{n,m}|} E_m(0, e_m)^2$. Thus we can obtain the original moment for each $l = (l_d, l_m) \in \mathscr{L}$ from the adjusted moments as follows:

$$E(l_d, l_m) = \frac{\tilde{E}(l_d, l_m)}{E_m(l_d, l_m)} = \frac{1}{\sqrt{|\mathsf{G}^{n,m}|}} \frac{\tilde{E}(l_d, l_m)}{\sqrt{\tilde{E}(0, l_m)}} . \tag{70}$$

All in all we obtain the following corollary to theorem 7.

**Corollary 18.** *Phenomenological data and measurement noise with error rates smaller than $\frac{1}{2}$ can be estimated from the measurements of a quantum data-syndrome code up to logical equivalence if the noise is not correlated over more than half the distance of the code.*

[1] M. Kliesch and I. Roth, *Theory of quantum system certification*, PRX Quantum **2**, 010201 (2021), arXiv:2010.05925 [quant-ph].

[2] J. Eisert, D. Hangleiter, N. Walk, I. Roth, D. Markham, R. Parekh, U. Chabaud, and E. Kashefi, *Quantum certification and benchmarking*, Nat. Rev. Phys. **2**, 382 (2020), arXiv:1910.06343 [quant-ph].

[3] M. Kliesch, R. Kueng, J. Eisert, and D. Gross, *Guaranteed recovery of quantum processes from few measurements*, Quantum **3**, 171 (2019), arXiv:1701.03135 [quant-ph].

[4] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, *Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography*, Nat. Commun. **8**, 14485 (2017), arXiv:1605.07674 [quant-ph].

[5] E. Nielsen, R. Blume-Kohout, L. Saldyt, J. Gross, T. L. Scholten, K. Rudinger, T. Proctor, J. K. Gamble, and A. Russo, *pygstio/pygsti: Version 0.9.9.3* (2020).

[6] R. Brieger, I. Roth, and M. Kliesch, *Compressive gate set tomography*, arXiv:2112.05176 [quant-ph].

[7] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, *Randomized benchmarking of quantum gates*, Phys. Rev. A **77**, 012307 (2008), arXiv:0707.0963 [quant-ph].

[8] E. Magesan, J. M. Gambetta, and J. Emerson, *Characterizing quantum gates via randomized benchmarking*, Phys. Rev. A **85**, 042311 (2012), arXiv:1109.6887.

[9] M. Sarovar, T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout, *Detecting crosstalk errors in quantum information processors*, Quantum **4**, 321 (2020), arXiv:1908.09855v3 [quant-ph].

[10] D. C. McKay, A. W. Cross, C. J. Wood, and J. M. Gambetta, *Correlated randomized benchmarking*, arXiv:2003.02354 [quant-ph].

[11] R. Harper, S. T. Flammia, and J. J. Wallman, *Efficient learning of quantum noise*, Nat. Phys. **16**, 1184 (2020), arXiv:1907.13022 [quant-ph].

[12] S. T. Flammia and J. J. Wallman, *Efficient estimation of Pauli channels*, ACM Transactions on Quantum Computing **1**, 1 (2020), arXiv:1907.12976 [quant-ph].

[13] S. T. Flammia and R. O'Donnell, *Pauli error estimation via Population Recovery*, Quantum **5**, 549 (2021), arXiv:2105.02885 [quant-ph].

[14] R. Harper, W. Yu, and S. T. Flammia, *Fast estimation of sparse quantum noise*, PRX Quantum **2**, 010322 (2021), arXiv:2007.07901 [quant-ph].

[15] R. Harper, S. T. Flammia, and J. J. Wallman, *Efficient learning of quantum noise*, Nat. Phys. **16**, 1184 (2020), arXiv:1907.13022 [quant-ph].

[16] J. J. Wallman and J. Emerson, *Noise tailoring for scalable quantum computation via randomized compiling*, Phys. Rev. A **94**, 052325 (2016), arXiv:1512.01098 [quant-ph].

[17] M. Ware, G. Ribeill, D. Ristè, C. A. Ryan, B. Johnson, and M. P. da Silva, *Experimental Pauli-frame randomization on a superconducting qubit*, Phys. Rev. A **103**, 042604 (2021), arXiv:1803.01818 [quant-ph].

[18] Y. Fujiwara, *Instantaneous quantum channel estimation during quantum information processing*, arXiv:1405.6267 [quant-ph] (2014).

[19] A. G. Fowler, D. Sank, J. Kelly, R. Barends, and J. M. Martinis, *Scalable extraction of error models from the output of error detection circuits*, arXiv:1405.1454 [quant-ph] (2014).

[20] M.-X. Huo and Y. Li, *Learning time-dependent noise to reduce logical errors: real time error rate estimation in quantum error correction*, New J. Phys. **19**, 123032 (2017), arXiv:1710.03636 [quant-ph].

[21] J. R. Wootton, *Benchmarking near-term devices with quantum error correction*, Quantum Science and Technology **5**, 044004 (2020), arXiv:2004.11037 [quant-ph].

[22] J. Florjanczyk and T. A. Brun, *In-situ adaptive encoding for asymmetric quantum error correcting codes*, arXiv:1612.05823 [quant-ph] (2016).

[23] S. T. Spitz, B. Tarasinski, C. W. J. Beenakker, and T. E. O'Brien, *Adaptive weight estimator for quantum error correction in a time-dependent environment*, Advanced Quantum Technologies **1**, 1870015 (2018), arXiv:1712.02360 [quant-ph].

[24] J. Combes, C. Ferrie, C. Cesare, M. Tiersch, G. J. Milburn, H. J. Briegel, and C. M. Caves, *In-situ characterization of quantum devices with error correction*, arXiv:1405.5656 [quant-ph] (2014).

[25] T. Wagner, H. Kampermann, D. Bruß, and M. Kliesch, *Optimal noise estimation from syndrome statistics of quantum codes*, Phys. Rev. Research **3**, 013292 (2021), arXiv:2010.02243 [quant-ph].

[26] T. Wagner, H. Kampermann, D. Bruß, and M. Kliesch, *Pauli channels can be estimated from syndrome measurements in quantum error correction*, Quantum **(accepted)**, arXiv:2107.14252 [quant-ph].

[27] E. H. Chen, T. J. Yoder, Y. Kim, N. Sundaresan, S. Srinivasan, M. Li, A. D. Córcoles, A. W. Cross, and M. Takita, *Calibrated decoders for experimental quantum error correction*, Physical Review Letters **128**, 10.1103/physrevlett.128.110504 (2022), arXiv:2110.04285 [quant-ph].

[28] Z. Chen, K. J. Satzinger, J. Atalaya, A. N. Korotkov, A. Dunsworth, D. Sank, C. Quintana, M. McEwen, R. Barends, P. V. Klimov, S. Hong, C. Jones, A. Petukhov, D. Kafri, S. Demura, B. Burkett, C. Gidney, A. G. Fowler, A. Paler, H. Putterman, I. Aleiner, F. Arute, K. Arya, R. Babbush, J. C. Bardin, A. Bengtsson, A. Bourassa, M. Broughton, B. B. Buckley, D. A. Buell, N. Bushnell, B. Chiaro, R. Collins, W. Courtney, A. R. Derk, D. Eppens, C. Erickson, E. Farhi, B. Foxen, M. Giustina, A. Greene, J. A. Gross, M. P. Harrigan, S. D. Harrington, J. Hilton, A. Ho, T. Huang, W. J. Huggins, L. B. Ioffe, S. V. Isakov, E. Jeffrey, Z. Jiang, K. Kechedzhi, S. Kim, A. Kitaev, F. Kostritsa, D. Landhuis, P. Laptev, E. Lucero, O. Martin, J. R. McClean, T. McCourt, X. Mi, K. C. Miao, M. Mohseni, S. Montazeri, W. Mruczkiewicz, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Newman, M. Y. Niu, T. E. O'Brien, A. Opremcak, E. Ostby, B. Pató, N. Redd, P. Roushan, N. C. Rubin, V. Shvarts, D. Strain, M. Szalay, M. D. Trevithick, B. Villalonga, T. White, Z. J. Yao, P. Yeh, J. Yoo, A. Zalcman, H. Neven, S. Boixo, V. Smelyanskiy, Y. Chen, A. Megrant, J. Kelly, and G. Q. AI, *Exponential suppression of bit or phase errors with cyclic error correction*, Nature **595**, 383 (2021), arXiv:2102.06132 [quant-

ph].
[29] G. Kalachev and S. Sadov, *A linear-algebraic and lattice-theoretical look at the cleaning lemma of quantum coding theory*, Linear Algebra and its Applications **649**, 96 (2022), arXiv:2204.04699 [quant-ph].
[30] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, 10th ed. (Cambridge University Press, USA, 2011).
[31] D. A. Lidar and T. A. Brun, eds., *Quantum Error Correction* (Cambridge University Press, 2013).
[32] Y. Mao and F. Kschischang, *On factor graphs and the Fourier transform*, IEEE Trans. Inf. Th. **51**, 1635 (2005).
[33] M. Aigner, *A Course in Enumeration*, Vol. 238 (Springer-Verlag Berlin Heidelberg, 2007).
[34] S. Bravyi and B. Terhal, *A no-go theorem for a two-dimensional self-correcting quantum memory based on stabilizer codes*, New Journal of Physics **11**, 043029 (2009), arXiv:0910.1983 [quantph].
[35] A. Ashikhmin, C.-Y. Lai, and T. A. Brun, *Quantum data-syndrome codes*, IEEE Journal on Selected Areas in Communications **38**, 449 (2020), arXiv:1907.01393 [quant-ph].
[36] D. Poulin, *Stabilizer formalism for operator quantum error correction*, Phys. Rev. Lett. **95**, 230504 (2005), arXiv:quant-ph/0508131 [quant-ph].
[37] D. S. Wang, A. G. Fowler, and L. C. L. Hollenberg, *Surface code quantum computing with error rates over 1%*, Phys. Rev. A **83**, 020302 (2011).
[38] N. Delfosse, B. W. Reichardt, and K. M. Svore, *Beyond single-shot fault-tolerant quantum error correction*, IEEE Transactions on Information Theory **68**, 287 (2022), arXiv:2002.05180 [quant-ph].
[39] D. Bacon, S. T. Flammia, A. W. Harrow, and J. Shi, *Sparse quantum codes from quantum circuits*, in *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '15 (Association for Computing Machinery, New York, NY, USA, 2015) p. 327–334, arXiv:1411.3334 [quant-ph].
[40] L. P. Pryadko, *On maximum-likelihood decoding with circuit-level errors*, Quantum **4**, 304 (2020), arXiv:1909.06732 [quant-ph].
[41] C. T. Chubb and S. T. Flammia, *Statistical mechanical models for quantum codes with correlated noise*, Ann. Inst. Henri Poincaré Comb. Phys. Interact. **8**, 296 (2021), arXiv:1809.10704 [quant-ph].
[42] W. Fulton and J. Harris, *Representation Theory: A First Course*, Graduate Texts in Mathematics (Springer New York, NY, 2013).
[43] S. Roman, *Field Theory* (Springer New York, NY, 2006).
[44] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques* (The MIT Press, 2009).
[45] P. Abbeel, D. Koller, and A. Y. Ng, *Learning Factor Graphs in Polynomial Time & Sample Complexity*, arXiv:1207.1366 [cs.LG] (2012).
[46] C. Vuillot, L. Lao, B. Criger, C. G. Almudéver, K. Bertels, and B. M. Terhal, *Code deformation and lattice surgery are gauge fixing*, New Journal of Physics **21**, 033028 (2019), arXiv:1810.10037 [quant-ph].
[47] Y. Fujiwara, *Ability of stabilizer quantum error correction to protect itself from its own imperfection*, Phys. Rev. A **90**, 062304 (2014), arXiv:1409.2559 [quant-ph].
[48] R. Chao and B. W. Reichardt, *Quantum error correction with only two extra qubits*, Physical Review Letters **121**, 10.1103/physrevlett.121.050502 (2018), arXiv:1705.02329 [quant-ph].

