Sampling Groups of Pauli Operators to Enhance Direct Fidelity Estimation

Júlia Barberà-Rodríguez[1], Mariana Navarro[1,2], and Leonardo Zambrano[1]

[1]ICFO - Institut de Ciencies Fotoniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels, Barcelona, Spain

[2]LuxQuanta Technologies S.L., Mediterranean Technology Park. Carrer d'Esteve Terradas, 1, Office 206, 08860 Castelldefels, Barcelona, Spain

arXiv:2501.19228v2 [quant-ph] 27 Jun 2025

Direct fidelity estimation is a protocol that estimates the fidelity between an experimental quantum state and a target pure state. By measuring the expectation values of Pauli operators selected through importance sampling, the method is exponentially faster than full quantum state tomography. We propose an enhanced direct fidelity estimation protocol that uses fewer copies of the experimental state by grouping Pauli operators before the sampling process. We derive analytical bounds on the measurement cost and estimator variance, showing improvements over the standard method. Numerical simulations validate our approach, demonstrating that for 8-qubit Haar-random states, our method achieves a one-third reduction in the required number of copies and reduces variance by an order of magnitude using only local measurements. These results underscore the potential of our protocol to enhance the efficiency of fidelity estimation in current quantum devices.

1 Introduction

The rapid advancements in quantum technologies have underscored the need for efficient methods to characterize quantum devices. Accurate characterization plays a crucial role in verifying hardware performance, diagnosing errors, and ensuring the reliability of quantum protocols. Among various characterization tasks, fidelity estimation is of particular importance, as it quantifies how closely a prepared quantum state matches a desired target state. Traditionally, this verification has relied on quantum state tomography (QST) [1, 2], which reconstructs the full quantum state of a physical system. However, QST is computationally expensive, requiring at least $\mathcal{O}(4^n)$ copies of the experimental state, where n is the number of qubits of the system. It also demands substantial classical resources for computation and storage, making it impractical for large quantum systems.

For fidelity estimation specifically, more targeted and efficient techniques can help address these challenges. One such method is direct fidelity estimation (DFE), which involves measuring the expectation values of Pauli operators selected via importance sampling [3–5]. DFE offers a practical and scalable approach for fidelity estimation, relying exclusively on local measurements and being exponentially faster than QST. However, in the worst case, the number of measurements required by DFE scales as $\mathcal{O}(2^n)$. Another promising approach are classical shadows (CS), which employ random measurements to estimate the expectation values of observables [6]. Although CS are effective for a variety of tasks, the scaling for fidelity estimation of arbitrary states is at least $\mathcal{O}(3^n)$ when constrained to local measurements. To mitigate this, a derandomized version of CS has been proposed, which reduces the measurement overhead for arbitrary observables [7]. Nevertheless, for the specific task of fidelity estimation with local measurements, DFE remains one of the most effective and practical methods available.

While DFE provides a more efficient framework for fidelity estimation, its practical implementation involves measuring a large number of Pauli operators and copies of the experimental state. This poses significant challenges for sys-

Leonardo Zambrano: leonardo.zambrano@icfo.eu

tems with many qubits, particularly in the context of noisy intermediate-scale quantum (NISQ) devices where measurement resources are limited [8]. To address these challenges, improved measurement strategies are necessary to optimize the extraction of information from each experimental setting. A widely studied solution involves grouping Pauli operators that commute, enabling them to be measured simultaneously within a single experimental setting [9]. Such grouping techniques, commonly used in estimating Hamiltonian expectation values, significantly reduce the number of measurements required without sacrificing accuracy. Over the past decade, various methods have been proposed to optimize Pauli operator groupings, leading to substantial improvements in the scalability of operator estimation [10–18]. In addition to their use in Hamiltonian expectation value estimation, grouping techniques have also been applied to arbitrary observable estimation within the framework of classical shadows [19–21]. However, these methods are primarily designed for general tasks and have not been particularly optimized for fidelity estimation.

In this work, we propose a method to reduce the number of measurements required in DFE based on Pauli grouping. Instead of evaluating individual Pauli strings, we group them into commuting families and perform weighted sampling over these groups. This approach allows us to derive tighter bounds on the number of copies of the state required for the protocol. Additionally, we prove that the variance of our estimator is upper bounded by the variance of the original DFE estimator under certain conditions. Our theoretical analysis demonstrates that this optimized protocol outperforms the standard DFE in terms of both measurement cost and statistical efficiency. We validate our theoretical results through numerical simulations, exploring various grouping techniques. The simulations reveal that even when restricted to local measurements, our method significantly reduces the number of copies required by one-third while achieving a reduction of one order of magnitude in the variance compared to the original DFE protocol.

This paper is structured as follows. Section 2 provides an overview of the DFE protocol and existing Pauli grouping techniques. In Section 3, we detail our optimized approach and present variance and measurement cost bounds. Section 4 discusses the results of numerical simulations, with conclusions drawn in Section 5.

2 Preliminaries

2.1 Direct fidelity estimation

Assume we have a n-qubit state σ prepared in a physical system and a target state ρ. The protocol of direct fidelity estimation (DFE) allows us to estimate the fidelity $F(\rho,\sigma) = \mathrm{Tr}(\rho\sigma)$, given that ρ is a pure state [3]. Both ρ and σ states can be expressed in terms of their Pauli expectation values as

$$\sigma = \sum_k \frac{a_k}{\sqrt{d}} P_k, \qquad \rho = \sum_k \frac{b_k}{\sqrt{d}} P_k. \tag{1}$$

Here, each P_k is a tensor product of n-qubit Pauli operators or the identity $\{\sigma_x, \sigma_y, \sigma_z, \mathbb{1}\}$, with associated coefficients $a_k = \mathrm{Tr}(\sigma P_k/\sqrt{d})$ and b_k defined analogously for ρ, where $d = 2^n$ is the dimension of the Hilbert space. Then, the fidelity between these two states reads

$$F(\rho,\sigma) = \sum_k a_k b_k = \sum_k b_k^2 \frac{a_k}{b_k}, \tag{2}$$

where the sum runs over all k for which $b_k \neq 0$. Since the target state ρ is pure, it is satisfied that $\sum_k b_k^2 = 1$. Consequently, the fidelity can be interpreted as the expected value of a random variable X, with values

$$X_k = \frac{a_k}{b_k} = \frac{\mathrm{Tr}(\sigma P_k/\sqrt{d})}{\mathrm{Tr}(\rho P_k/\sqrt{d})} \tag{3}$$

and probability distribution $\Pr(X = X_k) = b_k^2$.

By applying Monte Carlo importance sampling to the probability distribution b_k^2, we can estimate the observable X_k with additive error ϵ and failure probability δ. The steps to achieve this are outlined below, which form the core of the DFE protocol [3, 22]:

1. Generate ℓ independent and identically distributed random samples $k_1, ..., k_\ell$ from the importance sampling distribution $b_{k_i}^2$, where $i \in \{1, .., \ell\}$ and $\ell := \lceil (\epsilon^2\delta)^{-1} \rceil$.

