DFE Report

Ahmed Aman

April 2025

1 What is DFE and Why

DFE is a new technique that is implemented to estimate if any **arbitrary pure state** σ is close enough to the *desired/ideal* state ρ. The important requirement for being able to do useful computation is the ability to prepare the desired input states. A figure of merit relevant to input state preperation is to have minimum fidelity with which the state can be prepared. These desired states can be any, but we primarily work with Haar Random states, W states, and GHZ states. This estimation is done using a few "constant" numbers of Pauli measurements (in each *trial* run), selected at RANDOM according to an *importance weighted rule*. Roughly, Pauli operators that are most likely to detect deviations from the desired state ρ. While it's theoretically possible to measure the closeness of two quantum states, experimentally verifying this without first performing a full *Quantum State Tomography* (QST) is impossible. However, this new method is faster and requires fewer resources than QST. It allows for the validation of a quantum state's closeness to a desired state without needing to reconstruct the full state itself, and it can be applied to a wider range of states than the *Entanglement Witness* method.

- Quantum State Tomography is a 'process' by which a quantum state is reconstructed using known measurements (these have to form measurement operator basis on the Hilbert space, also known as tomographically complete) on an ensemble of identical quantum states which is produced by a source that consistently spits out either into pure states or general mixed state. This is very resource-intensive.
- In Quantum Information, Entanglement Witness is a *functional* which distinguishes a certain entangled state from a separable one.

We will also see later on that this method can be extended to channels, whereby you can quantify/analyze the action of the channel on a given state.

1.1 Setting of the problem:

- Hilbert Space $\mathcal{H}$ with $dim(\mathcal{H}) = d = 2^n$
- Assumptions:
  1. ρ is pure.
  2. We can measure n qubit Pauli observables, which are tensor products of single qubit Pauli operations.

3. There are no assumptions of additional structure or symmetry on ρ

- In order to obtain a distribution that falls within the predefined limits, we will execute the protocol for a specified desired state, $\mathcal{N}$ number of times (in my case, I conducted 1000 trials).
- A single run of estimation, which is at max 2ϵ far away from the true fidelity, will be obtained within a confidence interval of $(1-2\delta)\%$

2 Fidelity Estimation

The fidelity between ρ and σ,

$$F(\rho,\sigma) = tr([(\rho^{1/2}\sigma\rho^{1/2})^{1/2}]) = tr(\rho\sigma) \tag{1}$$

Here, we can write $tr(\rho\sigma)$ in terms of the Pauli expectation values of ρ and σ. Although there are $4n$ distinct Pauli operators, we only need to sample a constant number of them to estimate $F(\rho,\sigma)$ up to a constant additive error ϵ for arbitrary σ.

2.1 The Variables at hand:

- Let W_k $(k = 1, ..., d^2)$ denote all possible Pauli operators. In a d-dimensional Hilbert space, the space of all linear operators is itself d^2-dimensional. A convenient choice of orthonormal basis for these operators is the generalized Pauli basis, consisting of d^2 operators.
- We define the characteristic function

  $$\mathcal{X}_\rho(k) = tr(\rho W_k/d^{1/2}) \tag{2}$$

  which is basically the expectation value of ρ with W_k.
- We then have

  $$F(\rho,\sigma) = tr(\rho\sigma) = \sum_k \mathcal{X}_\rho(k)\mathcal{X}_\sigma(k) \tag{3}$$

  This can be proven as follows – For any Pauli operators W_k we have the identity $tr(W_k^\dagger W_l) = d\delta_{k,l}$. Since W_k forms an orthonormal basis, any operator A can be expanded as $A = \sum_{i=1}^{d^2} a_i W_i$, with the coefficients determined as $a_k = \frac{1}{d}tr(W_k^\dagger A)$. This is true for the following as well; $\rho = \sum_{k=1}^{d^2} r_k W_k$, and $\sigma = \sum_{j=l}^{d^2} s_l W_l$ with coefficients $r_k = \frac{1}{d}tr(W_k^\dagger \rho)$,and $s_l = \frac{1}{d}tr(W_l^\dagger \sigma)$ respectively. Using these, it is easy to see the required.
- As mentioned earlier, the estimation is done using a few randomly selected Pauli operators, which are selected according to an importance-weighted rule. Hence, the probability of selecting any one of $(k = 1, ..., d^2)$ is given by

  $$p(k) = (\mathcal{X}_\rho(k))^2 \tag{4}$$

  Now, this probability distribution might seem random, but it originates from the following,

  $$F(\rho,\sigma) = \sum_k r_k s_k = \sum_k r_k^2 \frac{s_k}{r_k} \tag{5}$$

  and the rest follows.