# E

# Quantum Grid States and Hybrid Graphs

| | |
|---|---|
| Title: | Quantum Grid States and Hybrid Graphs |
| Authors: | Biswash Ghimire, Thomas Wagner, Hermann Kampermann and Dagmar Bruß |
| Journal: | Physical Review A |
| Publication status: | Submitted |
| Contribution by TW: | Co-author (input approx. 10%) |

This publication corresponds to reference [Ghi+22]. In this work, new ways to construct quantum states from the Laplacians of several kinds of graphs are presented. The entanglement properties of these states are analyzed, and different entanglement criteria are presented. In particular, new constructions of bound entangled states are derived.

The main results of this work and the proofs were developed by BG. I checked all the proofs and suggested some simplifications and improvements to their structure. BG wrote the initial draft of the manuscript. I then helped proofread the manuscript and suggested some corrections and improvements.

# Quantum Grid States and Hybrid Graphs

Biswash Ghimire,[1*] Thomas Wagner,[1] Hermann Kampermann,[1] Dagmar Bruß[1]
[1] Institute for Theoretical Physics III, Heinrich-Heine-Universität Düsseldorf, D-40225 Düsseldorf, Germany
*Correspondence: biswash.ghimire@hhu.de

(Dated: July 21, 2022)

Using the signed Laplacian matrix, and weighted and hybrid graphs, we present additional ways to interpret graphs as grid states. Hybrid graphs offer the most general interpretation. Existing graphical methods that characterize entanglement properties of grid states are adapted to these interpretations. These additional classes of grid states are shown to exhibit rich entanglement properties, including bound entanglement. Further, we introduce graphical techniques to construct bound entangled states in a modular fashion. We also extend the grid states model to hypergraphs. Our work, on one hand, opens up possibilities for constructing additional families of mixed quantum states in the grid state model. On the other hand, it can serve as an instrument for investigating entanglement problems from a graph theory perspective.

## I. INTRODUCTION

The realization that quantum entanglement can be used as a resource [1] has garnered intense interest in the study and characterization of entanglement. A fundamental problem is to determine whether a given quantum state is entangled or separable – called the separability problem [2]. It has been proven that determining whether an arbitrary quantum system is separable is an NP-hard problem [3, 4]. However, it can still be worthwhile to explore the problem in the context of some particular family of quantum states instead of general states. In this paper, we focus on several families of quantum states that can be represented as combinatorial graphs, and determine entanglement properties of such states via graph theoretic methods.

Interest in interpreting so-called graph Laplacians as density matrices can be traced back to the work of Braunstein et at. [5], where it was shown that the normalized signed Laplacian matrix of a graph can be interpreted as a density matrix. This idea was refined by Lockhart et. al. in [6, 7] by imposing a grid structure on graphs, called grid-labelled graphs. We expand on this concept and provide additional interpretations of grid-labelled graphs as quantum states, using various Laplacian matrices.

We first summarize the concept and properties of quantum grid states. Grid states, introduced in [6], are mixed quantum states described by simple graphs called grid-labelled graphs. Note that these states are different from grid states in [8]. The vertices in a grid-labelled graph are arranged on a grid and are labelled with Cartesian indices $(i, j)$ row-wise from top-left to bottom-right. An edge $\{(i, j), (k, l)\}$ connecting vertices $(i, j)$ and $(k, l)$ is interpreted as the state $1/\sqrt{2}(|ij\rangle - |kl\rangle)$, called an edge state. For example, Fig. 1(a) shows the vertex labelling in a grid-labelled graph with the $|\phi^-\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle)$ Bell state. With this convention, the density matrix $\rho(G)$ of a grid state is defined as the equally weighted mixture of all projectors onto edge states in the corresponding grid-labelled graph $G$.

The (signed) Laplacian matrix of a grid-labelled graph, with a suitable normalization, is identical to its corresponding density matrix. In order to see this, remember that the signed Laplacian matrix $L$ of a graph on $n$ vertices is the $n \times n$ matrix defined as

$$L = D - A, \tag{1}$$

where $D$ is the degree matrix and $A$ the adjacency matrix [9]. The degree matrix $D$ is an $n \times n$ diagonal matrix, in which each diagonal entry $D_{\alpha\alpha}$, where $1 \leq \alpha \leq n$, indicates the number of edges connecting to vertex $v_\alpha$ – called the degree of vertex $v_\alpha$. The adjacency matrix $A$ is an $n \times n$ binary matrix such that if vertices $v_\alpha$ and $v_\beta$ are connected by an edge, the matrix entry $A_{\alpha\beta}$ is 1, otherwise it is 0 [9].

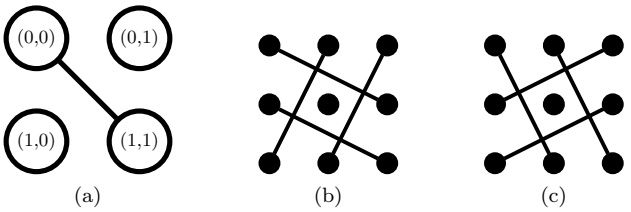
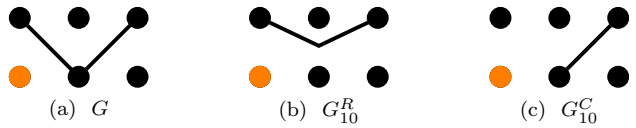



FIG. 1. (a) $L$-graph of the $|\phi^-\rangle$ Bell state. The pairs of integers indicate vertex indices. (b) A $3\times3$ cross-hatch graph, and (c) its partial transpose.

We call the grid-labelled graphs from [6] $L$-graphs. The degree criterion [5, 6] and the graph surgery procedure [6] characterize entanglement properties of grid states corresponding to $L$-graphs. The degree criterion is a graphical method that can be used to verify if the density matrix of an $L$-graph is positive under partial transpose. It makes use of the concept of partial transpose of a graph. The partial transpose of an $L$-graph $G$ is another $L$-graph $G^\Gamma$ such that an edge $\{(i, l), (k, j)\}$ exists in $G^\Gamma$ if and only if the edge $\{(i, j), (k, l)\}$ exists in $G$.

**Theorem 1** (Degree Criterion for $L$-graphs from [5, 6])**.** *The density matrix $\rho(G)$ of an $L$-graph $G$ is positive under partial transpose if and only if $D(G) = D(G^\Gamma)$.*

For example, the cross-hatch graph from [6], shown in Fig. 1(b), satisfies $D(G) = D(G^\Gamma)$. The corresponding



(a) $G$ (b) $G^R_{10}$ (c) $G^C_{10}$

FIG. 2. (a) An $L$-graph $G$. (b) and (c) Graphs $G^R_{10}$ and $G^C_{10}$ produced respectively by row and column surgeries on $G$ with vertex $(1,0)$ in orange as the selected isolated vertex. For row surgery, all edges connected to vertex $(1,1)$ are removed in the CUT step. As a result, the vertices $(0,0)$ and $(0,2)$ get disconnected, and then get reconnected in the STITCH step, which produces graph $G^R_{10}$ in (b). Likewise, for column surgery, all edges connected to vertex $(0,0)$ are removed. This does not disconnect any path between vertices not in column 0. The STITCH step is therefore not necessary. The graph $G^C_{10}$ in (c) is the result.

density matrix is therefore positive under partial transpose.

The graph surgery procedure [6] is a graphical method that allows to verify entanglement using the range criterion [10]. We restate the corollary of the range criterion from [6] as it also is the basis for graph surgery procedures presented in this paper.

**Corollary 2** ([6])**.** *If a rank $r$ density matrix has less than $r$ product vectors in its range, then it is entangled.*

Graph surgery involves performing a sequence of row and column surgeries on an $L$-graph. Row surgery is carried out by first selecting an isolated vertex, say $(i,j)$, in the $L$-graph and performing the "CUT" step, in which all edges connected to vertices in row $i$ are removed. This is followed by the "STITCH" step, which reconnects the path between every pair of vertices not in row $i$, if the "CUT" step severed the path [6]. In column surgery, the "CUT" and the "STITCH" steps are performed on the vertices in column $j$. The graph produced by a row / column surgery is denoted as $G^R_{ij}$ / $G^C_{ij}$, where the superscript indicates the type of surgery – $R$ for row surgery and $C$ for column surgery, and the subscript $ij$ denotes the isolated vertex chosen for the surgery. In effect, row / column surgery produces a simpler graph with fewer edges, unless vertices in the target row / column are all isolated vertices. Figure 2 shows an example of a row and a column surgery on an $L$-graph.