2. Measure each observable P_{k_i} a number of m_{k_i} times, where m_{k_i} is defined as

$$m_{k_i} := \left\lceil \frac{2}{b_{k_i}^2 d\ell\epsilon^2} \ln\frac{2}{\delta} \right\rceil. \tag{4}$$

3. For each i, calculate the empirical estimate of the expectation value $\mathrm{Tr}(\sigma P_{k_i})$ based on the measurements outcomes from step 2. Using these estimates, calculate the estimator $\hat{X}_{k_i}$ for $X_{k_i} := a_{k_i}/b_{k_i}$.

4. Calculate $\hat{Y} := \frac{1}{\ell}\sum_{i=1}^{\ell} \hat{X}_{k_i}$.

5. Return $\hat{Y}$ as a fidelity estimator.

The fidelity estimate $\hat{Y}$ serves as an unbiased estimator of $F(\rho,\sigma)$ with an accuracy of 2ϵ and a confidence level of $1-2\delta$. The protocol guarantees that the expected number of state preparations m_{k_i} is bounded and scales at most linearly with the Hilbert space dimension.

2.2 Grouping Pauli operators

Any n-qubit quantum state ρ can be expressed as a linear combination of Pauli strings as shown in Eq. (1). These strings can be grouped into sets of mutually commuting elements, such that

$$\rho = \sum_{k=1}^{g}\sum_{l=1}^{g_k} a_{kl} P_{kl}, \tag{5}$$

where g denotes the total number of groups, P_{kl} represents the Pauli strings within the k-th group, a_{kl} are their corresponding weights, and g_k is the number of Pauli strings in the k-th group. This grouping enables all operators within each set to be measured on a common basis, thereby reducing the number of experimental settings required and simplifying the measurement process [18].

There are two key aspects to consider in the grouping process. The first involves selecting an appropriate commutativity framework, with two primary approaches commonly employed. The full commutativity (FC) [13] technique groups Pauli strings when the associated operators commute, that is, $[P_k, P_l] = \left[\otimes_{j=1}^{n}\sigma_{kj}, \otimes_{j=1}^{n}\sigma_{lj}\right] = 0$. On the other hand, qubit-wise commutativity (QWC) [11] groups Pauli strings only if they commute at the level of single-qubit operators. Specifically, P_k and P_l can belong to the same group if $[\sigma_{kj}, \sigma_{lj}] = 0$ for all $j = 1, \ldots, n$. Although, in general, FC reduces the number of measurements needed and the variance of estimators compared to QWC, the implementation of FC requires two-qubit gates to perform the necessary basis changes for simultaneous measurements that increase the experimental overhead [17, 18, 23].

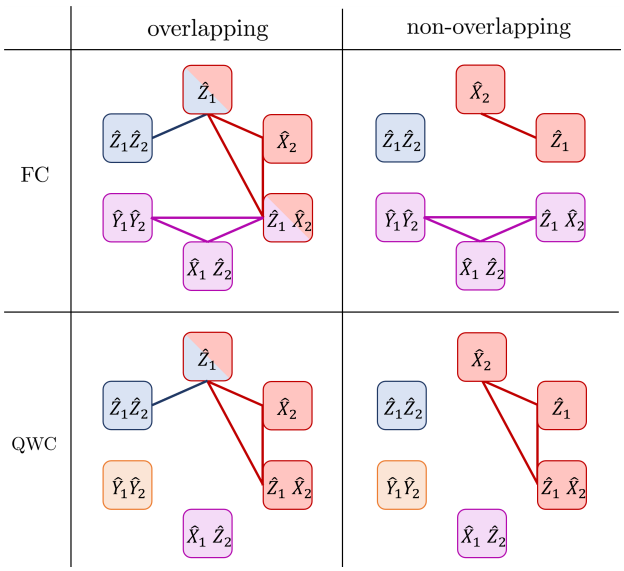


Figure 1: Diagram illustrating the different frameworks for grouping Pauli operators into commuting families. The x-axis distinguishes between overlapping groupings, where Pauli operators can belong to multiple groups, and non-overlapping groupings, where each Pauli operator is assigned to a single group. The y-axis differentiates between qubit-wise commutativity (QWC) and full commutativity (FC). Pauli strings in the same group share a color and are linked by solid lines.

The second consideration is whether to allow Pauli operators to appear in multiple groups. Initial methods, such as the sorted insertion (SI) algorithm [17], followed a non-overlapping framework, where each Pauli operator is assigned to a single group. Using these non-overlapping constraints and operating under QWC, SI minimizes variance by sorting Pauli operators by their weights in descending order before grouping them. While SI is simple, overlapping frameworks have shown improved performance by allowing Pauli operators to appear in multiple groups [21, 24]. Methods like iterative coefficient splitting (ICS) and iterative measurement allocation (IMA) [18] exploit overlaps to further reduce the number of measurements by redistributing coefficients among groups based on the covariance between commuting Pauli operators. Although ICS and IMA offer better performance in terms of reducing measurements and variance, ICS and IMA rely on SI to first generate non-overlapping groups, followed by additional steps for overlapping. This introduces extra computational costs, reducing their overall efficiency compared to SI. An illustrated comparison of the different grouping frameworks can be seen in Fig. 1.

3 Method

Considering an n-qubit system, the fidelity in Eq. (2) requires evaluating the expectation values of up to 4^n Pauli strings. This number can be significantly reduced by grouping the Pauli strings into g groups of mutually commuting operators, which can be measured simultaneously. For $k = 1, 2, \ldots, g$, the k-th group is defined as $\{P_{k1}, P_{k2}, \ldots P_{kg_k}\}$, where g_k is the number of elements in the group. For states σ and ρ we define the vectors $\mathbf{a}_k = (a_{k1}, a_{k2}, \ldots, a_{kg_k})$ and $\mathbf{b}_k = (b_{k1}, b_{k2}, \ldots, b_{kg_k})$, where $a_{kl} = \mathrm{Tr}(\sigma P_{kl}/\sqrt{d})$ and $b_{kl} = \mathrm{Tr}(\rho P_{kl}/\sqrt{d})$. Using this grouping, the fidelity between the target pure state ρ and the experimental state σ can be expressed as

$$F(\rho, \sigma) = \sum_{k=1}^{g} \mathbf{a}_k \cdot \mathbf{b}_k = \sum_{k=1}^{g} \|\mathbf{b}_k\|^2 \frac{\mathbf{a}_k \cdot \mathbf{b}_k}{\|\mathbf{b}_k\|^2}, \quad (6)$$

where $\mathbf{a}_k \cdot \mathbf{b}_k = \sum_{l=1}^{g_k} a_{kl} b_{kl}$ and

$$\|\mathbf{b}_k\| = \left(\sum_{l=1}^{g_k} b_{kl}^2 \right)^{1/2} \quad (7)$$

is the Euclidian norm of $\mathbf{b}_k$. Since ρ is a pure state, $\sum_{k=1}^{g} \|\mathbf{b}_k\|^2 = 1$.

Building on the original DFE protocol outlined in Sec. 2.1, we introduce a modified probability distribution $\|\mathbf{b}_k\|^2$, determined by the coefficients of the grouped Pauli strings associated with the target pure state ρ. Consequently, the values X_k in Eq. (3) of the random variable X are redefined as

$$X_k = \frac{\mathbf{a}_k \cdot \mathbf{b}_k}{\|\mathbf{b}_k\|^2} = \sum_{l=1}^{g_k} \frac{a_{kl} b_{kl}}{\sum_{j=1}^{g_k} b_{kj}^2}. \quad (8)$$

The exact value of X_k can be determined under the assumption that $\mathbf{a}_k$ is estimated perfectly, using as many copies of σ as necessary.