- We *construct* an estimator, which is considered a random variable as

$$X = \mathcal{X}_\sigma(k)/\mathcal{X}_\rho(k) \tag{6}$$

- The expectation value of this estimator X is $E(X) = \sum_k \ p(k)X = tr(\rho\sigma)$, which is basically the desired Fidelity that was mentioned earlier.

2.2 DFE Protocol summarized:

1. Generate ℓ independent and identically distributed random samples $k_1, ..., k_\ell$ from the importance sampling distribution $p(k_i)$, where $i \in \{1, .., \ell\}$ where

$$\ell := \lceil (\epsilon^2 \delta)^{-1} \rceil \tag{7}$$

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

$$m_i = \left\lceil \frac{2}{(\mathcal{X}_\rho(k_i))^2 d\ell\epsilon^2} \ln \frac{2}{\delta} \right\rceil. \tag{8}$$

   notice that m_i depends on k_i.

3. For each i, the measurement outcome of W_{k_i} on each m_i copies of σ we obtain an outcome $A_{ij} \in \{-1, 1\}$. We use these values of measurement to calculate the empirical estimate of the expectation value $tr\sigma W_{k_i}$. Using these estimates, we calculate the estimator $\tilde{X}_i$ given by

$$\tilde{X}_i = \frac{1}{m_i \sqrt{d} \chi_\rho(k_i)} \sum_{j=1}^{m_i} A_{ij}. \tag{9}$$

4. Calculate

$$\tilde{Y} = \frac{1}{\ell} \sum_{i=1}^{\ell} \tilde{X}_i \tag{10}$$

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

The fidelity estimate $\tilde{Y}$ 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_i is bounded and scales at most linearly with the Hilbert space dimension.

2.3 Couple of points to take note of:

- To estimate $tr(\rho\sigma)$ along with fixed additive error $\epsilon > 0$ and failure probability $\delta \ > 0$, we repeat the process a total of $\ell = \lceil 1/(\epsilon^2\delta) \rceil$ times: choose k_i independently, for which each time we have independent estimates X_i, that forms the random variable $Y = \frac{1}{\ell} \sum_{i=1}^{l} \ X_i$.

- From the appendix of the paper, it was shown that $Var(X_i) = tr(\sigma^2) - [tr(\rho\sigma)]^2 \leq 1$ which in turn implies $Var(Y) = Var(\frac{1}{d}\sum_{i=1}^{l} X_i) = \frac{1}{l^2}\sum_{i=1}^{l} Var(X_i) \leq \frac{1}{l^2}.l = \frac{1}{l}$. Now, according to Chebyshev's inequality,

$$Pr[|Y - tr(\rho\sigma)| \geq \frac{\lambda}{\sqrt{\ell}}] \leq \frac{1}{\lambda^2} \tag{11}$$

  then we set $\lambda = 1/\sqrt{\delta}$ and $\ell = \lceil 1/(\epsilon^2\delta) \rceil$. Which gives us,

$$Pr[|Y - tr(\rho\sigma)| \geq \epsilon] \leq \delta \tag{12}$$

  This inequality just shows us how far away our independent random variables X_i are from the mean/expectation, i.e, the fidelity.

- This ideal "infinite-precision" estimator Y can be approximated by an estimator $\tilde{Y}$ that uses a finite number of copies of the state σ.

2.4 Calculation of Required Sample Counts

This section details the derivation for the number of measurement settings, ℓ, and the number of measurements per setting, m_i, required to estimate the fidelity with a given precision and confidence. The calculations rely on two fundamental concentration inequalities: *Chebyshev's inequality* for ℓ and *Hoeffding's inequality* for m_i.

2.4.1 Number of Pauli Settings (ℓ)

First, we determine the number of distinct Pauli operators, ℓ, that we need to sample. This calculation is based on bounding the error of an estimated mean using **Chebyshev's inequality**.

Let's consider a simplified problem: estimating the expectation value $\mu = \mathrm{Tr}(\rho\sigma)$ of a single Pauli observable σ for a quantum state ρ. We perform ℓ independent measurements. Let X_i be the random variable for the outcome of the i-th measurement. The outcomes can only be $+1$ or -1.

The expectation value and variance of a single measurement outcome X_i are:

$$\mathbb{E}[X_i] = \mathrm{Tr}(\rho\sigma) \tag{13}$$

$$\mathrm{Var}(X_i) = \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2 \tag{14}$$

$$= \sum_k [\chi_\sigma(k)]^2 - [\mathrm{Tr}(\rho\sigma)]^2 \tag{15}$$

Since the measurement outcomes are either $+1$ or -1, we have $X_i^2 = 1$, which means $\mathbb{E}[X_i^2] = 1$. The expectation value $\mathrm{Tr}(\rho\sigma)$ is bounded by $[-1, 1]$, so its square is bounded by $[0, 1]$.