It was shown in [6] that any product vector in the range of the density matrix $\rho(G)$ of an $L$-graph $G$ – and thereby in the range of $L(G)$ – must also be in the range of either $L(G^R_{ij})$ or of $L(G^C_{ij})$. Since $G^R_{ij}$ and $G^C_{ij}$ are $L$-graphs, further row / column surgeries can be performed on them, and on the resulting graphs, and so on. Therefore, if iterated graph surgery on an $L$-graph $G$ always leads to the empty graph $G_E$, then any product vector in the range of $L(G)$ must also be in the range of $L(G_E)$, which is the zero matrix. This is clearly a contradiction, which means there are no product vectors in the range of $L(G)$. And, the corresponding density matrix is entangled according to Corollary 2.

The degree criterion and the graph surgery procedure connect entanglement properties of grid states to structural properties of $L$-graphs. Together, they enable the construction of bound entangled grid states. Furthermore, genuine multipartite entanglement is found in higher-dimensional grid states [6]. Such rich entanglement properties raise further questions. Are there other ways to interpret grid-labelled graphs as quantum states? Would the states also exhibit entanglement properties such as bound entanglement? Can the degree criterion and the graph surgery procedure be extended to such new interpretations of grid-labelled graphs? In this paper, we investigate these questions using additional types of Laplacian matrices. Specifically, the notion of grid-states is extended using the signless Laplacian matrix, and the weighted signed and signless Laplacian matrices. These interpretations lead us to conceive hybrid graphs, which represent density matrices that are mixtures of edge states corresponding to the signed and the signless Laplacian matrices. For these new states, we derive the analogous degree criteria and graph surgery procedures, and use them to construct bound entangled states. As a proof of concept, we also show that a degree criterion can be derived for grid-labelled hypergraphs.

We largely follow the nomenclature from [6]. For clarity, we occasionally prefix certain terms with the letter symbols of corresponding Laplacian matrices. For example, we call the grid-labelled graphs from [6] $L$-graphs. Further, we make no distinction between Laplacian matrices and density matrices when normalization is irrelevant. Similarly, since only bipartite quantum systems are considered in this paper, the partial transpose of a matrix $M$ is denoted by $M^\Gamma$ without loss of generality, as it is only used in relation to the Peres-Horodecki (also PPT) criterion [11], which does not depend on the transposed subsystem. We write a graph as $G = (V, E)$, where $V$ and $E$ are the vertex and the edge sets. Throughout this paper, we always assume that the density matrices are normalized. Additionally, depending on the context, we may use both a boldface letter or the braket notation for representing vectors. For example, for product vectors, the braket notation is the clearer notation.

With the following observation it is possible to check if the degree criterion can be adapted to a new interpretation of grid-labelled graphs.

**Observation 3.** *Let $G$ be a grid-labelled graph on $n$ vertices and $\rho(G)$ be the corresponding density matrix via any of the interpretations mentioned previously. If a vector $\mathbf{v}$ with all its components equal to $\pm 1$ (henceforth $\mathbf{v} \in \{-1,1\}^n$) exists in the kernel of $\rho(G^\Gamma)$, and if $\rho(G)$ is separable, then $D(G) = D(G^\Gamma)$.*

The observation is proven in Appendix A.

(a) $G_1$ (b) $G_2$

FIG. 3. (a) Graph $G_1$. (b) Graph $G_2$. Graphs $G_1$ and $G_2$ are partial transposes of each other.

## II. $Q$-GRID STATES

In this section, grid-labelled graphs are interpreted with the signless Laplacian matrix. The signless Laplacian of a graph $G$ is defined as $Q = D+A$, where $D$ and $A$ are the degree and the adjacency matrices of $G$. Normalized, the signless Laplacian is a proper density matrix. We call the quantum states described by the normalized signless Laplacian $Q$-grid states. The corresponding graphs are called $Q$-graphs. Graph features such as grid structure and vertex labelling are unchanged for $Q$-graphs, while the interpretation of edges $\{(i,j),(k,l)\}$ changes. A $Q$-edge state has the form $1/\sqrt{2}(|ij\rangle + |kl\rangle)$. The density matrix of a $Q$-grid state represented by a $Q$-graph $G = (V, E)$ is defined as

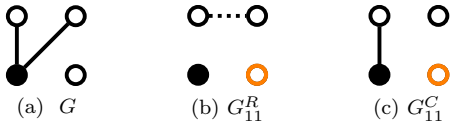
$$\rho_Q(G) = \frac{1}{|E|}\sum_{e\in E} |e\rangle\langle e| = \frac{1}{|E|}Q(G), \tag{2}$$

where $\{|e\rangle\}$ are the $Q$-edges states of edges in $E$. The notion of partial transpose of $L$-graphs in [6] is directly applicable to $Q$-graphs because it does not depend on the sign of the Laplacian matrix.

In the following, we adapt the degree criterion and the graph surgery procedure to $Q$-graphs. We use Observation 3 to identify $Q$-graphs for which the degree criterion is applicable. The observation requires that for a $Q$-graph $G$ on $n$ vertices the signless Laplacian $Q(G^\Gamma)$ of its partial transpose graph must have a vector $\mathbf{v} \in \{-1,1\}^n$ in its kernel. This is only fulfilled for bipartite graphs (see Lemma B.5). Therefore, we require this condition on the partial transpose of the graph. Remember that a graph is bipartite if its vertex set can be divided into two disjoint subsets such that no edge in the graph connects vertices in the same subset.

**Theorem 4** (Degree Criterion for $Q$-graphs)**.** *Let $G$ be a $Q$-graph. If $\rho_Q(G)$ is separable and $G^\Gamma$ is bipartite, then $D(G) = D(G^\Gamma)$.*

The proof of Theorem 4 is found in Appendix B. The degree criterion for $Q$-graphs, like its counterpart for $L$-graphs, is necessary and sufficient for $2\times 2$ and $2\times 3$ systems, due to the PPT criterion. The bipartite condition for the graph transpose in the degree criterion for $Q$-graphs has an important implication. There exist grid-labelled graphs that, if interpreted as $Q$-graphs, are separable, but are entangled if interpreted as $L$-graphs. For example, the graphs $G_1$ and $G_2$ in Fig. 3, if treated as $L$-graphs, represent entangled states because $D(G_1) \neq D(G_2)$. If instead both are treated as $Q$-graphs, $G_1$ still represents an entangled state because $G_2$ is bipartite and $D(G_1) \neq D(G_2)$. On the other hand, the degree criterion is not applicable to $G_2$ because its partial transpose $G_1$ is not bipartite. It is easily verified that the density matrix $\rho_Q(G_2)$ is separable.



(a) $G$ (b) $G^R_{11}$ (c) $G^C_{11}$

FIG. 4. Vertices are colored in black and white to show that the graphs are bipartite. Then vertex chosen for graph surgery is indicated in orange. Solid and dashed edges indicate $Q$- and $L$-edges, respectively. (a) $Q$-graph $G$. (b) Graph $G^R_{11}$. The CUT step splits the connected component with vertices $(0,0)$, $(0,1)$, and $(1,0)$. Since vertices $(0,0)$ and $(0,1)$ are in the same partition, it is not possible to reconnect them with a $Q$-edge. So an $L$-edge is used. (c) Graph $G^C_{11}$.

We now extend the graph surgery procedure to $Q$-graphs. We call graph surgery on $Q$-graphs $Q$-surgery. To understand $Q$-surgery, we need the concept of connected components. A connected component of a graph is a subgraph that has a path between any two of its vertices, and no paths between any of its vertices and the remaining vertices of the original graph. An isolated vertex trivially satisfies the definition and is considered a connected component. For example, the graph $G_1$ in Fig. 3 has two connected components and the graph $G_2$ has one. Like $L$-surgery, $Q$-surgery is a sequence of row and/or column surgeries. For simplicity, $Q$-surgery is only defined for bipartite $Q$-graphs. Row surgery is performed as follows:

- CUT: Select an isolated vertex $(i,j)$ and remove all edges attached to vertices in row $i$.
- STITCH: If the CUT step splits any connected component and vertices in the split constituents, excluding the ones in row $i$, all belong to the same partition, reconnect the constituents with $L$-edge(s). Otherwise, reconnect the constituents with $Q$-edge(s).

Note that in the STICH step, if $Q$-edge(s) are used for reconnection, each $Q$-edge must connect vertices in opposing partitions.

Likewise, column surgery is performed on vertices in column $j$. The graph resulting from a row/column surgery on vertex $(i,j)$ is denoted as $G^R_{ij}$ / $G^C_{ij}$.

An iteration of row/column surgery on an $L$-graph always produces an $L$-graph. In contrast, the analogous case is not necessarily true for $Q$-graphs. Suppose a connected component of a $Q$-graph split in the CUT step is reconnected in the STITCH step with $Q$-edge(s), while another split connected component is reconnected with $L$-edge(s). The resulting graph is then not a $Q$-graph because it has both $L$- and $Q$-edges in it. Nonetheless,

it still holds for $Q$-graphs that any product vector in the range of the density matrix of the original $Q$-graph must be in the range of the density matrix of the graph produced after an iteration of a row and column surgery. This is formalized in the following observation.

**Observation 5.** *Let $G$ be a bipartite $Q$-graph on $n$ vertices with an isolated vertex $(i, j)$. If a product vector $|\mu\,\nu\rangle \in R[\rho_Q(G)]$, where $R$ denotes the range, then*

- $|\mu\,\nu\rangle \in R[\rho_Q(G^R_{ij})]$ *or* $R\big[\rho_L(G^R_{ij})\big]$, *or*
- $|\mu\,\mu\rangle \in R\left[\rho_Q(G^C_{ij})\right]$ *or* $R\left[\rho_L(G^C_{ij})\right]$, *or*
- $|\mu\,\nu\rangle \in R\left[\rho(G')\right]$,

*where $G'$ is a hybrid graph (see Section IV).*

The proof of Observation 5 is found in Appendix B. An example of row and column surgeries on a $Q$-graph is shown in Fig. 4. With Observation 5, $Q$-surgery, like $L$-surgery, can be used in connection with Corollary 2. Therefore, if $Q$-surgery on a $Q$-graph always produces the empty graph, the associated density matrix is entangled.