To apply Monte Carlo importance sampling with the probability distribution $P(X = X_k) = \|\mathbf{b}_k\|^2$, we would like to satisfy two conditions. First, we demand X to be an unbiased estimator of $\mathrm{Tr}(\rho\sigma)$ with respect to the probability distribution $\|\mathbf{b}_k\|^2$, that is, $\mathbb{E}_{X \sim \|\mathbf{b}_k\|^2}[X] = \mathrm{Tr}(\rho\sigma)$. This follows directly from Eq. (6). The second condition is that the variance of the estimator is bounded. This can be seen from

$$\begin{aligned} \mathrm{Var}(X) &= \sum_{k=1}^{g} \|\mathbf{b}_k\|^2 \left(\frac{\mathbf{a}_k \cdot \mathbf{b}_k}{\|\mathbf{b}_k\|^2} \right)^2 - \mathrm{Tr}(\rho\sigma)^2 \\ &\leq \sum_{k=1}^{g} \frac{1}{\|\mathbf{b}_k\|^2} \|\mathbf{a}_k\|^2 \|\mathbf{b}_k\|^2 \\ &\leq 1, \end{aligned} \quad (9)$$

where in the second line we have used the Cauchy–Schwarz inequality.

We aim to estimate $F(\rho, \sigma)$ with additive error 2ϵ and confidence $1 - 2\delta$. To achieve this, we generate $\ell = \lceil (\epsilon^2 \delta)^{-1} \rceil$ independent and identically distributed samples $k_1, \ldots, k_\ell$ according to the probability distribution $\|\mathbf{b}_k\|^2$. Each k_i corresponds to an independent realization X_{k_i} of the random variable X. Assuming that all the X_{k_i} can be known perfectly, the empirical mean is

$$Y = \frac{1}{\ell} \sum_{i=1}^{\ell} X_{k_i}. \quad (10)$$

By applying Chebyshev's inequality, we have that Y satisfies

$$\Pr(|Y - \mathrm{Tr}(\rho\sigma)| \geq \epsilon) \leq \frac{1}{\ell \epsilon^2}. \quad (11)$$

Thus, the estimator Y is ϵ-close to the fidelity $F(\rho, \sigma)$ with probability at least $1 - \delta$, for any arbitrarily small δ.

In a more realistic setting, the ideal estimator Y, which assumes infinite precision and unlimited resources, is approximated by a finite-sample estimator $\tilde{Y}$ that uses a limited number of copies of the state σ. Then we need to establish a bound for the approximation error $|\tilde{Y} - Y|$ using Hoeffding's inequality. For this, we perform an eigendecomposition of the Pauli operators P_{kl} within group k. Since all the operators in each group commute, they share a common eigenbasis, denoted by $\{|r_k\rangle\}_{r=1}^{d}$. Then,

$$P_{kl} = \sum_{r=1}^{d} c_{kl}^{(r)} |r_k\rangle\langle r_k|, \quad (12)$$

where $\{c_{kl}^{(r)}\} \in \{-1, 1\}$ are the corresponding eigenvalues.

For any realization $k_1, \ldots, k_\ell$ of the experiment, the value k_i indicates that the group of Pauli strings $\{P_{k_i l}\}_{l=1}^{g_{k_i}}$ must be measured in the i-th round of the experiment. We prepare m_{k_i}

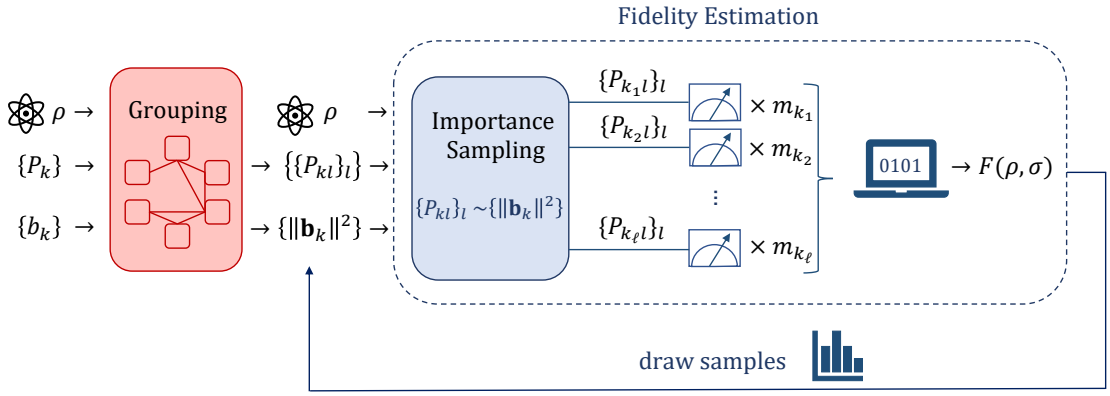


Figure 2: Improved protocol for estimating the fidelity between an unknown experimental state σ and a pure state ρ. The algorithm inputs include the n-qubit target state ρ, a list of n-qubit Pauli strings $\{P_k\}$, and their corresponding weights $\{b_k\}$. First, the Pauli strings are grouped into commuting families $\{\{P_{kl}\}_l\}$, and the new weights $\{\|\mathbf{b}_k\|^2\}$ are calculated using the chosen technique. The DFE protocol is then performed, using the groups as elements for the sampling process with the probability distribution $\{\|\mathbf{b}_k\|^2\}$. Each sampled group of Pauli strings is measured m_{k_i} times, requiring a total of $\sum_{k_i} m_{k_i}$ copies of the target state ρ.

(defined below) copies of the state σ, and measure them using the POVM $\{|r_{k_i}\rangle\langle r_{k_i}|\}_{r=1}^{d}$. For each measurement $j = 1, \dots m_{k_i}$, we obtain an outcome $r_{k_i j}$ with probability $p(r_{k_i j}|k_i) = \langle r_{k_i}|\sigma|r_{k_i}\rangle$. Then, we define the estimator $\tilde{X}_{k_i}$ of X_{k_i} as

$$\tilde{X}_{k_i} = \frac{1}{m_{k_i}\|\mathbf{b}_{k_i}\|^2\sqrt{d}} \sum_{j=1}^{m_{k_i}} C_{k_i r_j}, \tag{13}$$

where $C_{k_i r_j} = \sum_l c_{k_i l}^{(r_j)} b_{k_i l}$ and $c_{k_i l}^{(r_j)}$ denotes the r_j-th eigenvalue of the Pauli operator $P_{k_i l}$ obtained during the j-th round of the experiment. The expectation value of the measurement outcomes $c_{k_i l}^{(r_j)}$ is

$$\begin{aligned}\mathbb{E}\left[c_{k_i l}^{(r_j)}\right] &= \sum_{r=1}^{d} c_{k_i l}^{(r)} \langle r_{k_i}|\rho|r_{k_i}\rangle \\ &= \mathrm{Tr}\left(\sigma P_{k_i l}\right) \\ &= \sqrt{d}\, a_{k_i l}.\end{aligned} \tag{14}$$

Then, we have $\mathbb{E}\left[C_{k_i r_j}\right] = \sqrt{d}\sum_l a_{k_i l} b_{k_i l} = \sqrt{d}\mathbf{a}_{k_i}\cdot\mathbf{b}_{k_i}$ and since expectation values are linear, $\tilde{X}_k$ is an unbiased estimator of X_k.