$$\mathrm{Var}(X_i) = 1 - [\mathrm{Tr}(\rho\sigma)]^2 \leq 1 \tag{16}$$

We use the sample mean of ℓ measurements as our estimator, Y:

$$Y = \frac{1}{\ell}\sum_{i=1}^{\ell} X_i \tag{17}$$

The expectation of our estimator Y is the true expectation value, making it an unbiased estimator:

$$\mathbb{E}[Y] = \mathbb{E}\left[\frac{1}{\ell}\sum_{i=1}^{\ell} X_i\right] = \frac{1}{\ell}\sum_{i=1}^{\ell}\mathbb{E}[X_i] = \frac{1}{\ell}\sum_{i=1}^{\ell}\mathrm{Tr}(\rho\sigma) = \mathrm{Tr}(\rho\sigma) \tag{18}$$

The variance of the estimator Y is:

$$\mathrm{Var}(Y) = \mathrm{Var}\left(\frac{1}{\ell}\sum_{i=1}^{\ell} X_i\right) = \frac{1}{\ell^2}\sum_{i=1}^{\ell}\mathrm{Var}(X_i) \leq \frac{1}{\ell^2}\sum_{i=1}^{\ell}(1) = \frac{\ell}{\ell^2} = \frac{1}{\ell} \tag{19}$$

Our goal is to ensure that our estimate Y is within a desired **precision** $\varepsilon > 0$ of the true value $\mathrm{Tr}(\rho\sigma)$, with a probability of at least $1-\delta$. In other words, we want the failure probability to be at most δ:

$$P\left[|Y - \mathrm{Tr}(\rho\sigma)| \geq \varepsilon\right] \leq \delta. \tag{20}$$

To find the required ℓ, we apply Chebyshev's inequality, which states for a random variable Y with mean μ_Y (Eqn. 54) and variance σ_Y^2:

$$P(|Y - \mu_Y| \geq \varepsilon) \leq \frac{\sigma_Y^2}{\varepsilon^2} \tag{21}$$

Substituting our known quantities: $\mu_Y = \mathrm{Tr}(\rho\sigma)$ and $\sigma_Y^2 = \mathrm{Var}(Y) \leq 1/\ell$.

$$P\left[|Y - \mathrm{Tr}(\rho\sigma)| \geq \varepsilon\right] \leq \frac{\mathrm{Var}(Y)}{\varepsilon^2} \leq \frac{1/\ell}{\varepsilon^2} = \frac{1}{\ell\varepsilon^2} \tag{22}$$

To satisfy our (Eqn. 56), we set this bound to be less than or equal to δ:

$$\frac{1}{\ell\varepsilon^2} \leq \delta \implies \ell \geq \frac{1}{\varepsilon^2\delta} \tag{23}$$

Since ℓ must be an integer, we take the ceiling of this value.

$$\ell = \lceil 1/(\varepsilon^2\delta) \rceil \tag{24}$$

This gives the minimum number of Pauli settings required.

2.4.2 Number of Measurements per Setting (m_i)

For each of the ℓ Pauli settings W_{k_i} chosen, we must perform a number of measurements, m_i, to estimate the corresponding expectation value. The number of state copies measured for each setting does not have to be the same. We use **Hoeffding's inequality** for this part of the calculation because it provides a tighter bound than Chebyshev's for variables with known finite bounds (like our measurement outcomes).

Given any choice of operators $W_{k_1}, \ldots, W_{k_\ell}$, we proceed as follows. For each $i = 1, \ldots, \ell$, we use m_i copies of the state ρ to estimate an intermediate quantity. We measure the Pauli observable W_{k_i} on each of these copies, getting measurement outcomes $A_{ij} \in \{1, -1\}$ for $j = 1, \ldots, m_i$. Note that $\mathbb{E}[A_{ij}] = \mathrm{Tr}(\rho W_{k_i})$.

Let's define an estimator $\tilde{Y}$ for a target quantity Y. The estimator $\tilde{Y}$ is constructed as a sum of scaled measurement outcomes.

$$\tilde{Y} = \sum_{i=1}^{\ell} \sum_{j=1}^{m_i} c_i A_{ij} \quad \text{where} \quad c_i = \frac{1}{\ell m_i \sqrt{d} \chi_\sigma(k_i)} \tag{25}$$

Here, $\chi_\sigma(k_i)$ are the coefficients of the target state σ in the Pauli basis, and d is the dimension of the system. The expectation value of our estimator is:

$$\mathbb{E}[\tilde{Y}] = \sum_{i=1}^{\ell} \sum_{j=1}^{m_i} c_i \mathbb{E}[A_{ij}] = \sum_{i=1}^{\ell} m_i c_i \mathrm{Tr}(\rho W_{k_i}) \tag{26}$$

$$= \sum_{i=1}^{\ell} m_i \left(\frac{1}{\ell m_i \sqrt{d} \chi_\sigma(k_i)} \right) \mathrm{Tr}(\rho W_{k_i}) \tag{27}$$

$$= \frac{1}{\ell} \sum_{i=1}^{\ell} \frac{\mathrm{Tr}(\rho W_{k_i})}{\sqrt{d} \chi_\sigma(k_i)} =: Y \tag{28}$$