In general, the $Q$- and $L$-grid states of the same grid-labelled graph are not unitarily equivalent. In the following observation, we identify a condition when that is the case.

**Observation 6.** *Let $G$ be a grid-labelled graph. If $G$ is not bipartite, then $\rho_L(G)$ and $\rho_Q(G)$ are not unitarily equivalent.*

A proof of Observation 6 is given in Appendix B.

## III. GRID STATES CORRESPONDING TO WEIGHTED GRAPHS

Weighted graphs generalize the notion of edges in graphs and allow non-zero, positive weights to be associated with each edge in the graph [12]. In this section, the weighted signed and signless Laplacian matrices are interpreted as quantum states that correspond to the respective weighted $L$- and $Q$-graphs.

Edge states in a weighted $L$- or a $Q$-graph have the same form as in their unweighted counterparts. However, the density matrix is defined as

$$\rho(G_w) = \frac{1}{\sum_e w_e} \sum_{e\in E} w_e\, |e\rangle\langle e| \tag{3}$$

where $G_w$ is a weighted grid-labelled graph, $\{|e\rangle\}$ are the edge states of edges in $G_w$, and $\{w_e\}$ the respective non-zero, positive edge weights. If the edges denote $L$-edge states ($Q$-edge states), the density matrix is the normalized signed (signless) Laplacian of the weighted graph. The signed and the signless Laplacian matrices of weighted graphs are defined as $L = D - A$ and $Q = D + A$, respectively. The degree of a vertex in a weighted graph is the sum of edge weights of all edges that connect to it, and the degree matrix $D$ is a diagonal matrix with degrees of vertices as its diagonal entries. Likewise, the adjacency matrix $A$ also accounts for edge weights. The matrix entry $A_{\alpha\beta}$ is $w_{\alpha\beta}$ if vertices $v_\alpha$ and $v_\beta$ are connected by an edge weighted $w_{\alpha\beta}$, otherwise it is 0 [12]. Notice that in an unweighted graph all edge weights are implicitly 1.

The edges in the partial transpose graph $G^\Gamma$ of a weighted grid-labelled graph $G$ carry the weights of the corresponding edges in $G$. The degree criteria and the graph surgery procedures on unweighted $L$- and $Q$-graphs directly apply to weighted graphs. Lemma 7 justifies this claim.

**Lemma 7.** *If the vertex and the edge sets of two weighted $L$-graphs (resp. $Q$-graphs) are identical, their signed (resp. signless) Laplacians have identical kernels.*

The proof of Lemma 7 is found in Appendix C. With Lemma 7 and Observation 3, the degree criteria for unweighted $L$- and $Q$-graphs are also valid for weighted $L$- and $Q$-graphs. Likewise, $L$- and $Q$ surgeries also directly apply to weighted graphs. Since Laplacian matrices are hermitian, Lemma 7 implies that Laplacians of weighted graphs with identical vertex and edge sets have identical ranges. This means if graph surgery on an unweighted $L$- or $Q$-graph always yields the empty graph, it must be that graph surgery on any other weighted graph with the same vertex and edge sets must also always yield the empty graph. Therefore, edge weights are irrelevant for graph surgery and the graph surgery procedures for unweighted $L$- and $Q$-graphs can be used on weighted $L$- and $Q$-graphs. Edge weights alone also do not determine if the density matrix corresponding to a weighted $L$- or $Q$-graph is entangled or separable.

Moreover, Observation 6 can be applied to weighted $Q$-graphs as formalized in the following corollary.

**Corollary 8.** *Let $G_w$ be a weighted grid-labelled graph. If $G_w$ is not bipartite, then $\rho_L(G_w)$ and $\rho_Q(G_w)$ are not unitarily equivalent.*

The corollary is proved in Appendix C.

## IV. GRID STATES WITH HYBRID GRAPHS

In this section, we approach the idea of interpreting graphs as quantum states from a physical point of view. A density matrix that is a mixture of both $L$- and $Q$-edge states is not unphysical. Is it then possible to represent such density matrices using grid-labelled graphs? We answer this question in the affirmative by introducing the notion of hybrid graphs and describing analogous degree criteria and graph surgery procedures for them.

A hybrid graph contains both $L$- and $Q$-edges and is written as $G = (V, E_L + E_Q)$, where $V$ is the vertex set, and $E_L$ and $E_Q$ are the sets of $L$- and $Q$-edges, respectively. Its $L$- and $Q$-subgraphs are the graphs $S_l = (V, E_L)$ and $S_q = (V, E_Q)$. Hybrid graphs slightly resemble signed graphs [13], where each edge in a graph is given either a positive or a negative sign. However, we do not use the Laplacian matrix in [13] to derive the density matrix of hybrid graphs. Instead, we treat hybrid

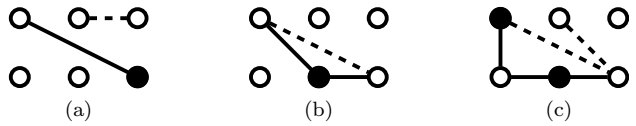



FIG. 5. Three types of hybrid graphs. Vertices are colored in black and white to show that the graphs are bipartite. (a) An NOI-graph. (b) A COI-graph. (c) A GI graph. Solid and dashed edges indicate $Q$- and $L$-edges, respectively.

graphs as compositions of $L$ and $Q$-graphs and define the hybrid Laplacian matrix as $\mathcal{L}(G) = L(S_l) + Q(S_q)$. The normalized hybrid Laplacian is a density matrix that is the equally weighted mixture of all $L$- and $Q$-edge states in the corresponding graph.

Coexistence of $L$- and $Q$-edges limit general results on entanglement properties, because Observation 3 imposes different conditions on $L$- and $Q$-graphs. Considering that, hybrid graphs are divided into three categories based on their edge-vertex characteristics:

- Non-Overlapping Incidence (NOI): A hybrid graph with NOI has a bipartite $Q$-subgraph and no vertex in it is connected by both a $Q$-edge and an $L$-edge.
- Conditionally Overlapping Incidence (COI): A hybrid graph with COI has a bipartite $Q$-subgraph and every $L$-edge in it connects vertices that are both in the same partition.
- General Incidence (GI): Hybrid graphs with GI have no restrictions on incidences of $L$- and $Q$-edges.

We call a hybrid graph with NOI a NOI-graph, and likewise for graphs with COI and GI. An example each of a NOI-, a COI-, and a GI-graph is given in Fig. 5. Note that a NOI-graph is a special case of a COI graph, because vertices connected by $L$-edges in a NOI-graph can all be put in one of the two vertex partitions.

As before, we adapt the degree criteria and graph surgery procedures to hybrid graphs. GI-graphs are too general for Observation 3 to be applicable. Therefore, only NOI- and COI-graphs are considered.

**Theorem 9** (Degree Criterion)**.** *If the density matrix $\rho(G)$ of a hybrid graph $G$ is separable and $G^\Gamma$ is a NOI- or a COI-graph, then $D(G) = D(G^\Gamma)$.*

A proof for Theorem 9 is provided in Appendix D.

Graph surgery on a NOI-graph involves both $L$- and $Q$-surgeries. Any connected component in a NOI-graph has either all $L$-edges or all $Q$-edges. One can thus perform $L$- and $Q$-surgery independently on the respective connected components.

Graph surgery on a COI-graph however is not as straightforward. The non-identical STITCH steps of $L$- and $Q$-surgery are equally valid for any vertex with simultaneous incidences of $L$- and $Q$-edges. This ambiguity is resolved by a proxy graph.

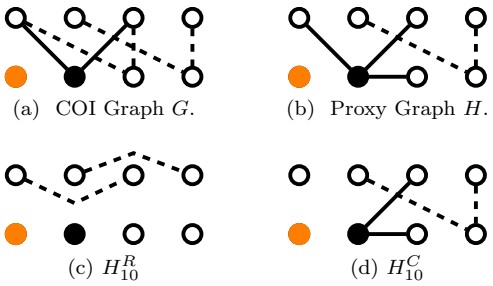



FIG. 6. Vertices are colored in black and white to show that the graphs are bipartite. Solid and dashed edges are $Q$- and $L$-edges, respectively. (a) Graph surgery on a COI-graph $G$ with vertex $(1, 0)$, colored orange, as the selected isolated vertex. (b) Graph $H$, a proxy graph of $G$, as described in Section IV. To derive $H$ from $G$, two $L$-edges $\{(0, 0), (1, 2)\}$ and $\{(0, 2), (1, 2)\}$ are removed and a $Q$-edge $\{(1, 1), (1, 2)\}$ is added. (c) Graph $H^R_{10}$. Vertices $(0, 0)$ and $(0, 2)$ cannot be connected by a $Q$-edge because they belong to the same partition. So an $L$-edge is used. (d) Graph $H^C_{10}$.

A *proxy graph* of a COI-graph is a NOI-graph such that the kernels of their hybrid Laplacians are identical. It is constructed with a two-step process: first, by removing $L$-edges from all vertices on which both $L$- and $Q$- edges are incident; then, by reconnecting split connected components, if any, using $Q$-edges only.

**Observation 10.** *Every COI-graph has a proxy graph.*

The proof of Observation 10 is found in Appendix D. Deriving a proxy graph is akin to graph sparsification, which removes edges from a dense graph while preserving certain spectral properties of the Laplacian of the original graph [14]. In the case of proxy graphs, only $L$-edges are removed and the preserved spectral property is the kernel of the hybrid Laplacian. Given Observation 10, graph surgeries on a proxy NOI-graph and on the original COI-graph are equivalent. Therefore, graph surgery on a COI-graph is performed by first constructing a proxy NOI-graph and performing graph surgery on it. One iteration each of row and column surgeries on a COI-graph are shown in Fig. 6.

The implication of graph surgery on hybrid graphs is the same as on $L$- and $Q$-graphs: If graph surgery on a hybrid graph always produces the empty graph, then the corresponding density matrix is entangled.

Hybrid graphs can also have weighted edges. As in the case of weighted $L$- and $Q$-graphs, the degree criteria and the graph surgery procedures on unweighted hybrid graphs also apply to weighted hybrid graphs, as justified by the following lemma.

**Lemma 11.** *If the vertex and the edge sets of two weighted hybrid graphs are identical, their hybrid Laplacians have identical kernels.*

The proof of Lemma 11 is found in Appendix D.