We now define an unbiased estimator for the random variable $\tilde{Y}$ as follows

$$\begin{aligned}\ell\tilde{Y} &= \sum_{i=1}^{\ell} \tilde{X}_{k_i} \\ &= \sum_{i=1}^{\ell}\sum_{j=1}^{m_{k_i}} \frac{C_{k_i r_j}}{m_{k_i}\|\mathbf{b}_{k_i}\|^2\sqrt{d}}.\end{aligned} \tag{15}$$

Using Hölder's inequality, we observe that $|C_{k_i r_j}| \leq \|\mathbf{b}_{k_i}\|_1 \max_l |c_{k_i l}^{(r_j)}| = \|\mathbf{b}_{k_i}\|_1$, which provides a bound for the argument of the sum in Eq. (15):

$$\left|\frac{C_{k_i r_j}}{m_{k_i}\|\mathbf{b}_{k_i}\|^2\sqrt{d}}\right| \leq \frac{\|\mathbf{b}_{k_i}\|_1}{m_{k_i}\|\mathbf{b}_{k_i}\|^2\sqrt{d}}. \tag{16}$$

Applying Hoeffding's inequality for $\epsilon\ell > 0$ and using the bounds from Eq. (16), we obtain the following inequality

$$\begin{aligned}&\Pr(|\tilde{Y} - Y| \geq \epsilon) \\ &\leq 2\exp\left(\frac{-2\epsilon^2\ell^2}{\sum_{i=1}^{\ell}\sum_{j=1}^{m_{k_i}}\left(\frac{2\|\mathbf{b}_{k_i}\|_1}{m_{k_i}\|\mathbf{b}_{k_i}\|^2\sqrt{d}}\right)^2}\right) \\ &= 2\exp\left(\frac{-\epsilon^2\ell^2}{\sum_{i=1}^{\ell}\frac{2\|\mathbf{b}_{k_i}\|_1^2}{m_{k_i}\|\mathbf{b}_{k_i}\|^4 d}}\right).\end{aligned} \tag{17}$$

We want this probability to be smaller than a given δ, that is,

$$\Pr(|\tilde{Y} - Y| \geq \epsilon) \leq \delta. \tag{18}$$

For this to hold, we need

$$\ln\frac{2}{\delta} \leq \frac{\epsilon^2\ell^2}{\sum_{i=1}^{\ell}\frac{2\|\mathbf{b}_{k_i}\|_1^2}{m_{k_i}\|\mathbf{b}_{k_i}\|^4 d}}, \tag{19}$$

from which we obtain the number of copies of the state needed for each group,

$$m_{k_i} = \left\lceil \frac{2\|\mathbf{b}_{k_i}\|_1^2}{\|\mathbf{b}_{k_i}\|^4 d\ell\epsilon^2} \ln\frac{2}{\delta} \right\rceil. \tag{20}$$

We have established a bound on the approximation error $|Y - \tilde{Y}|$. By combining this result with the bound in Eq. (11) and using the union bound, we can conclude that

$$\Pr[\tilde{Y} - F(\rho, \sigma) \leq 2\epsilon] \geq 1 - 2\delta. \tag{21}$$

Thus, when employing our protocol (see Fig. 2), the fidelity is guaranteed to be within the interval $[\tilde{Y} - 2\epsilon, \tilde{Y} + 2\epsilon]$ with confidence $1 - 2\delta$.

Let us now calculate a bound for the average number of copies in our protocol. We sample $\ell = \lceil 1/(\epsilon^2\delta) \rceil$ groups, with each group measured m_{k_i} times. Then, the average number of copies per group is bounded by

$$\mathbb{E}[m_{k_i}] \leq 1 + \sum_{k=1}^{g} \frac{2\|\mathbf{b}_{k_i}\|_1^2}{\|\mathbf{b}_{k_i}\|^2 d\ell\epsilon^2} \ln\frac{2}{\delta}. \tag{22}$$

The total number of copies of the state is $m = \sum_{i=1}^{\ell} m_{k_i}$. Therefore, for the expected total number of copies of our protocol, we have

$$\mathbb{E}\left[\sum_{i=1}^{\ell} m_{k_i}\right] \leq 1 + \frac{1}{\epsilon^2\delta} + \sum_{k=1}^{g} \frac{2\|\mathbf{b}_{k_i}\|_1^2}{\|\mathbf{b}_{k_i}\|^2 d\epsilon^2} \ln\frac{2}{\delta}. \tag{23}$$

Since $\|\mathbf{b}_{k_i}\|_1 \leq \sqrt{g_{k_i}}\|\mathbf{b}_{k_i}\|$ and $\sum_{k=1}^{g} g_{k_i}$ is at most d^2 for an arbitrary grouping technique, we find that the total number of copies for our protocol is bounded by that of the original DFE protocol:

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{\ell} m_{k_i}\right] &\leq 1 + \frac{1}{\epsilon^2\delta} + \frac{2}{d\epsilon^2}\ln\frac{2}{\delta}\sum_{k=1}^{g} g_{k_i} \\ &\leq 1 + \frac{1}{\epsilon^2\delta} + \frac{2d}{\epsilon^2}\ln\frac{2}{\delta}.\end{aligned} \tag{24}$$

Moreover, in App. B, we show that the variance of our estimator, $\mathrm{Var}(\tilde{Y})$, is upper bounded by the variance of the original DFE estimator, $\mathrm{Var}(\hat{Y})$, when the coefficients satisfy $\|\mathbf{b}_k\|_1^2/\|\mathbf{b}_k\|^4 \gg 1$ (e.g. for Haar-random states when d is large, since $|b_{kl}| \approx 1/d$). In this regime, $m_k \gg 1$, and the approximation $\lceil m_k \rceil \approx m_k$ holds. Consequently, the fidelity estimates obtained with our protocol are more accurate than those from the original protocol.

Fidelity estimation can also be performed using shadow-based methods, such as classical shadows and their derandomized variant [6, 7]. For non-stabilizer states and local measurements, both classical shadows and the derandomized version exhibit a worst-case sample complexity of $\mathcal{O}(4^n/\epsilon^2)$. In contrast, standard DFE and our protocol require $\mathcal{O}(2^n/\epsilon^2)$ copies. While this scaling is an average-case analysis, by Markov's inequality, the actual cost is unlikely to deviate significantly from the average [3]. In the classical shadows framework with entangled (Clifford) measurements, the sample complexity becomes independent of the Hilbert space dimension. However, implementing entangled operations remains challenging and noise-prone on current quantum devices [25]. Moreover, as shown in Ref. [26], the associated classical post-processing overhead can exceed the total resource cost of DFE for non-stabilizer target states. Therefore, our protocol offers favorable resource scaling among the methods discussed in this paragraph.