Our goal is to ensure the estimation error $|\tilde{Y} - Y|$ is small. We require:

$$\Pr\big[|\tilde{Y} - Y| \geq \varepsilon\big] \leq \delta \tag{29}$$

We apply Hoeffding's inequality for a sum of independent random variables. For a sum $S = \sum_{k=1}^{N} Z_k$ where Z_k is a random variable bounded by $[a_k, b_k]$, the inequality is:

$$P\left(|S - \mathbb{E}[S]| \geq t\right) \leq 2 \exp\left(- \frac{2t^2}{\sum_{k=1}^{N} (b_k - a_k)^2} \right) \tag{30}$$

In our case, the estimator $\tilde{Y}$ is a sum of $N = \sum m_i$ independent random variables $Z_{ij} = c_i A_{ij}$. Since $A_{ij} \in \{-1, 1\}$, each variable Z_{ij} is bounded by $[-c_i, c_i]$. Thus, the range is $b_{ij} - a_{ij} = c_i - (-c_i) = 2c_i$.

The sum of the squared ranges in the denominator of the exponent is:

$$C = \sum_{i=1}^{\ell} \sum_{j=1}^{m_i} (2c_i)^2 = \sum_{i=1}^{\ell} m_i (4c_i^2) = 4 \sum_{i=1}^{\ell} m_i c_i^2 \quad \textit{with} \quad c_i = \frac{1}{\ell m_i \sqrt{d} \chi_\rho(k_i)}. \tag{31}$$

Plugging this into Hoeffding's inequality with our desired precision $t = \varepsilon$:

$$\Pr\big[|\tilde{Y} - Y| \geq \varepsilon\big] \leq 2 \exp\left(\frac{-2\varepsilon^2}{C} \right) \tag{32}$$

We want this probability to be at most δ. So we set $2 \exp(-2\varepsilon^2 / C) = \delta$, which gives the condition:

$$C = \frac{2\varepsilon^2}{\log(2/\delta)} \tag{33}$$

Now we substitute the expressions for C and c_i:

$$C = 4\sum_{i=1}^{\ell} m_i \left(\frac{1}{\ell m_i \sqrt{d}\chi_\sigma(k_i)}\right)^2 = \sum_{i=1}^{\ell} \frac{4}{\ell^2 m_i d \chi_\sigma(k_i)^2} \tag{34}$$

Combining these, we get the constraint on our choices of m_i:

$$\sum_{i=1}^{\ell} \frac{4}{\ell^2 m_i d \chi_\sigma(k_i)^2} \leq \frac{2\varepsilon^2}{\log(2/\delta)} \tag{35}$$

To satisfy this inequality, we make a simplifying choice by distributing the error budget equally across all ℓ terms in the sum. We enforce that each term must be less than or equal to $\frac{1}{\ell}$ of the total bound on the right-hand side.

$$\frac{4}{\ell^2 m_i d \chi_\sigma(k_i)^2} \leq \frac{1}{\ell}\left(\frac{2\varepsilon^2}{\log(2/\delta)}\right) = \frac{2\varepsilon^2}{\ell \log(2/\delta)} \tag{36}$$

Now, we can solve for a single m_i:

$$\frac{4}{\ell m_i d \chi_\sigma(k_i)^2} \leq \frac{2\varepsilon^2}{\log(2/\delta)} \tag{37}$$

$$4\log(2/\delta) \leq 2\varepsilon^2 \ell m_i d \chi_\sigma(k_i)^2 \tag{38}$$

$$m_i \geq \frac{4\log(2/\delta)}{2\ell d \chi_\sigma(k_i)^2 \varepsilon^2} = \frac{2\log(2/\delta)}{\ell d \chi_\sigma(k_i)^2 \varepsilon^2} \tag{39}$$

Since m_i must be an integer, we take the ceiling.

$$m_i = \left\lceil \frac{2}{d\chi_\sigma(k_i)^2 \ell \varepsilon^2} \log(2/\delta) \right\rceil. \tag{40}$$

This result specifies the number of measurements needed for each of the ℓ selected Pauli operators to achieve the desired overall accuracy.