## V. CONSTRUCTION OF BOUND ENTANGLED STATES

In [6, 7], bound entangled $L$-grid sates are constructed using the degree criterion to verify a positive partial transpose of the density matrix, and the graph surgery procedure to verify entanglement. This method can be used to construct new families of bound entangled states with the grid states presented in this paper.

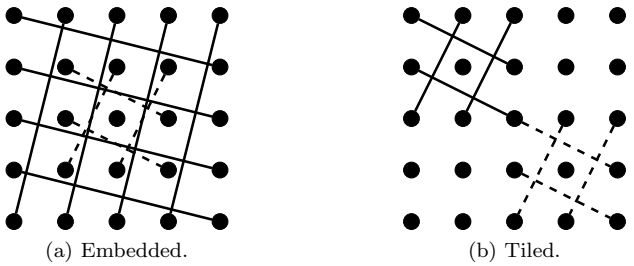
**Observation 12.** *If a grid-labelled graph $G$ satisfies $D(G) = D(G^\Gamma)$, the corresponding density matrix has a positive partial transpose, independent of whether the graph is interpreted as an $L$-graph, a $Q$-graph, a weighted graph, or a hybrid graph.*

A proof of Observation 12 is provided in Appendix E. According to the observation, the degree criterion verifies that a grid-state is positive under partial transpose, and graph surgery verifies that it is entangled. Given that, bound entangled $Q$-grid states can be constructed using the degree criterion and the graph surgery procedure defined in Section II if both the Q-graph and its partial transpose graph are bipartite. The cross-hatch pattern from [6] satisfies these conditions. The pattern is in fact applicable not only to $Q$-graphs, but also to weighted and hybrid graphs.

**Theorem 13.** *The density matrix of an $m \times n$ cross-hatch graph with $m, n \geq 3$ is bound entangled for all grid states independent of whether the graph is interpreted as an $L$-graph, a $Q$-graph, a weighted graph, or a hybrid graph.*

Theorem 13 is proved in Appendix E. Moreover, the cross-hatch pattern can be composed. For example, irrespective of the Laplacian matrix used to interpret the resulting graph, a smaller cross-hatch graph can be embedded inside a bigger one as shown in Fig. 7(a) to produce new bound entangled states. Likewise, the pattern can be tiled as shown in Fig. 7(b). Both graphs in Fig. 7 satisfy the degree criterion, because the constituent graphs in each graph individually satisfy the degree criterion. Therefore, they represent grid-states whose density matrices are positive under partial transpose.

Graph surgery on both graphs is carried out by first performing graph surgery on one of the constituent graphs and then on the remaining edges of the other one. In the tiled composition, the STITCH step adds a diagonal edge, which can be treated as a part of another cross-hatch graph and be removed. In addition, the embedded and tiled compositions like in Fig. 7 can also be composed to produce more bound entangled states, as long as the compositions satisfy the respective degree criterion and are reducible to empty graphs via graph surgery.



(a) Embedded. (b) Tiled.

FIG. 7. Examples of composing cross-hatch graphs. Solid and dashed edges distinguish constituent graphs. The states corresponding to both graphs are bound entangled, irrespective of their interpretation as weighted or unweighted $L$- or $Q$-graphs, or as hybrid graphs.

## VI. GRID STATES CORRESPONDING TO HYPERGRAPHS

With hybrid graphs, we showed that it is possible to generate density matrices from a mixture of $Q$- and $L$-edge states. By defining a suitable Laplacian matrix, we derived degree criteria and graph surgery procedures. As a proof of concept, we follow the same approach to extend the grid state model to hypergraphs.

Hypergraphs generalize graphs and allow edges to contain more than two vertices [15]. Here, we only consider hypergraphs in which all hyperedges contain exactly three vertices. In the literature, various approaches to extend graph matrices to hypergraphs are found, which range from matrices in [15–17] to tensors in [18]. None of these previous approaches leads to a density matrix that can be elegantly represented by a grid-labelled hypergraph. Therefore, we first extend the notion of edge states and define hyperedge states, from which we define the density matrix and the hypergraph Laplacian matrix. As such, the hyperedge state is chosen to be of the form $1/\sqrt{3}\,(|ij\rangle + |kl\rangle + |mn\rangle)$. The density matrix is the equal mixture of all hyperedge states in a hypergraph, and the Laplacian matrix is the unnormalized density matrix. Split into a diagonal and a non-diagonal matrix, the Laplacian of a hypergraph $H$ is written as

$$L(H) = D(H) + A(H), \tag{4}$$

where the diagonal matrix $D(H)$ and the non-diagonal matrix $A(H)$ matrix are the degree and adjacency matrices, respectively. The diagonal entries of the degree matrix are the degrees of vertices in the hypergraph. The degree of a vertex is the number of hyperedges incident on the vertex. The adjacency matrix is defined as

$$A_{\alpha\beta} = \begin{cases} \mathrm{adj}(v_\alpha, v_\beta), \text{ if } \alpha \neq \beta, \\ 0, \text{ otherwise,} \end{cases} \tag{5}$$

where $\mathrm{adj}(v_\alpha, v_\beta)$ is the number of hyperedges connecting vertices $v_\alpha$ and $v_\beta$.

### A. Weighted Graph Model for hypergraph

A hypergraph can be modeled with a weighted graph, and its Laplacian matrix can be connected to the signless Laplacian matrix of the weighted graph.

Consider a hypergraph $H$ with two hyperedges in Figs. 8(a, b). Each hyperedge is turned into a clique as in Figs. 8(c, d). A clique is a subset of vertices of a graph such that every vertex in the set is connected to every other vertex in the set [19]. The cliques are combined into a weighted graph as in Fig. 8(e) such that the edge weight of an edge connecting a vertex pair is the cumulative number of edges in all cliques that connect the vertex pair. In Fig. 8(e), the weights of black edges are all 1 and the orange edge is weighted 2. We call the weighted graph derived in this fashion the graph of a hypergraph. Formally, the graph of a hypergraph $H$ is a weighted graph $G$ such that any vertex pair $\{v_\alpha, v_\beta\}$ connected by a hyperedge in $H$ is connected in $G$ by an edge with weight $A(H)_{\alpha\beta}$.

With this construction, the adjacency matrix of a hypergraph and of its graph are the same matrix. But the degree matrices are different. Consider a hypergraph $H$ and its graph $G$. The degree of a non-isolated vertex $v_\alpha$ in $H$ is $D(H)_\alpha < \sum_\beta A(H)_{\alpha\beta}$. However, in the graph $G$ the degree of the same vertex by definition is $D(G)_\alpha = \sum_\beta A(G)_{\alpha\beta}$. The degree matrices of a hypergraph and its graph thus are offset by a diagonal non-negative matrix, which we call the offset matrix and define it as

$$O(H) = D(G) - D(H), \tag{6}$$

where $H$ is a hypergraph, $G$ its graph, and $O(H)$ the offset matrix. With these observations, the hypergraph Laplacian of a hypergraph $H$ can be written as

$$L(H) = Q(G) - O(H), \tag{7}$$

where $Q(\cdot)$ indicates the signless Laplacian.

With the weighted graph model, we can derive a degree criterion for hypergraph grid states.

**Theorem 14.** *Let $H$ be a hypergraph and $G$ be its graph. If $\rho(H)$ is separable and $G^\Gamma$ is bipartite, then $D(G) = D(G^\Gamma)$.*

For the proof of Theorem 14, see Appendix F. Unlike the degree criteria for grid-labelled graphs, it is not clear that the hypergraph degree criterion is sufficient for the positive partial transpose of the hypergraph density matrix. Suppose $H$ is a hypergraph and $G$ is its graph, and $D(G) = D(G^\Gamma)$. Then,

$$\begin{aligned} Q^\Gamma(G) &= D^\Gamma(G^\Gamma) + A^\Gamma(G) \\ &= D(G^\Gamma) + A(G^\Gamma) = Q(G^\Gamma), \end{aligned} \tag{8}$$

and from Equation (7)

$$\begin{aligned} &Q(G) = L(H) + O(H). \\ \implies &Q^\Gamma(G) = L^\Gamma(H) + O^\Gamma(H) = L^\Gamma(H) + O(H), \end{aligned} \tag{9}$$

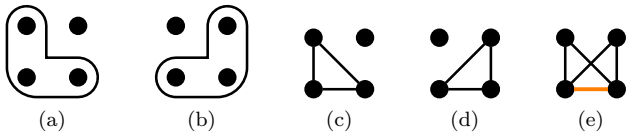



FIG. 8. Weighted graph model of a hypergraph. (a) and (b) Two hyperedges. (c) and (d) Their respective cliques. (e) Weighted graph derived from the cliques.

from which it follows

$$L^\Gamma(H) = Q(G^\Gamma) - O(H). \tag{10}$$

From Equation (10), it is not clear if $Q(G^\Gamma) - O(H)$ is always positive semidefinite. On the other hand, consider a $2 \times 2$ hypergraph $H$ with a single hyperedge shown in Fig. 8(a). The graph $G$ of the hypergraph is the graph in Fig. 8(c). It is easily seen that $D(G) \neq D(G^\Gamma)$, and also verified using the PPT criterion that $\rho(H)$ is entangled.

Graph surgery cannot be extended to hypergraphs via the weighted graph model. The graph surgery procedure for weighted $Q$-graphs requires the graphs to be bipartite. The graph of a hypergraph, although a weighted $Q$-graph, is not bipartite, because cliques are inherently not bipartite.

Even though this interpretation of hypergraph grid states does not allow graph surgery, it illustrates the flexibility of the grid-state model. We were not only able to define a hypergraph laplacian matrix in an ad-hoc manner to suit our purpose, but also integrate the weighted Laplacian to derive a degree criterion for hypergraph grid states.

## VII. CONCLUSION AND OUTLOOK

This paper reveals a rich interplay between graphs and quantum states. Using a variety of interpretations of graphs as density matrices, we have identified additional families of grid states beyond the ones originally suggested in [6] and shown that their entanglement properties relate to properties of the corresponding graphs. In particular, we investigated signless Laplacians and weighted graphs. We introduced the concept of hybrid graphs, containing two different types of edges, and derived the entanglement properties of the corresponding grid states. Additionally, we constructed new families of bound entangled states with these new grid states, using the method from [6]. We showed that the crosshatch pattern is not only bound entangled for the new families of grid states, but it could also be composed to construct more bound entangled states. We noted two additional links between graph theory and grid states – resemblance between hybrid graphs and signed graphs, and between proxy graph construction and graph sparsification. Further work into these links would be interesting. For example, one could investigate if proxy graphs

can be connected to the concept of local graph isomorphism discussed in [7].