4 Numerical simulations

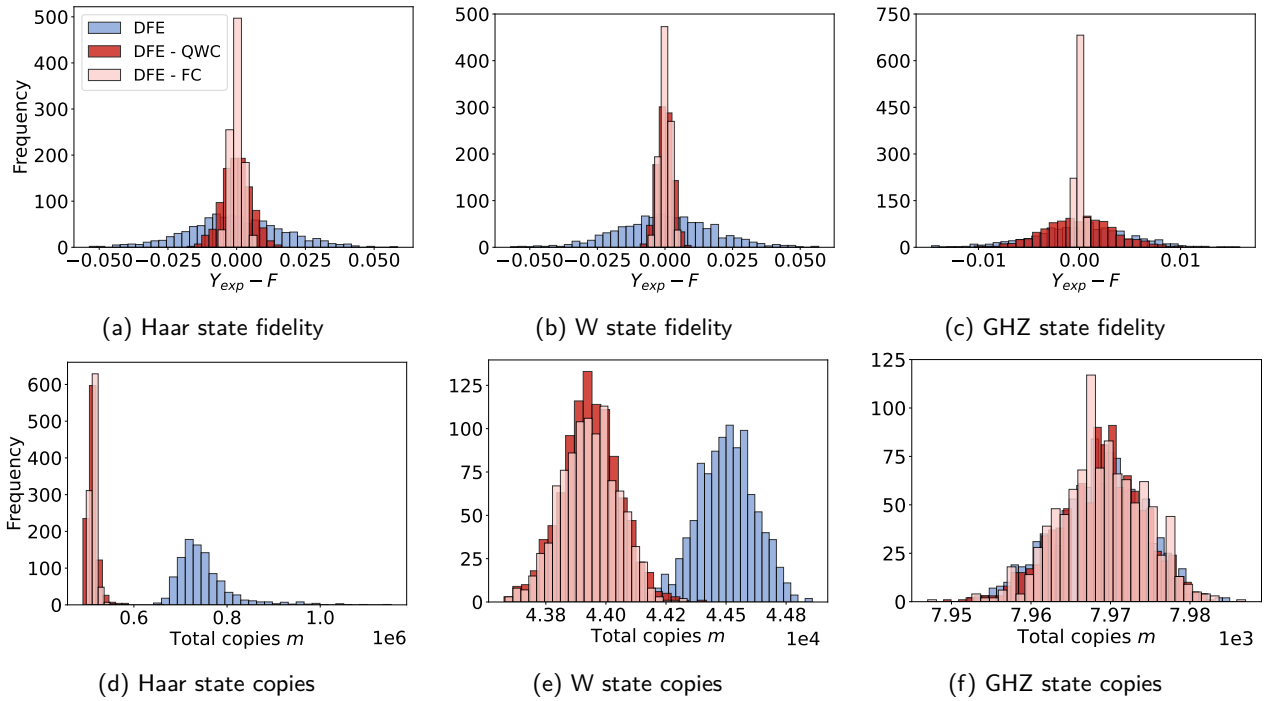
To study the performance of the proposed method, we carried out numerical simulations and benchmarked it against DFE protocol [3]. We tested fidelity estimation on three types of states: Haar-random states, W states, and the GHZ state. These ideal input states ρ were subjected to a depolarizing channel, producing the noisy state

$$\sigma := (1 - p)\rho + p\frac{\mathbb{1}}{d}, \tag{25}$$

where p is the noise level.

Our protocol differentiates from DFE by incorporating a preprocessing step, polynomial in d, that involves grouping Pauli strings. To achieve this, we employed the SI algorithm, within the non-overlapping framework, for both full and qubit-wise commutation. This algorithm minimizes the estimator variance among other grouping techniques while remaining computationally efficient [18]. The algorithm first sorts Pauli strings in decreasing order based on their weights, $|b_{kl}|$, and then forms the groups (see App. A for the pseudocode). This approach ensures that the largest groups consist of Pauli strings with the highest weights, thereby reducing the number of copies and the variance of the expected value of the operator under consideration.

In Fig. 3, we show the distribution of the residual error of the fidelity, represented as the frequency with which different error values occur across 1000 simulated 8-qubit states. We also report the total number of copies required. Both



(a) Haar state fidelity (b) W state fidelity (c) GHZ state fidelity

(d) Haar state copies (e) W state copies (f) GHZ state copies

Figure 3: The plots are obtained using 1000 samples of 8-qubit states with parameters set to $p = 0.1, \varepsilon = 0.05$, $\delta = 0.05$, $\ell = 8000$. In (a), (b), and (c), we show the distribution of the residual error of the fidelity (Y_{exp} - F), represented by the frequency of occurrence of different error values across samples. Results are shown for the DFE protocol (blue) and our method with QWC (red) and FC (blue) using Haar-random states, W states and GHZ states. Figures (d), (e), and (f) show the number of copies required by each protocol. For Haar-random states, the variance of Y_{exp} is reduced by 92% (99%) with QWC (FC) and the mean number of copies is reduced by 32%. In Fig. 3(d), we omit the last 5 samples from the dataset used in the original DFE and QWC analysis for better visualization.

metrics are benchmarked against the direct fidelity estimation protocol outlined in Sec. 2.1. The simulations were executed with parameters set to $p = 0.1, \varepsilon = 0.05$, $\delta = 0.05$, which resulted in $\ell = 1/\varepsilon^2\delta = 8000$. For Haar-random states, the results show that using grouping techniques for DFE significantly reduces the variance of the fidelity estimator by 92% for QWC and 99% for FC. Similarly, Fig. 3(d) highlights that the number of copies required decreases with grouping techniques. Both QWC and FC frameworks achieve a 32% reduction in the mean number of copies, demonstrating that grouping effectively reduces the number of copies keeping the same precision of fidelity estimation.

The results in Fig. 3(b) and Fig. 3(e) indicate that for the W state, the variance of Y_{exp} is reduced by 97% with QWC and 99% with FC, while the mean number of copies decreases by 1.27% compared to DFE with both grouping techniques. Finally, for the GHZ state, the variance of the fidelity estimator decreased by 48% with QWC and 99% with FC, as illustrated in Fig. 3(c). However, in Fig. 3(f), we see that the mean number of copies remained unchanged. These results on the number of copies align with observations in Ref. [3], which noted that their method was optimal when these states were used as input.