2.4.3 An upper bound for m_i and total m:

It requires m copies of the state σ, where $m = \sum_{i=1}^{\ell} m_i$. Though this depends on the random choices k_i (there are d^2 of them). However, we can bound it in expectation:

$$\mathbb{E}(m_i) = \sum_{k_i} (\chi_\rho(k_i))^2 m_i \tag{41}$$

$$= \sum_{k_i} (\chi_\rho(k_i))^2 \left\lceil \frac{2}{d\chi_\sigma(k_i)^2 \ell \varepsilon^2} \log(2/\delta) \right\rceil \tag{42}$$

$$\leq 1 + \frac{2d}{\ell\varepsilon^2}\log(2/\delta), \quad \{\textit{by using the property } \lceil x \rceil \leq x+1\} \tag{43}$$

Similarly, we can calculate the expected number of total copies as

$$\mathbb{E}(m) \leq 1 + \frac{1}{\varepsilon^2 \delta} + \frac{2d}{\varepsilon^2}\log(2/\delta). \tag{44}$$

3 Numerics

- Tried to recreate the graph from the DFE paper.
- I have used $\ell = \lceil 1/(\varepsilon^2\delta) \rceil$ with $\varepsilon = 0.05$ and $\delta = 0.05$, which results in $\ell = 8000$.
- The number of qubits n varies for certain graphs, but nonetheless, it will be either 3 or 4.
- The states that I have used/worked with are:
  1. *Haar Random States*
  $$|\psi_{\text{Haar}}\rangle = \sum_{i=0}^{2^n-1} c_i|i\rangle, \quad \text{with} \quad \sum_{i=0}^{2^n-1} |c_i|^2 = 1, \tag{45}$$
  2. *W States*
  $$|W_n\rangle = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} |0\rangle^{\otimes(k-1)} \otimes |1\rangle \otimes |0\rangle^{\otimes(n-k)}, \tag{46}$$
  3. *GHZ States*
  $$|\text{GHZ}_n\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle^{\otimes n} + |1\rangle^{\otimes n}\right), \tag{47}$$
- The errors(10%) that I have used/worked with are:
  1. *Depolarizing Noise*
  $$\mathcal{E}_{\text{depol}}(\rho) = (1-p)\rho + p\,\frac{I^{\otimes n}}{2^n}. \tag{48}$$
  2. *Phase Damping Noise* (The equation below is for a single qubit.)
  $$\mathcal{E}_{\text{phase}}(\rho) = \sum_{k=0}^{1} E_k \rho E_k^\dagger, \; E_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{bmatrix}, \quad E_1 = \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{\gamma} \end{bmatrix}, \tag{49}$$
  3. *Amplitude Damping Noise* (The equation below is for a single qubit.)
  $$\mathcal{E}_{\text{amp}}(\rho) = \sum_{k=0}^{1} E_k \rho E_k^\dagger, \; E_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{bmatrix}, \quad E_1 = \begin{bmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{bmatrix}, \tag{50}$$
  4. *General Amplitude Damping* (The equations below are for a single qubit.): The effect of $\mathcal{E}_{\text{GAD}}$ on a single qubit is described by a set of four Kraus operators, E_0, E_1, E_2, E_3. These operators characterize the different ways the environment can interact with the qubit.
     - $E_0 = \sqrt{p} \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{bmatrix}$
       * This represents the scenario where the qubit remains in its state (or undergoes a small dephasing/amplitude damping related to $\sqrt{1-\gamma}$) and the environment remains in its excited state.

- $E_1 = \sqrt{p}\begin{bmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{bmatrix} \equiv \sqrt{p}\sqrt{\gamma}|0\rangle\langle 1|$
  - This describes the process where the qubit transitions from the excited state $|1\rangle \to |0\rangle$ (energy loss, "amplitude damping"), with the environment absorbing this energy.
- $E_2 = \sqrt{1-p}\begin{bmatrix} 0 & 0 \\ \sqrt{\gamma} & 0 \end{bmatrix} \equiv \sqrt{1-p}\sqrt{\gamma}|1\rangle\langle 0|$
  - This operator describes the reverse process: the qubit transitions from the ground state $|0\rangle \to |1\rangle$ (energy gain), because it absorbs energy from the environment. This is the crucial part that distinguishes GAD from simple amplitude damping, as it accounts for a finite-temperature environment.
- $E_3 = \sqrt{1-p}\begin{bmatrix} \sqrt{1-\gamma} & 0 \\ 0 & 1 \end{bmatrix}$
  - This operator represents the scenario where the qubit remains in its state (or undergoes a small dephasing/amplitude damping related to $\sqrt{1-\gamma}$) and the environment remains in its ground state.

- The left side of each graph gives the residual $(Y - F)$, where Y is our *estimation* of the *true fidelity* F. The right side just gives us the total number of copies of σ used throughout the whole trial.