We demonstrated with hypergraph grid states that our approach for hybrid graphs can be applied in other contexts. Similar approaches could be used to incorporate more general edge states, for example, with the normalized Laplacian defined in [20] and with complex Laplacian matrices.

Since genuine multipartite entanglement has been found in $L$-grid states [6], for further work, one could investigate if the same is the case for grid states presented above. Finally, as the graph surgery procedure is not possible without isolated vertices, it would be desirable to improve graph surgery or find alternative procedures that do not require isolated vertices.

## ACKNOWLEDGMENTS

We acknowledge financial support by the QuantERA grant QuICHE and the German ministry of education and research (BMBF, grant no. 16KIS1119K), and by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – Cluster of Excellence Matter and Light for Quantum Computing (ML4Q) EXC 2004/1 – 390534769.

## Appendix A: Additional Graph Theory Concepts

The appendices contain the proofs of results stated in the main text. The statements are repeated before each proof. In this section, in addition to the proof of Observation 3, we present graph concepts used in the proofs.

The *unoriented incidence matrix* of a graph $G = (V, E)$ is the $|V| \times |E|$ matrix $R$ such that $R_{ij} = \sqrt{w_j}$ if edge $e_j$ with weight $w_j$ is incident on vertex $v_i$, and $R_{ij} = 0$ otherwise. The *oriented incidence matrix* $B$ results from negating one of the two non-zero entries in each column of the matrix $R$ [12]. The signed and the signless Laplacian matrices satisfy $L = BB^T$ and $Q = RR^T$ [9, 21].

For the proof of Observation 3, the following lemma is needed. Hereafter, $K(\cdot)$ denotes the kernel of a matrix.

**Lemma A.1** ([5])**.** *Let $M$ and $\Delta$ be $n \times n$ real matrices. Let $M$ be symmetric and positive semidefinite, and $\Delta$ be nonzero, diagonal, and traceless. If a vector $\mathbf{v} \in \{-1, 1\}^n$ exists in $K(M)$, then $M + \Delta \not\geq 0$.*

*Proof of Lemma A.1 .* Given the nature of matrix $\Delta$, at least one of its diagonal entries, say $\Delta_{ii} = \delta$, is positive and nonzero. Let $\mathbf{v} \in \{-1, 1\}^n$ be in $K(M)$. Let $\mathbf{w} \coloneqq \mathbf{v} + a\mathbf{x}$, with $a \in \mathbb{R}$, and $\mathbf{x}$ be the $i$-th standard basis vector. Consider the inner product

$$\begin{aligned} I \coloneqq & \langle \mathbf{w}, (M + \Delta)\mathbf{w} \rangle \\ = & \langle \mathbf{v}, M\mathbf{v} \rangle + a\langle \mathbf{v}, M\mathbf{x} \rangle + a\langle \mathbf{x}, M\mathbf{v} \rangle + a^2 \langle \mathbf{x}, M\mathbf{x} \rangle \\ & + \langle \mathbf{v}, \Delta\mathbf{v} \rangle + a\langle \mathbf{v}, \Delta\mathbf{x} \rangle + a\langle \mathbf{x}, \Delta\mathbf{v} \rangle + a^2 \langle \mathbf{x}, \Delta\mathbf{x} \rangle. \end{aligned} \tag{A1}$$

The scalars $\langle \mathbf{v}, M\mathbf{v} \rangle$, $\langle \mathbf{v}, M\mathbf{x} \rangle$ and $\langle \mathbf{x}, M\mathbf{v} \rangle$ are all 0, because $M\mathbf{v} = 0$. And, $\langle \mathbf{x}, M\mathbf{x} \rangle = M_{ii}$ and $\langle \mathbf{x}, \Delta\mathbf{x} \rangle = \delta$. The remaining terms are

$$\langle \mathbf{v}, \Delta\mathbf{v} \rangle = \sum_{j=1}^{n} (\mathbf{v}_j)^2 \Delta_{jj} = \mathrm{tr}(\Delta) = 0 \tag{A2}$$

and

$$\langle \mathbf{v}, \Delta\mathbf{x} \rangle = \langle \mathbf{x}, \Delta\mathbf{v} \rangle = \pm\delta, \text{ if } \mathbf{v}_i = \pm 1. \tag{A3}$$

Equation (A1) thus reduces to

$$I = a^2(\delta + M_{ii}) \pm 2\delta a, \text{ if } \mathbf{v}_i = \pm 1. \tag{A4}$$

Notice that all diagonal entries of the matrix $M$ are non-negative, because $M$ is positive semidefinite. Equation (A4) therefore always has distinct roots, because $M_{ii} + \delta > 0$. This implies that there exists $a$ for which $I < 0$, meaning $M + \Delta \not\geq 0$. □

We now prove the observation.

**Observation 3.** *Let $G$ be a grid-labelled graph on $n$ vertices and $\rho(G)$ be the corresponding density matrix via any of the interpretations mentioned previously. If a vector $\mathbf{v}$ with all its components equal to $\pm 1$ (henceforth $\mathbf{v} \in \{-1, 1\}^n$) exists in the kernel of $\rho(G^\Gamma)$, and if $\rho(G)$ is separable, then $D(G) = D(G^\Gamma)$.*

*Proof of Observation 3.* Let $G$ be a grid-labelled graph on $n$ vertices, and $D(G)$ and $A(G)$ be the degree and the adjacency matrices of $G$, respectively. Let $\mathrm{L}(G) = D(G) \pm A(G)$ be a generic Laplacian matrix representative of the Laplacian matrices used in this paper. Let the corresponding density matrix $\rho(G)$ be the normalized $\mathrm{L}(G)$. Then

$$\mathrm{L}^\Gamma(G) = D^\Gamma(G) \pm A^\Gamma(G) = D(G) \pm A(G^\Gamma),$$

which implies

$$\begin{aligned} \mathrm{L}^\Gamma(G) &= D(G) + \mathrm{L}(G^\Gamma) - D(G^\Gamma) \\ &= \mathrm{L}(G^\Gamma) + \Delta, \end{aligned} \tag{A5}$$

where the matrix $\Delta = D(G) - D(G^\Gamma)$ is traceless and diagonal. If $\Delta$ is nonzero, then since $\mathrm{L}(G^\Gamma) \geq 0$, Lemma A.1 implies $\mathrm{L}(G^\Gamma) + \Delta \not\geq 0$. But $\rho(G)$ is separable and the PPT criterion requires $\rho^\Gamma(G) \geq 0$, meaning $\mathrm{L}^\Gamma(G) \geq 0$. This is a contradiction. Then, it must be that $\Delta = D(G) - D(G^\Gamma) = 0$. □

## Appendix B: $Q$-Grid States

Proof of results stated in Section II are given here. Several supporting observations are needed for the proof of Lemma B.5, which is then used to prove the degree criterion.

**Observation B.1** ([21])**.** *The least eigenvalue of the signless Laplacian of a connected graph is equal to 0 if and only if the graph is bipartite. In this case 0 is a simple eigenvalue.*

Next, we deduce a property of the kernel of the signless Laplacian matrix of connected bipartite graphs.

**Observation B.2.** *For any connected bipartite graph $G$ on $n$ vertices there exists a vector $\mathbf{v} \in \{-1, 1\}^n$ in $K[Q(G)]$.*

*Proof of Observation B.2.* Let the two vertex partitions in $G$ be $P_1$ and $P_2$. From Observation B.1, $Q(G)$ has a non-trivial kernel because $G$ is bipartite. Suppose a vector $\mathbf{v} \in \{-1, 1\}^n$ is constructed as follows: if the $k^{\text{th}}$ vertex is in $P_1$ then the component $v_k = 1$, otherwise $v_k = -1$. Given that the vertices connected by any edge in $G$ belong to opposite partitions, from the definition of the incidence matrix $R$, we see that $R(G)^T\mathbf{v} = \mathbf{0}$. Then $Q(G)\mathbf{v} = R(G)R(G)^T\mathbf{v} = \mathbf{0}$. □

Finally, with another result from [21], we derive a corollary to prove Lemma B.5.

**Observation B.3** ([21])**.** *In any graph, the (algebraic) multiplicity of the eigenvalue 0 of the signless Laplacian is equal to the number of bipartite (connected) components.*

**Corollary B.4.** *Each connected component in a bipartite graph $G$ on $n$ vertices corresponds to a basis vector $\mathbf{v} \in \{-1, 0, 1\}^n$ of $K[Q(G)]$.*

*Proof of Corollary B.4.* Observation B.4 applies to connected components, because they are connected subgraphs. If $G$ is not a connected graph, the vectors from Observation B.4 are extended by setting vector components to 0 for vertices not in the connected component. Let $\mathbf{v}_k$ denote the vector associated in this way to the connected component $C_k$ of $G$. Then the set of vectors $\{\mathbf{v}_k\}$ is linearly independent, because the vectors have disjoint support.

Since $Q(G)$ is diagonalizable, the algebraic and geometric multiplicities of its eigenvalues are equal [22]. From Observation B.3 and the previous statement, the geometric multiplicity of the 0 eigenvalue of $Q(G)$ is the number of connected components in $G$, which is equal to the cardinality of $\{\mathbf{v}_k\}$. Suppose $|\{\mathbf{v}_k\}| = m$. Then we have $m$ linearly independent vectors in the $m$-dimensional kernel of $Q(G)$. The vectors therefore span $K[Q(G)]$. □

The next lemma allows us to use Observation 3 on $Q$-graphs.

**Lemma B.5.** *A vector $\mathbf{v} \in \{-1, 1\}^n$ exists in the kernel $Q(G)$ of a graph $G$ on $n$ vertices if and only if it is bipartite.*

*Proof of Lemma B.5.* Let $G$ be a bipartite graph on $n$ vertices. Let $\{\mathbf{v}_k\}$ be vectors derived from connected components, including isolated vertices, of $G$ as described in Corollary B.4. Then, because the vectors $\{\mathbf{v}_k\}$ have disjoint support, the sum $\sum_k \mathbf{v}_k =: \mathbf{v} \in \{-1, 1\}^n$ and $Q(G)\mathbf{v} = 0$.