To further benchmark our protocol, we conducted experiments with 1000 samples of Haar-random states for different system dimensions, using the same simulation parameters as in Fig. 3: $p = 0.1$, $\varepsilon = 0.05$, $\delta = 0.05$, and $\ell = 8000$. The data obtained is shown in Tab. 1, where we observe that the number of groups scales polynomially with d, with fewer groups required for the FC technique. This observation lines up with the analysis made in Ref. [18] that numerically demonstrates that the total number of groups g under FC is significantly smaller than for QWC. We also note that the variance of Y_{exp} is consistently lower when FC is employed, and this reduction increases with d. In particular, the variance decreases by 53% and 63% for smaller systems, reaching reductions of 92% and 99% for larger systems, for QWC and FC, respectively. Similarly, using both commutating frameworks, the reduction of the mean number of copies also

n	g		Variance (1×10^{-5})			$\mathbb{E}[m]$		
	QWC	FC	DFE	QWC	FC	DFE	QWC	FC
2	10	7	12.2	7.2	4.8	13803	12630	12914
3	28	15	18.4	6.2	4.0	26660	22705	23545
4	82	33	25.3	6.3	2.5	50572	40126	42523
5	244	74	26.8	5.3	1.8	98439	73166	77357
6	730	172	28.7	4.1	1.2	191706	138117	142610
7	2188	399	33.1	3.3	0.7	380112	262896	268826
8	6562	924	33.6	2.6	0.4	754950	513139	513995

Table 1: Number of groups g of Pauli operators in the sorted insertion (SI) method for full and qubit-wise commutativity (FC and QWC) across number of qubits n, along with the variance of the fidelity estimator Y_{exp} and the mean number of copies $\mathbb{E}[m]$ for the original protocol and the two grouping techniques, with $m = \sum_{k_i} m_{k_i}$ and m_{k_i} given in Eq. (4) and Eq. (20).

grows with d, starting at approximately 6% for $n = 2$ and reaching 33% for $n = 8$. Notably, for smaller systems, QWC yields slightly greater reductions in the number of copies, which may be attributed to a more uneven distribution of the weights associated with the Pauli operators, favoring fewer copies.

Another potential approach to reduce the number of copies is through overlapping grouping techniques [18], such as the ICS algorithm introduced in Sec. 2.2. ICS forms overlapping groups based on the covariances of commuting Pauli operators but relies on the prior identification of non-overlapping groups using the SI algorithm [18]. However, this added classical complexity makes ICS a less efficient option for reducing resources in DFE.

5 Conclusions

Given the importance of the DFE protocol, it is crucial to develop improvements that reduce its measurement cost while maintaining precision. One promising approach is to optimize the measurement strategy by grouping commuting Pauli operators, which reduces the number of distinct measurement settings and improves estimation efficiency [10–21]. While such techniques have been widely explored for general observable estimation, their application to DFE has remained largely unexplored.

In this work, we integrated Pauli operator grouping techniques into the DFE protocol to reduce the total number of measurements required while maintaining precision. In our method, we group the Pauli strings appearing in the decomposition of the target state into commuting families and perform weighted sampling over these groups. The expectation values of these operators are then estimated by measuring in the common eigenbases of each group. We proved that, for a fixed error ϵ and confidence $1 - \delta$, the total number of copies required by our protocol is upper-bounded by that of the standard DFE protocol. Furthermore, we also upper-bounded the variance of our protocol by that of standard DFE under certain conditions.

Numerical simulations on 8-qubit Haar-random states confirmed the theoretical predictions. For local measurements, our approach achieved a 92% reduction in variance and a 32% reduction in the number of copies required compared to the standard DFE protocol. Additional reductions in variance were achieved when allowing for entangled measurements, though these come with higher implementation costs that are often avoided in practice.

While our protocol and simulations assume ideal measurements, experimental imperfections, such as gate and readout errors, can introduce bias and lead to systematic deviations in fidelity estimates. Fortunately, the protocol, when implemented within the QWC framework, relies only on single-qubit gates, which are among the most accurate operations available on current quantum hardware [27, 28]. Although readout errors re-

main a relevant source of noise, they could be mitigated by incorporating existing error mitigation techniques into our protocol [29–31]. Moreover, recent advances in error mitigation for randomized measurement protocols [32–34] may be adaptable to our setting, potentially further enhancing robustness. Investigating these directions could yield improved practical performance and pave the way for more noise-resilient implementations of our protocol.

Code availability

Simulations were run using `Pyhton` 3.11.9 on an Intel Core i5-7500 3.4 GHz 6 MB L3 cache. The computations could be optimized by utilizing a GPU or parallelization. The scripts to run the experiments presented in this work can be found in GitHub in the repository `juliabarbera/DFE-sampling-pauli-groups`.

Acknowledgments

This work was supported by the Government of Spain (Severo Ochoa CEX2019-000910-S, FUNQIP, Misiones CUCO Grant MIG-20211005, European Union NextGenerationEU PRTR-C17.I1 and Quantum Spain), Fundació Cellex, Fundació Mir-Puig and Generalitat de Catalunya (CERCA program and Grup reconegut 2021 SGR 01440). J.B.R. has received funding from the "Secretaria d'Universitats i Recerca del Departament de Recerca i Universitats de la Generalitat de Catalunya" under grant FI-1 00096, as well as the European Social Fund Plus. M.N. acknowledges funding from the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 101081441.

References

[1] D. F. V. James, P. G. Kwiat, W. J. Munro and A. G. White, Measurement of qubits, Phys. Rev. A **64**, 052312 (2001).

[2] R. T. Thew, K. Nemoto, A. G. White and W. J. Munro, Qudit quantum-state tomography, Phys. Rev. A **66**, 012303 (2002).

[3] S. T. Flammia and Y.-K. Liu, Direct Fidelity Estimation from Few Pauli Measurements, Phys. Rev. Lett. **106** (2011).

[4] M. P. da Silva, O. Landon-Cardinal and D. Poulin, Practical Characterization of Quantum Devices without Tomography, Phys. Rev. Lett. **107**, 210404 (2011).

[5] O. Fawzi, A. Oufkir and R. Salzmann, Optimal Fidelity Estimation from Binary Measurements for Discrete and Continuous Variable Systems (2024), arXiv:2409.04189 .

[6] H.-Y. Huang, R. Kueng and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nat. Phys. **16**, 1050 (2020).

[7] H.-Y. Huang, R. Kueng and J. Preskill, Efficient Estimation of Pauli Observables by Derandomization, Phys. Rev. Lett. **127**, 030503 (2021).

[8] K. Bharti et al., Noisy intermediate-scale quantum algorithms, Rev. Mod. Phys. **94**, 015004 (2022).

[9] A. Kandala et al., Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature **549**, 242 (2017).

[10] A. F. Izmaylov, T.-C. Yen and I. G. Ryabinkin, Revising the measurement process in the variational quantum eigensolver: is it possible to reduce the number of separately measured operators?, Chem. Sci. **10**, 3746 (2019).

[11] V. Verteletskyi, T.-C. Yen and A. F. Izmaylov, Measurement optimization in the variational quantum eigensolver using a minimum clique cover, J. Chem. Phys. **152** (2020).

[12] A. Jena, S. N. Genin and M. Mosca, Optimization of variational-quantum-eigensolver measurement by partitioning Pauli operators using multiqubit Clifford gates on noisy intermediate-scale quantum hardware, Phys. Rev. A **106**, 042443 (2022).

[13] T.-C. Yen, V. Verteletskyi and A. F. Izmaylov, Measuring all compatible operators in one series of single-qubit measurements using unitary transformations, J. Chem. Theory Comput. **16**, 2400 (2020).

[14] A. F. Izmaylov, T.-C. Yen, R. A. Lang and V. Verteletskyi, Unitary partitioning approach to the measurement problem in the variational quantum eigensolver method, J. Chem. Theory Comput. **16**, 190 (2019).

[15] A. Zhao et al., Measurement reduction in

variational quantum algorithms, Phys. Rev. A **101**, 062322 (2020).