- In the simulations, I have tried two different ways of simulating m_i measurement outcomes of the Pauli operators with the state σ.
  - First, I tried brute force calculation for each copy of σ by using NumPy and Qutip packages. This was very slow and took around 30 minutes for Haar Random states. This way of calculation also yielded very accurate estimates, so measurement errors had to be added.
  - Second, I tried a *Vectorized simulation of measurements* so that we can handle large m_i shots efficiently, rather than looping through each shot. This did produce a more accurate result, like that predicted by the theory. This method avoids looping through individual measurements. It first calculates the probability of a specific outcome (e.g., +1) from the quantum state's expectation value with that of the chosen Pauli observable for the $i'th$ iteration. It then uses a **binomial distribution** to directly compute the total number of +1 and -1 outcomes for all m_i measurements at once. This approach is equivalent to simulating each measurement individually.

- The two graphs below are without the vectorized measurements. **Figure 1** is without the measurement error and **Figure 2** is with measurement error of 5%. Do note that measurement errors reduce observed Pauli expectations $\langle W_{k_i} \rangle$ toward zero. This leads to:
  - Systematic underestimation of fidelity,
  - Cumulative bias across many Pauli terms.

  To correct this, we follow the steps below:

  1. Compute the attenuation factor:
  $$\beta = 1 - 2\alpha$$

in our case $\alpha = 0.05$

2. Correct each observed Pauli expectation value:

$$\langle P \rangle_{\text{corrected}} = \frac{\langle P \rangle_{\text{observed}}}{\beta}$$

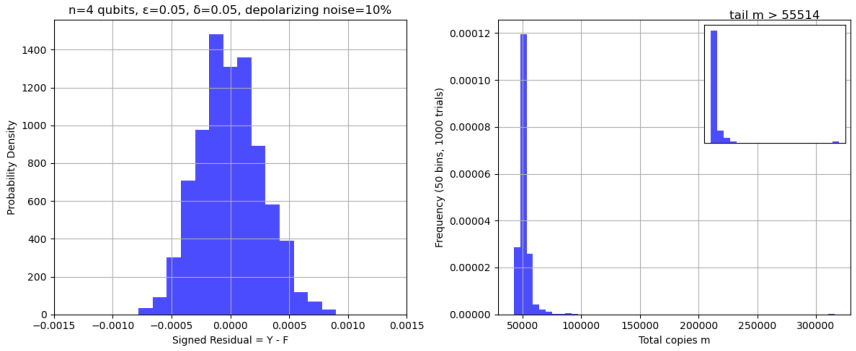

Figure 1: This is for the Haar state of $n = 4$ qubits without measurement error.

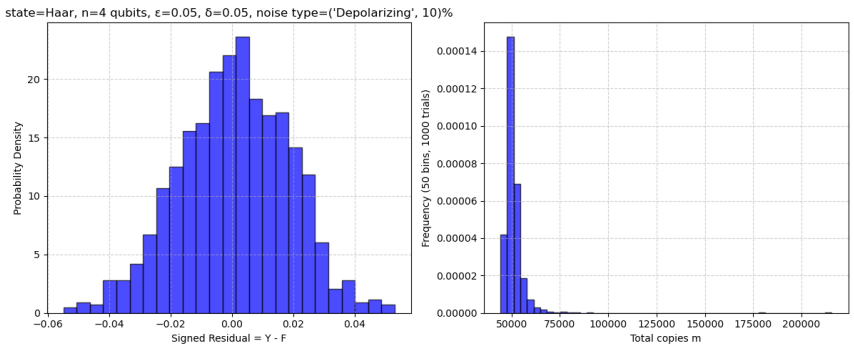

Figure 2: This is for the Haar state of $n = 4$ qubits with measurement error.

- Along with using the vectorized method, I conducted Monte Carlo sampling on 1,000 trials and extended it to 10,000 trials, as each trial had independent Fidelity estimates. I identified the best-fitting probability distribution from the 1,000 trials and used it to sample an additional 9,000 points to complete the curve.