If a vector $\mathbf{v} \in \{-1, 1\}^n$ is in $K[Q(G)]$, then $Q(G)\mathbf{v} = 0$, meaning $R^T\mathbf{v} = 0$. It then follows from Proposition 2.1 in [21] that $G$ is bipartite. □

Finally, we prove the degree criterion for $Q$-graphs.

**Theorem 4** (Degree Criterion for $Q$-graphs)**.** *Let $G$ be a $Q$-graph. If $\rho_Q(G)$ is separable and $G^\Gamma$ is bipartite, then $D(G) = D(G^\Gamma)$.*

*Proof of Theorem 4.* Using Lemma B.5, the proof follows from applying Observation 3 to $Q$-graphs. □

For the proof of Observation 5, we assign a notion of vectors to vertices in a grid-labelled graph. The vector of a vertex is the standard basis vector corresponding to its index. In a grid-labelled graph, the vertices are indexed row-wise from top-left to bottom-right. Thus, in an $m \times n$ grid-labelled graph, the vertex $(0, 0)$ is the first vertex and is assigned the standard basis vector $\mathbf{e}_1$. The vertex $(m-1, n-1)$ is the last vertex and is assigned the vector $\mathbf{e}_{m\cdot n}$. This is convenient because the state vector of the state $|0\,0\rangle$ is $\mathbf{e}_1$ and of $|m-1, n-1\rangle$ is $\mathbf{e}_{mn}$. With this convention, we can say vertex $(i, j)$ corresponds to the state $|i\,j\rangle$.

**Observation 5.** *Let $G$ be a bipartite $Q$-graph on $n$ vertices with an isolated vertex $(i, j)$. If a product vector $|\mu\,\nu\rangle \in R[\rho_Q(G)]$, where $R$ denotes the range, then*

- $|\mu\,\nu\rangle \in R[\rho_Q(G^R_{ij})]$ *or* $R\big[\rho_L(G^R_{ij})\big]$*, or*
- $|\mu\,\mu\rangle \in R\left[\rho_Q(G^C_{ij})\right]$ *or* $R\left[\rho_L(G^C_{ij})\right]$*, or*
- $|\mu\,\nu\rangle \in R\left[\rho(G')\right]$*,*

*where $G'$ is a hybrid graph (see Section IV).*

*Proof of Observation 5.* Given vertex $(i, j)$ is an isolated vertex and thus a connected component, by Corollary B.4, $\rho_Q(G)\,|i\,j\rangle = 0$. Since $\rho_Q(G)$ is hermitian, $\langle\mu\,\nu|i\,j\rangle = 0$, which implies either $\langle i|\mu\rangle = 0$ or $\langle j|\nu\rangle = 0$. We first consider the case $\langle i|\mu\rangle = 0$, from which it follows that the inner product $\langle\mu\,\nu|i\,j_c\rangle = 0$ for all $c$. This means $|\mu\,\nu\rangle$ is orthogonal to states corresponding to all vertices in row $i$.

Let $C_k$ be a connected component in $G$ and $|C_k\rangle \coloneqq \mathbf{v}_k$ be the basis vector from Corollary B.4 of $K[\rho_Q(G)]$. Then $\langle\mu\,\nu|C_k\rangle = 0$.

Consider the vector $|C'_k\rangle \coloneqq |C_k\rangle + |L\rangle$, where $|L\rangle \coloneqq \sum_c \lambda_c\,|i\,j_c\rangle$ is a linear combination of vectors of all vertices in row $i$. A suitable set of scalars $\{\lambda_c\}$ can always be chosen to make $|C'_k\rangle_c = 0$ for all $c$. Using Corollary B.4, the vector $|C'_k\rangle$ can be interpreted as the vector of a connected component $C'_k$ that includes all vertices in $C_k$ except the ones in row $i$. Vertices in $C'_k$ have the same relative partitioning as in $C_k$. Further, $\langle\mu\,\nu|C'_k\rangle = 0$, because $\langle\mu\,\nu|L\rangle = 0$ as $\langle\mu\,\nu|i\,j_c\rangle = 0$ for all $c$, and $\langle\mu\,\nu|C_k\rangle = 0$.

Let $G'$ be a grid-labelled graph with the same vertex set as $G$. For every connected component $C_k$ in $G$, let the graph $G'$ have the connected component $C'_k$ derived

from $C_k$ as described above. Notice that the isolated vertices $\{(i_o, j_o)\}$ in $G$ remain isolated in $G'$, and that $G'$ has additional isolated vertices – the vertices in row $i$. The graph $G'$ thus can be produced via row surgery on $G$ with isolated vertex $(i, j)$. It can therefore be labelled as $G^R_{ij}$.

Depending the nature of the vectors $\{|C'_k\rangle\}$, we have three possibilities:

- If the vectors $\{|C'_k\rangle\}$ are all in $\{1, 0\}^n$, then $G^R_{ij}$ is an $L$-graph. The kernel of $L(G^R_{ij})$ is spanned by the vectors $\{|C'_k\rangle\}$, $\{|i_o\, j_o\rangle\}$, and $\{|i, j_c\rangle\}$ of its connected components, to all of which $|\mu\, \nu\rangle$ is orthogonal. Thus $|\mu\, \nu\rangle$ is in the range of $L(G^R_{ij})$ and also of $\rho_L(G^R_{ij})$. This case is identical to $L$-surgery.

- If the vectors $\{|C'_k\rangle\}$ are all in $\{1, 0, -1\}^n$, then by Corollary B.4 and arguments analogous to above, the vector $|\mu\, \nu\rangle$ is in the range of $\rho_Q(G^R_{ij})$.

- Finally, if some vectors in $\{|C'_k\rangle\}$ are in $\{1, 0\}^n$ and others in $\{1, 0, -1\}^n$, then $G'$ is a hybrid graph. Graph surgery on hybrid graphs are presented in Section IV.

It can be shown with analogous arguments that if instead $\langle l|j\rangle = 0$, then $|\mu\, \nu\rangle$ is in the range of $\rho_L(G^C_{ij})$ or of $\rho_Q(G^C_{ij})$ or of the density matrix of an analogous hybrid graph. □

We now show the unitary inequivalence of the $L$- and the $Q$-grid states corresponding to the same non-bipartite grid-labelled graph.

**Observation 6.** *Let $G$ be a grid-labelled graph. If $G$ is not bipartite, then $\rho_L(G)$ and $\rho_Q(G)$ are not unitarily equivalent.*

*Proof of Observation 6.* Let $G$ be a non-bipartite grid-labelled graph. The dimension of $K[L(G)]$ is the number of connected components in $G$ (see Section 3.13.5 in [23]). From Corollary B.4, the dimension of $K[Q(G)]$ is the number of bipartite connected components in $G$. At least one connected component in $G$ is not bipartite. This means the dimensions of $K[L(G)]$ and of $K[Q(G)]$ are not equal. Then from the rank-nullity theorem, the ranks of $L(G)$ and of $Q(G)$ are not equal. Therefore, $\rho_L(G)$ and $\rho_Q(G)$ cannot be unitarily equivalent. □

## Appendix C: Weighted Graphs

This section consists of proof of results stated for weighted grid-labelled graphs in the main text.

**Lemma 7.** *If the vertex and the edge sets of two weighted $L$-graphs (resp. $Q$-graphs) are identical, their signed (resp. signless) Laplacians have identical kernels.*

*Proof of Lemma 7.* Let $G = (V, E)$ be a weighted graph and edge weights of edges in $G$ be $\{w_1, \ldots w_m\}$, where $m = |E|$. If $Q\mathbf{v} = \mathbf{0}$, then

$$\left[R^T\mathbf{v}\right]_i = \sqrt{w_i}\,(\mathbf{v}_{i1} + \mathbf{v}_{i2}) = \mathbf{0},\ \forall i \in \{1, \ldots, m\}, \quad \text{(C1)}$$

because $Q = RR^T$, where $R$ is the unoriented incidence matrix. The vector components $\{\mathbf{v}_{i1}, \mathbf{v}_{i2}\}$ correspond to vertices connected by edge $e_i \in E$. The solutions of Equation (C1) are independent of the edge weights. Therefore, any vector $\mathbf{v} \in K[Q(G)]$ must also be in the kernels $\{K[Q(G')]\}$ of all graphs $\{G'\}$ with the same edge and vertex sets. The same arguments apply to the signed Laplacian $L(G)$. □

**Corollary 8.** *Let $G_w$ be a weighted grid-labelled graph. If $G_w$ is not bipartite, then $\rho_L(G_w)$ and $\rho_Q(G_w)$ are not unitarily equivalent.*

*Proof of Corollary 8.* Let $G_w = (V, E)$ be a non-bipartite weighted grid-labelled graph and $G = (V, E)$ be its unweighted counterpart. From the proof Observation 6, we know $\rho_L(G)$ and $\rho_Q(G)$ are not unitarily equivalent because their ranks are not equal. According to Lemma 7, $K[\rho_L(G)] = K[\rho_L(G_w)]$ and $K[\rho_Q(G)] = K[\rho_Q(G_w)]$. This means that the ranks of $\rho_L(G_w)$ and $\rho_Q(G_w)$ are not equal. Therefore, the density matrices cannot be unitarily equivalent.

□

## Appendix D: Hybrid Graphs

Here, we prove the results for grid-states derived from the grid-labelled hybrid graphs. To proceed, we need a notion of incidence matrix. The *incidence matrix* of a hybrid graph $G = (V, E)$ is the $|V| \times |E|$ matrix $\mathcal{R} = [B_l \quad R_q]$, where $B_l$ and $R_q$ are the unoriented and the oriented incident matrices of its $L$- and $Q$-subgraphs, respectively. The hybrid Laplacian satisfies $\mathcal{L} = \mathcal{R}\mathcal{R}^T$.

Like in the case of $Q$-grid states, we need supporting lemmas to prove the degree criterion for NOI- and COI-graphs.