[16] I. Hamamura and T. Imamichi, Efficient evaluation of quantum observables using entangled measurements, npj Quantum Inf. **6**, 56 (2020).

[17] O. Crawford et al., Efficient quantum measurement of Pauli operators in the presence of finite sampling error, Quantum **5**, 385 (2021).

[18] T.-C. Yen, A. Ganeshram and A. F. Izmaylov, Deterministic improvements of quantum measurements with grouping of compatible operators, non-local transformations, and covariance estimates, npj Quantum Inf. **9**, 14 (2023).

[19] C. Hadfield, S. Bravyi, R. Raymond and A. Mezzacapo, Measurements of quantum Hamiltonians with locally-biased classical shadows, Commun. Math. Phys. **391**, 951 (2022).

[20] A. Gresch and M. Kliesch, Guaranteed efficient energy estimation of quantum many-body Hamiltonians using ShadowGrouping, Nat. Commun. **16** (2025).

[21] B. Wu, J. Sun, Q. Huang and X. Yuan, Overlapped grouping measurement: A unified framework for measuring quantum states, Quantum **7**, 896 (2023).

[22] M. Kliesch and I. Roth, Theory of Quantum System Certification, PRX Quantum **2**, 010201 (2021).

[23] T.-C. Yen and A. F. Izmaylov, Cartan subalgebra approach to efficient measurements of quantum observables, PRX Quantum **2**, 040320 (2021).

[24] A. Shlosberg et al., Adaptive estimation of quantum observables, Quantum **7**, 906 (2023).

[25] J. Preskill, Quantum Computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[26] L. Leone, S. F. E. Oliviero and A. Hamma, Nonstabilizerness determining the hardness of direct fidelity estimation, Phys. Rev. A **107** (2023).

[27] R. Barends et al., Superconducting quantum circuits at the surface code threshold for fault tolerance, Nature **508**, 500 (2014).

[28] M. C. Smith, A. D. Leu, K. Miyanishi, M. F. Gely and D. M. Lucas, Single-Qubit Gates with Errors at the 10^{-7} Level, Phys. Rev. Lett. **134**, 230601 (2025).

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

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

[31] S. Bravyi, S. Sheldon, A. Kandala, D. C. Mckay and J. M. Gambetta, Mitigating measurement errors in multiqubit experiments, Phys. Rev. A **103**, 042605 (2021).

[32] A. Seshadri, M. Ringbauer, J. Spainhour, T. Monz and S. Becker, Theory of versatile fidelity estimation with confidence, Phys. Rev. A **110**, 012431 (2024).

[33] S. Chen, W. Yu, P. Zeng and S. T. Flammia, Robust Shadow Estimation, PRX Quantum **2**, 030348 (2021).

[34] H.-Y. Hu et al., Demonstration of robust and efficient quantum property learning with shallow shadows, Nat. Commun. **16** (2025).

A Sorted insertion

In this section, we present the pseudocode for the sorted insertion (SI) algorithm introduced in Ref. [17]. The SI algorithm is designed to group Pauli strings into commuting families while considering their variances. The goal is to minimize the total number of measurements required and ensure low variances in the resulting estimators.

Algorithm 1: SI Algorithm

Input: `pauli_list`: List of strings representing the Pauli operators, `x_ki_list`: List of coefficients corresponding to the Pauli operators, `condition`: Specifies the commutation framework, which can be either fully commuting or qubit-wise commuting.

Output: `no_groups`: List of groups, each containing commuting Pauli operators, `x_ki_no_groups`: A list of weights grouped in correspondence with the Pauli strings

Initialize: `no_groups` $\leftarrow \emptyset$, `x_ki_no_groups` $\leftarrow \emptyset$

Sort `pauli_list` in decreasing order based on the absolute values of their corresponding weights, `x_ki_list`

foreach *pauli in pauli_list with index i* **do**
  `added_to_any_group` $\leftarrow$ `False`
  foreach *group, x_ki_group in zip(no_groups, x_ki_no_groups)* **do**
    if *commutes(pauli, group, condition)* **then**
      Append `pauli` to `group`
      Append `x_ki_list[i]` to `x_ki_group`
      `added_to_any_group` $\leftarrow$ `True`
      break
    end
  end
  if ***not*** *added_to_any_group* **then**
    Append `[pauli]` to `no_groups`
    Append `[x_ki_list[i]]` to `x_ki_no_group`
  end
end
return *no_groups, x_ki_no_group*

B Variance of the original and grouped protocols

In this section, we demonstrate that, under the approximation $\lceil m_k \rceil \approx m_k$, the variance of our protocol is bounded by the variance of the original protocol. This suggests that our approach provides an improvement over the original method in terms of variance. The approximation holds when $\|\mathbf{b}_k\|_1^2/\|\mathbf{b}_k\|^4$ is sufficiently large, which is the case for Haar-random states. In such states, the coefficients $|b_k|$ are approximately $1/d$ for large d.

Let $\hat{Y}$ and $\tilde{Y}$ represent the fidelity estimators of DFE and our protocol, respectively. The objective is to show that $\mathrm{Var}(\tilde{Y}) \leq \mathrm{Var}(\hat{Y})$. Since both estimators are sums of independent and identically distributed random variables $\hat{X}_{k_i}$ and $\tilde{X}_{k_i}$, this is equivalent to proving that $\mathrm{Var}(\tilde{X}_{k_i}) \leq \mathrm{Var}(\hat{X}_{k_i})$. Furthermore, as the expected value of both estimators equals the theoretical fidelity (i.e., $\mathbb{E}[\tilde{X}_{k_i}]^2 = \mathbb{E}[\hat{X}_{k_i}]^2 = \mathrm{Tr}(\rho\sigma)^2$), the problem reduces to showing that $\mathbb{E}[\tilde{X}_{k_i}^2] \leq \mathbb{E}[\hat{X}_{k_i}^2]$. Let k_i be the index of the Pauli string measured in experiment round i, and $r = r_1 \dots r_{m_{k_i}}$ specifies which one of the d eigenstates of $\hat{X}_{k_i}$, given the Pauli string k_i, is obtained for each of the m_{k_i} measurement shots. Since for each $\hat{X}_{k_i}$ we are considering m_{k_i} shots, the random variables $\hat{X}_{k_i}$ in the DFE protocol follow the probability

distribution

$$p(k_i, r) = p(k_i)p(r|k_i) = p(k_i)p(r_1|k_i)\dots p(r_{m_{k_i}}|k_i) = b_{k_i}^2\langle r_1|\rho|r_1\rangle\dots\langle r_{m_{k_i}}|\rho|r_{m_{k_i}}\rangle. \tag{26}$$

We start analyzing the squared expected value of the estimator in the original protocol

$$\mathop{\mathbb{E}}_{k_i,r}[\hat{X}_{k_i}^2] = \mathop{\mathbb{E}}_{k_i}\mathop{\mathbb{E}}_{r|k_i}\left[\left(\frac{1}{m_{k_i}b_{k_i}\sqrt{d}}\sum_{j=1}^{m_{k_i}} c_{k_i r_j}\right)^2\right] e = \mathop{\mathbb{E}}_{k_i}\mathop{\mathbb{E}}_{r|k_i}\left[\frac{1}{m_{k_i}^2 b_{k_i}^2 d}\left(\sum_{j=1}^{m_{k_i}} c_{k_i r_j}^2 + \sum_{j\neq l}^{m_{k_i}} c_{k_i r_j}c_{k_i r_l}\right)\right]. \tag{27}$$