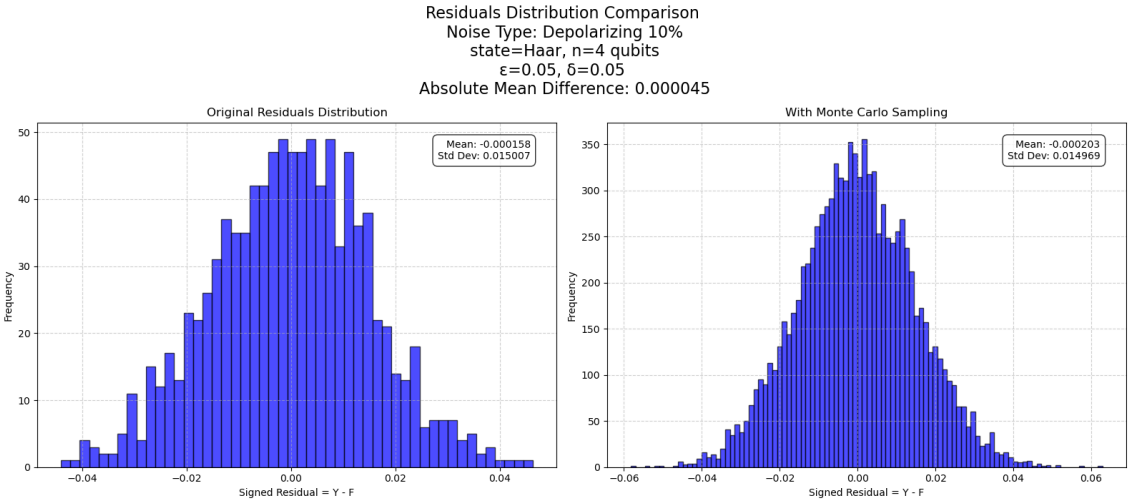

Figure 3: Distribution for the Residuals

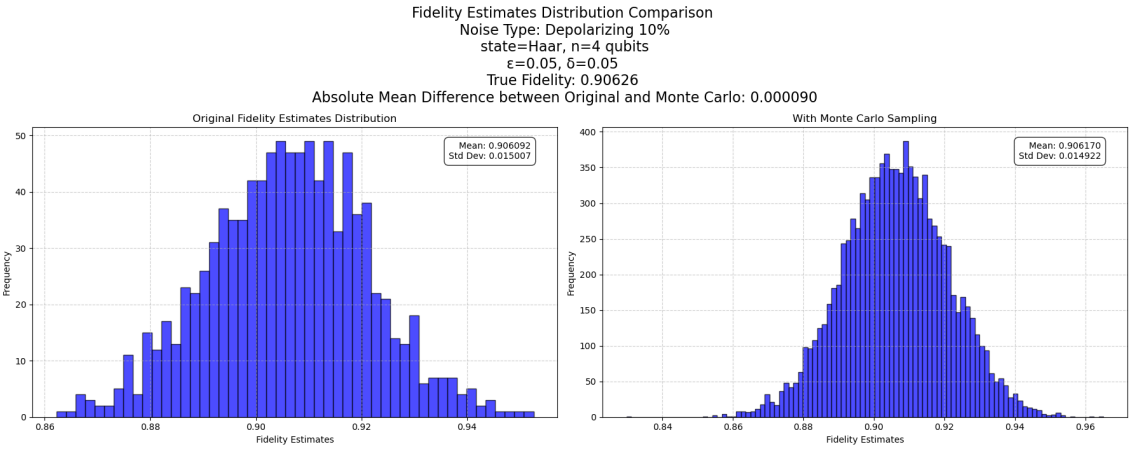

Figure 4: Distribution for the Fidelity Estimates

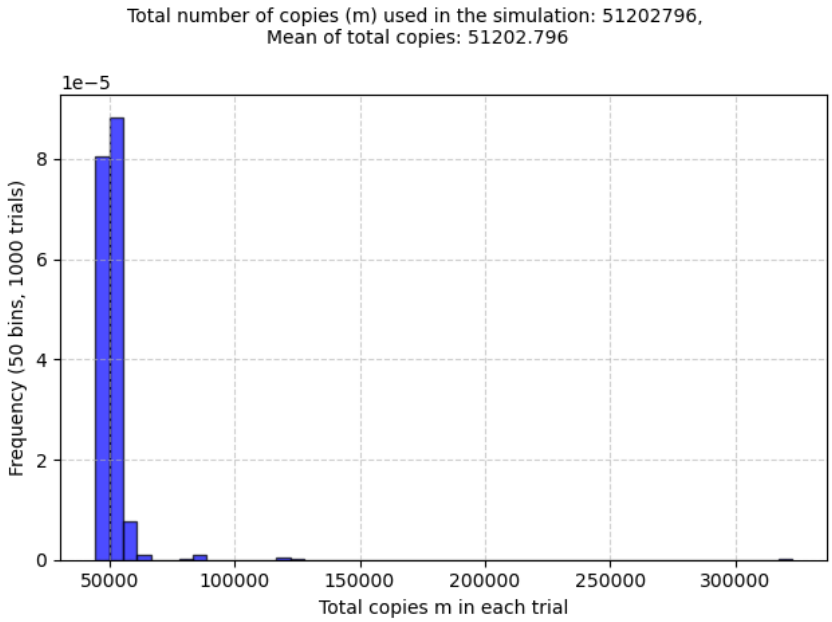

Figure 5: Distribution of the total copies

- ***Click here*** for the rest of the plots with different noise types, and the code for the simulation.
- I also ran the simulation for on a loop with increasing noise percentages from 15% to 25% with 1% increment in each loop. And the following plots were obtained. The plots are not unique to any errors. The estimates do not have Monte Carlo sampling, though they do use vectorized measurements.