**Lemma D.1.** *Each connected component in a NOI- or a COI-graph $G$ on $n$ vertices corresponds to a basis vector $\mathbf{v} \in \{-1, 0, 1\}^n$ of $K[\mathcal{L}(G)]$.*

*Proof of Lemma D.1.* The proof follows for adapting the arguments in the proof of Corollary B.4 to NOI- and COI-graphs. □

**Lemma D.2.** *For any NOI- or COI-graph $G$ on $n$ vertices there exists a vector $\mathbf{v} \in \{-1, 1\}^n$ in the kernel of $\mathcal{L}(G)$.*

*Proof of Lemma D.2.* With Lemma D.1, arguments analogous to the ones given in the proof Lemma B.5 prove this lemma.

□

**Theorem 9** (Degree Criterion)**.** *If the density matrix $\rho(G)$ of a hybrid graph $G$ is separable and $G^\Gamma$ is a NOI- or a COI-graph, then $D(G) = D(G^\Gamma)$.*

*Proof of Theorem 9.* Using Lemma D.2, the proof follows from applying Observation 3 to a NOI- or a COI-graph. □

We now prove the claim that every COI graph has a proxy graph.

**Observation 10.** *Every COI-graph has a proxy graph.*

*Proof of Observation 10.* Let $G$ be a COI-graph with two vertex partitions $P_1$ and $P_2$ determined by its $Q$-subgraph.

First, note that any connected component that contains a $Q$-edge must contain at least one vertex in partition $P_1$, since $Q$-edges connect vertices in opposite partitions. Second, by definition, the pair of vertices connected by any $L$-edge in $G$ must both be in the same partition. Using these observations, we can construct the proxy graph as follows:

- For each connected component that contains a $Q$-edge, choose two designated vertices – one in partition $P_1$ and the other is partition $P_2$.
- Then, for all vertices in the graph that have both an $L$-edge and a $Q$-edge incident, remove the $L$-edge.
- If a vertex belonging to partition $P_1$ (resp. $P_2$) is isolated from its previous connected component, reconnect it with a $Q$-edge to the corresponding designated vertex in partition $P_2$ (resp. $P_1$).

The above steps not only yield a NOI-graph, say $G'$, but also guarantee that the relative vertex partitioning of the vertices in $G$ and in $G'$ remain identical, and that all connected components in $G'$ have the same vertices as in their counterpart in $G$. Therefore, the vectors associated to connected components in $G$ and to connected components in $G'$ are identical. Then, from Lemma D.1, it follows that the kernels of $\mathcal{L}(G)$ and of $\mathcal{L}(G')$ are identical. □

Finally, we show that in the case of hybrid graphs as well the edge weights alone do not affect the kernel of the hybrid laplacian.

**Lemma 11.** *If the vertex and the edge sets of two weighted hybrid graphs are identical, their hybrid Laplacians have identical kernels.*

*Proof of Lemma 11.* Let $G$ be a weighted hybrid graph and $\mathcal{L}$ be its hybrid Laplacian matrix. Its incidence matrix is $\mathcal{R} = [B_l \quad R_q]$, where $B_l$ and $R_q$ are the signed and the signed Laplacian matrices of its $L$- and $Q$-subgraphs, respectively. Since $\mathcal{L} = \mathcal{R}\mathcal{R}^T$, by the same arguments as in the proof of Lemma 7, the solutions to the equation $\mathcal{L}\mathbf{v} = 0$ are independent of the edge weights. □

## Appendix E: Construction of Bound Entangled States

The proofs of two results related to construction of bound entangled states are given here.

**Observation 12.** *If a grid-labelled graph $G$ satisfies $D(G) = D(G^\Gamma)$, the corresponding density matrix has a positive partial transpose, independent of whether the graph is interpreted as an $L$-graph, a $Q$-graph, a weighted graph, or a hybrid graph.*

*Proof of Observation 12.* Normalization is ignored as it has no effect on the definiteness of a matrix. Let $G$ be a $Q$-graph and $G^\Gamma$ be its partial transpose. Given $D(G) = D(G^\Gamma)$,

$$D(G) = Q(G) - A(G) = D(G^\Gamma). \tag{E1}$$

Thus,

$$\begin{aligned} Q(G) &= D(G^\Gamma) + A(G). \\ \implies Q^\Gamma(G) &= D^\Gamma(G^\Gamma) + A^\Gamma(G) \\ &= D(G^\Gamma) + A(G^\Gamma) \\ &= Q(G^\Gamma) \geq 0. \end{aligned} \tag{E2}$$

The same arguments apply to weighted and to hybrid graphs. □

**Theorem 13.** *The density matrix of an $m \times n$ cross-hatch graph with $m, n \geq 3$ is bound entangled for all grid states independent of whether the graph is interpreted as an $L$-graph, a $Q$-graph, a weighted graph, or a hybrid graph.*

*Proof of Theorem 13.* An $m \times n$ cross-hatch $L$-graph is entangled for all $m, n \geq 3$ [7]. Graph surgery procedures on $Q$- and $L$-graphs only differ in the STITCH step, which is not required for graph surgery on cross-hatch graphs, because connected components in cross-hatch graphs are either isolated vertices or single edges. Therefore, the proof for $L$-graphs is sufficient for $Q$-graphs.

By Lemma 7, weighted cross-hatch $L$- and $Q$-graphs are entangled. Since graph surgery on hybrid graphs is based on $L$- and $Q$-surgeries, hybrid cross-hatch graphs are entangled. All cross-hatch graphs satisfy the degree criterion. Thus they are bound entangled. □

## Appendix F: Hypergraphs

The degree criterion for hypergraph grid-states is proved below.

**Theorem 14.** *Let $H$ be a hypergraph and $G$ be its graph. If $\rho(H)$ is separable and $G^\Gamma$ is bipartite, then $D(G) = D(G^\Gamma)$.*

*Proof of Theorem 14.* Let $H$ be a hypergraph on $n$ vertices and $G$ be its graph. From Equation (7)

$$L(H) = Q(G) - O(H),$$

where $O(H)$ is the offset matrix. Then

$$\begin{aligned} L^\Gamma(H) &= Q^\Gamma(G) - O^\Gamma(H) \\ &= Q(G^\Gamma) + \Delta - O(H), \end{aligned} \tag{F1}$$

where $\Delta = D(G) - D(G^\Gamma)$, and the second equality follows from applying Equation (A5) to $G$.

The offset matrix $O(H)$ is positive semidefinite because it is a real, diagonal matrix with non-negative diagonal entries. And from the PPT criterion, $L^\Gamma(H) \geq 0$, because $H$ represents a separable state. This means

$$L^\Gamma(H) + O(H) = Q(G^\Gamma) + \Delta \geq 0. \tag{F2}$$

Since $G^\Gamma$ is bipartite, from Lemma B.5 and Lemma 7, there exists a vector $\mathbf{v} \in \{-1, 1\}^n$ in $K[Q(G^\Gamma)]$. The matrix $\Delta$ is traceless and diagonal matrix. Thus, from Lemma A.1, the matrix $Q(G^\Gamma) + \Delta \not\geq 0$. This is a contradiction. Therefore, $\Delta = D(G) - D(G^\Gamma) = 0$. □


[1] W. K. Wootters and W. S. Leng, Philosophical Transactions: Mathematical, Physical and Engineering Sciences **356**, 1717 (1998).
[2] D. Bruß, J. I. Cirac, P. Horodecki, F. Hulpke, B. Kraus, M. Lewenstein, and A. Sanpera, Journal of Modern Optics **49**, 1399 (2002), https://doi.org/10.1080/09500340110105975.
[3] L. Gurvits, in *Proceedings of the Thirty-fifth Annual ACM Symposium on Theory of Computing*, STOC '03 (ACM, New York, NY, USA, 2003) pp. 10–19.
[4] S. Gharibian, Quantum Info. Comput. **10**, 343–360 (2010).
[5] S. L. Braunstein, S. Ghosh, T. Mansour, S. Severini, and R. C. Wilson, Phys. Rev. A **73**, 012320 (2006).
[6] J. Lockhart, O. Gühne, and S. Severini, Physical Review A 10.1103/PhysRevA.97.062340 (2018), arXiv:1705.09261.
[7] J. Lockhart, *Combinatorial Structures in Quantum Information*, Ph.D. thesis, University College London, London (2019).
[8] J. Hastrup, K. Park, J. B. Brask, R. Filip, and U. L. Andersen, npj Quantum Information **7**, 17 (2021).
[9] R. Diestel, *Graph Theory*, Graduate Texts in Mathematics (Springer Berlin Heidelberg, 2010).
[10] P. Horodecki, Physics Letters, Section A: General, Atomic and Solid State Physics 10.1016/S0375-9601(97)00416-7 (1997).
[11] M. Horodecki, P. Horodecki, and R. Horodecki, Physics Letters, Section A: General, Atomic and Solid State Physics 10.1016/S0375-9601(96)00706-2 (1996).
[12] B. Mohar, Some applications of laplace eigenvalues of graphs, in *Graph Symmetry: Algebraic Methods and Applications*, edited by G. Hahn and G. Sabidussi (Springer Netherlands, Dordrecht, 1997) pp. 225–275.
[13] Y. Hou, J. Li, and Y. Pan, Linear & Multilinear Algebra **51**, 21 (2003).
[14] S. Apers and R. de Wolf, in *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)* (2020) pp. 637–648.
[15] F. Chung, DIMACS Series in Discrete Mathematics and Theoretical Computer Science **10**, 21 (1993).
[16] A. Banerjee, Linear Algebra and its Applications https://doi.org/10.1016/j.laa.2020.01.012 (2020).
[17] J. Rodri´guez, Linear and Multilinear Algebra **50**, 1 (2002), https://doi.org/10.1080/03081080290011692.
[18] S. Hu and L. Qi, Journal of Combinatorial Optimization **29**, 331 (2015).
[19] A. Bondy and U. Murty, *Graph Theory*, Graduate Texts in Mathematics (Springer London, 2009).
[20] F. Chung, *Spectral graph theory* (Published for the Conference Board of the mathematical sciences by the American Mathematical Society, Providence, R.I, 1997).
[21] D. Cvetković, P. Rowlinson, and S. K. Simić, Linear Algebra and its Applications **423**, 155 (2007), special Issue devoted to papers presented at the Aveiro Workshop on Graph Spectra.
[22] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra* (Society for Industrial and Applied Mathematics, USA, 2000).
[23] A. Brouwer and W. Haemers, *Spectra of Graphs*, Universitext (Springer New York, 2011).