The first term in the last line simplifies to m_{k_i}, since $c_{k_i r_j} \in \{-1, 1\}$. Calculating the conditional expectation, the second term can be rewritten as $\sum_{j\neq l}\mathbb{E}_{r|k_i}[c_{k_i r_j}c_{k_i r_l}] = \sum_{j\neq l}\mathbb{E}_{r_j|k_i}[c_{k_i r_j}]\,\mathbb{E}_{r_l|k_i}[c_{k_i r_l}]$, since the j-th and l-th samples are independent conditioned on k_i. Also,

$$\mathop{\mathbb{E}}_{r_j|k_i}[c_{k_i r_j}] = \sum_{r_j=1}^{d} c_{k_i r_j}\langle r_j|\rho|r_j\rangle = a_{k_i}\sqrt{d}. \tag{28}$$

Thus, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{k_i,r}[\hat{X}_{k_i}^2] &= \mathop{\mathbb{E}}_{k_i}\left[\frac{1}{m_{k_i}b_{k_i}^2 d}\right] + \mathop{\mathbb{E}}_{k_i}\left[\frac{1}{m_{k_i}^2 b_{k_i}^2 d}\sum_{j\neq l}^{m_{k_i}}\mathop{\mathbb{E}}_{r_j|k_i}\left[c_{k_i r_j}\right]\mathop{\mathbb{E}}_{r_l|k_i}\left[c_{k_i r_l}\right]\right] \\
&= \mathop{\mathbb{E}}_{k_i}\left[\frac{1}{m_{k_i}b_{k_i}^2 d}\right] + \mathop{\mathbb{E}}_{k_i}\left[\frac{1}{m_{k_i}^2 b_{k_i}^2}\sum_{j\neq l}^{m_{k_i}} a_{k_i}^2\right] \\
&= \mathop{\mathbb{E}}_{k_i}\left[\frac{1}{m_{k_i}b_{k_i}^2 d}\right] + \mathop{\mathbb{E}}_{k_i}\left[\frac{1}{m_{k_i}b_{k_i}^2}(m_{k_i}-1)a_{k_i}^2\right] \\
&\approx \frac{\ell\epsilon^2}{2\ln(2/\delta)}\left(\sum_{k_i} b_{k_i}^2 - d\sum_{k_i} b_{k_i}^2 a_{k_i}^2\right) + \sum_{k_i} a_{k_i}^2.
\end{aligned} \tag{29}$$

Here, we use the fact that the number of copies in the original protocol is given by Eq. (4), and assume the approximation $\lceil m_{k_i}\rceil \approx m_{k_i}$ holds.

We now compute the same quantity for our protocol. The calculation follows a similar approach, but with the key difference that the sampling is performed over groups of Pauli operators. Here, k_i corresponds to the selected group, and $c_{k_i r_j} = \sum_l c_{k_i l}^{(r_j)} b_{k_i l}$, where the sum is taken over the elements in the group, with $b_{k_i l}$ representing the coefficient associated to the Pauli l of the group g_k. One result we need is $\mathbb{E}_{r_j|k_i}[c_{k_i l}^{(r_j)}] = a_{k_i l}$. Then, we have that

$$\begin{aligned}
\mathop{\mathbb{E}}_{k_i,r}[\tilde{X}_{k_i}^2] &= \mathop{\mathbb{E}}_{k_i,r}\left[\left(\frac{1}{m_{k_i}\sqrt{d}\|\mathbf{b}_{k_i}\|^2}\sum_{j=1}^{m_{k_i}}\sum_{l=1}^{g_k} c_{k_i l}^{(r_j)} b_{k_i l}\right)^2\right] \\
&= \mathop{\mathbb{E}}_{k_i}\left[\frac{1}{m_{k_i}^2 d\|\mathbf{b}_{k_i}\|^4}\left(\sum_{j=1}^{m_{k_i}}\sum_{l=1}^{g_k}\mathop{\mathbb{E}}_{r_j|k_i}\left[\left(c_{k_i l}^{(r_j)} b_{k_i l}\right)^2\right] + \sum_{j\neq l=1}^{m_{k_i}}\sum_{l,t=1}^{g_k}\mathop{\mathbb{E}}_{r_j|k_i}\left[c_{k_i l}^{(r_j)}\right]\mathop{\mathbb{E}}_{r_l|k_i}\left[c_{k_i t}^{(r_l)}\right] b_{k_i l}b_{k_i t}\right)\right] \\
&= \mathop{\mathbb{E}}_{k_i}\left[\frac{1}{m_{k_i}d\|\mathbf{b}_{k_i}\|^4}\sum_{l=1}^{g_k} b_{k_i l}^2\right] + \mathop{\mathbb{E}}_{k_i}\left[\frac{1}{m_{k_i}^2\|\mathbf{b}_{k_i}\|^4}\sum_{j\neq l}^{m_{k_i}}\sum_{l,t=1}^{g_k} a_{k_i l}a_{k_i t}b_{k_i l}b_{k_i t}\right] \\
&\leq \mathop{\mathbb{E}}_{k_i}\left[\frac{1}{m_{k_i}d\|\mathbf{b}_{k_i}\|^2}\right] + \mathop{\mathbb{E}}_{k_i}\left[\frac{1}{m_{k_i}\|\mathbf{b}_{k_i}\|^2}(m_{k_i}-1)||\mathbf{a}_{k_i}||^2\right] \\
&\leq \frac{\ell\epsilon^2}{2\ln(2/\delta)}\left(\sum_{k_i}||\mathbf{b}_{k_i}||^2 - d\sum_{k_i}||\mathbf{b}_{k_i}||^2||\mathbf{a}_{k_i}||^2\right) + \sum_{k_i}||\mathbf{a}_{k_i}||^2 \\
&\leq \frac{\ell\epsilon^2}{2\ln(2/\delta)}\left(\sum_{k_i,l} b_{k_i l}^2 - d\sum_{k_i,l} b_{k_i l}^2 a_{k_i l}^2\right) + \sum_{k_i,l} a_{k_i l}^2.
\end{aligned} \tag{30}$$

In the fourth line, we used the definition of $||\mathbf{b}_{k_i}||^2$ from Eq. (7) for the first term, and applied the Cauchy-Schwarz inequality, $\left(\sum_{l=1}^{g_k} a_{k_i l} b_{k_i l}\right)^2 \leq ||\mathbf{b}_{k_i}||^2 ||\mathbf{a}_{k_i}||^2$, for the second term. In the fifth line, we substituted the expression for the number of copies m_{k_i} derived in Eq. (20), considering the approximation $\lceil m_{k_i} \rceil \approx m_{k_i}$. Additionally, we utilized the inequality $||\mathbf{b}_{k_i}||_2 \leq ||\mathbf{b}_{k_i}||_1$. From Eq.(30), it follows that the variance of our protocol is upper-bounded by the variance of the standard DFE protocol in the limit $m_{k_i} \gg 1$.