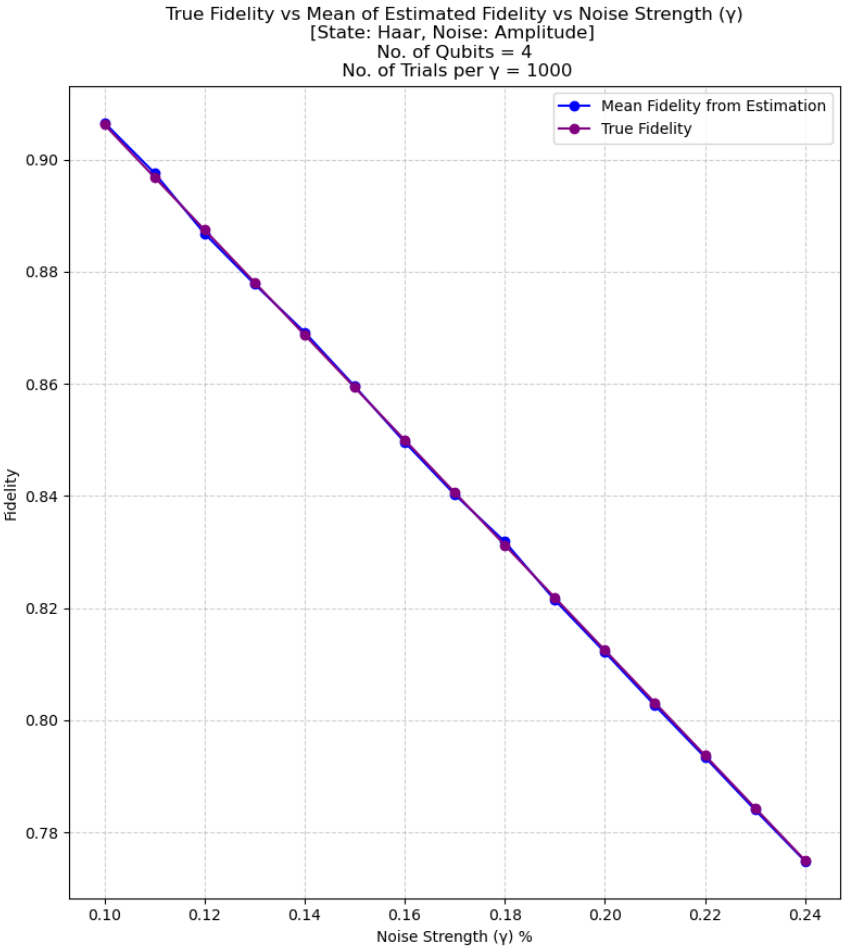

Figure 6: We can see that the estimates give a mean that is very close to the true value even for increased error rate.

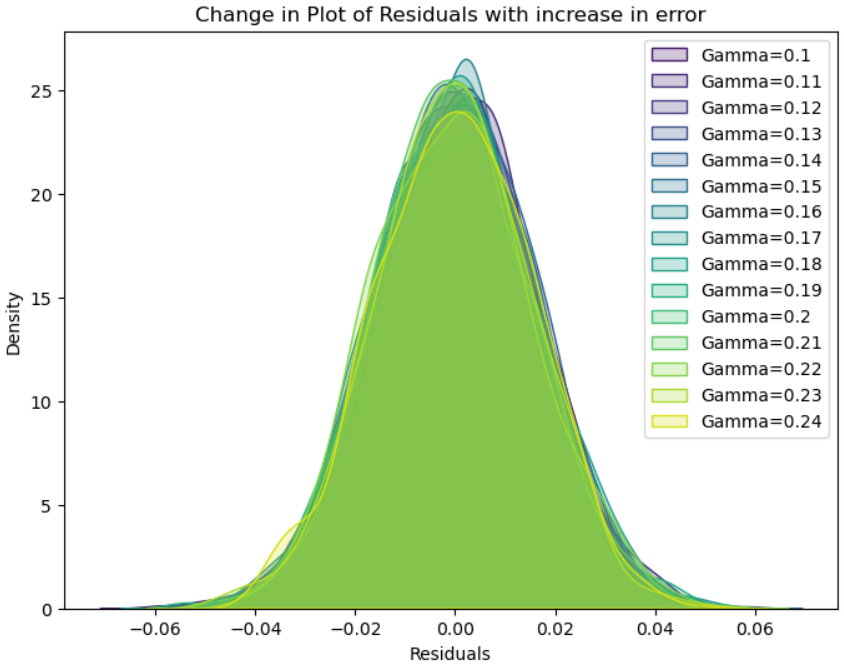

Figure 7: There is a general trend for the distribution to flatten with increase in error rate.