# 1 Scaling up protection

(Generalization of the previous PDD section; CDD content from Secs. IVA-D of draft by Jiaan & Xiansong. The main goal here is to show that an accuracy threshold for DD, under the CDD scheme for scaling up, fails to exist. This means that there is a maximum level for CDD, beyond which further levels only make things worse. To discuss the differences from using different measures.) We can distinguish between two operational considerations: to preserve states in a quantum memory (no computational gates), or to reduce noise in the course of carrying out a quantum computation. In the memory setting, the goal is to preserve the quantum informational state for some time $T$ without damage. During this time $T$, we can do DD pulses, and ask if the noise at the end of each complete DD sequence is lower than the noise over the same time period if there were no DD pulses. In other words, for DD to work well, we require

$$\text{Memory setting:} \quad \epsilon_{\text{DD}} < \epsilon(L), \tag{1}$$

where $\epsilon_{\text{DD}}$ quantifies the noise after the DD sequence, while $\epsilon(L)$ is the noise after bare evolution for the time taken for the $L$ DD pulses to be applied. In the computational setting, the comparison of with and without DD is different. With DD, computational gates can only be applied at the completion of each iteration of the DD sequence, as opposed to every time step without DD. The condition for DD to be effective in this case is then

$$\text{Computational setting:} \quad \epsilon_{\text{DD}} < \epsilon(1), \tag{2}$$

where $\epsilon(1)$ is the noise after bare evolution for a single time step.

These two perspectives result in very different overall decoupling strategies. Especially if one is using CDD to "scale up protection". As illustrated in Fig. 1, for the quantum memory setting, as total time is fixed $T$, one only need to come up with finer and finer slicing of the time evolution pieces; but for the quantum computation setting, one need to join the DD control gates to form a dilated time evolution map. The past papers in favor of CDD had taken the former perspective. It is undeniable that with finer and finer slicing of time evolution, one gets better and better performance. But this picture has some series drawbacks. First the total evolution time $T$ is typically unknown, even for quantum memory. Hence
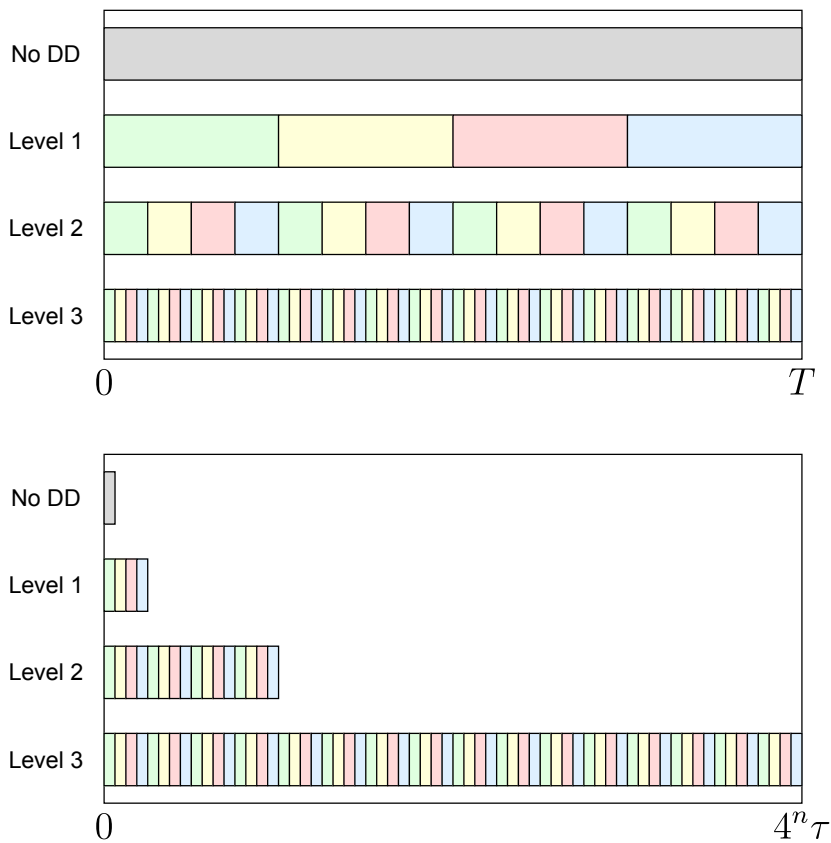
Figure 1: An illustration of the CDD time evolution maps under increasing concatenation levels. We compare the quantum memory setting (above) to the quantum computation setting (under).

there is no way of knowing in advance which gate to perform at what time. Second, there exists a minimal switch/cool-down time between consecutive gates due to technological limitations. Hence it is impractical to arbitrarily suppress noise by increasing concatenation level.

In our work, we focus on the computational setting, the more stringent one among the two, though it is straightforward to modify our analysis for the memory case. As we will see, Condition (2) will yield a requirement on the noise parameters characterizing the noise in the system and the DD pulses. We refer to that requirement as the "break-even point" for DD, borrowing terminology from quantum error correction and fault-tolerant quantum computing. (There was a comment in the .tex file that some past papers have analyzed the memory case. What are those papers? Do you just mean the Khodjasteh-Lidar 2007 paper? That paper looks only at imperfections due to finite-width pulses, not control errors as we do here. Their paper hence only has a single-parameter $\Phi_{SB}$, not our two-parameter $\Phi_{SB}$ and $\eta$ situation. In fact, in current experiments $\eta$ is likely the larger problem; memory noise is usually much smaller than gate noise. Or are there other earlier papers?) (Yes, that paper in particular. Their conclusion that CDD decouples to order $n$ and dramatically outperforms PDD come from this perspective. Our paper is focused on the quantum computation case which is more practical. See my new added picture 1). And let me quote their words (bottom left page 10) on how they do the comparison: "The higher order data for PDD are obtained by simply repeating the basic sequence and shrinking the pulse interval. This is necessary to get the long-term behavior of decoupling. . . ." I move this memory/computation argument from the previous section as it is more appropriate when we discuss the scaling behavior. —Jiaan)

(To be merged with the new text.) The conditions for achieving noise suppression are already derived previously for the idealistic PDD and CDD case. In general, we find that smaller $(\alpha, \beta)$ values would lead to to better noise suppression rate. Therefore the gates should be performed as frequently as possible. But this may leads to suboptimal outcomes in practice given the extra noise introduced by imperfect gate operations. To study the fault-tolerance of DD, we must account for the additional noise associated with the control gates.

We can certainly discuss the break-even point for a particular $\mathsf{CDD_n}$

sequence, treating it simply as a "flattened" sequence of pulses, and ask when the error phase after $\mathsf{CDD_n}$ is reduced:

$$\Phi_{\mathrm{SB},n} < \phi_{\mathrm{SB}}, \tag{3}$$

as was done for PDD in the previous section. However, what is more interesting is the question of the so-called accuracy threshold for CDD. As described in the Preliminaries section, CDD achieves higher and higher decoupling order by concatenating the same basic DD scheme—PDD for our discussion here—to higher and higher levels. We are interested in understanding how the noise removal improves as the DD scheme scales up. An accuracy threshold for CDD exists if the parameters governing the noise of the situation satisfying a condition $\eta < \eta_0$, such that

$$\Phi_{\mathrm{SB},n+1} < \Phi_{\mathrm{SB},n} \qquad \forall n = 1, 2, \ldots. \tag{4}$$

In fault-tolerant quantum computing, there is the concept of a fault tolerance noise threshold, a strength of the noise below which scaling up the QEC code leads to improved protection against noise, and hence more accurate quantum computation. This threshold is referred to as the accuracy threshold. Here, we ask the analogous question of CDD, whether there is a condition on the noise parameter of the problem such that increasing the concatenation level always leads to improved noise removal capabilities. If such condition can not be satisfied, we ask the question of whether there exists an optimal concatenation level and how the imperfection in gate control could impact it.

## 1.1 Ideal case

Should the accuracy threshold exist, then Eq. (4) must first and foremost apply in the case where the gate noise is non-existent. Below, we examine this condition for the ideal CDD before examining on the noisy version.

Following the approach of Khodjasteh and Lidar, the $\mathsf{CDD_n}$ generator in the ideal case can be estimated as:

$$\begin{aligned}
\Omega_n \approx\ & \sigma_0 \otimes 4^n B_0 \\
& + \sigma_1 \otimes (-\mathrm{i})^n 2^{n(n+1)} \, \mathrm{ad}_{B_0}^n(B_1) \\
& + \sigma_2 \otimes (-\mathrm{i})^n 2^{n^2} \, \mathrm{ad}_{B_0}^{n-1}\big(\, [B_0, B_2] - \mathrm{i}\{B_1, B_3\} \,\big).
\end{aligned} \tag{5}$$

4

It is shown in the appendix A that the above expression indeed captures the leading order behavior of $\Omega_n$. Hence $\mathsf{CDD_n}$ indeed achieves $n$th-order decoupling. Different from the leading order term in the Magnus series, Eq. (5) combines the leading order pure bath part and the interaction part. The pure bath part, whose size can be estimated by $\Phi_{\mathrm{B},n} \simeq 4^n \phi_\mathrm{B}$, is of first order; whereas the interaction part is of $(n+1)$th order. Focusing on the leading order interaction part, we can derive an upper bound for for the error phase:

$$\Phi_{\mathrm{SB},n} \lesssim 2^{n(n+2)} \phi_\mathrm{B}^n \, \phi_{\mathrm{SB}} \, f_n\!\left(\frac{\phi_{\mathrm{SB}}}{\phi_\mathrm{B}}\right), \tag{6}$$

where the auxillary $f_n$ function is defined by:

$$f_n(x) \equiv \begin{cases} 1, & x \le \sqrt{4^n - 1}, \\[2mm] \dfrac{1}{2^n} \sqrt{\left(\dfrac{4^n - 1}{2x} + \dfrac{x}{2}\right)^2 + 1}, & x > \sqrt{4^n - 1}. \end{cases} \tag{7}$$

As mentioned in the Preliminaries section, $\mathsf{CDD_n}$ achieves $n$th-order decoupling. This can be understood from the error phase expression: $\Phi_{\mathrm{SB},n}$ after $\mathsf{CDD_n}$ is of order $\phi^{n+1}$, for $\phi \sim \phi_\mathrm{B}, \phi_{\mathrm{SB}}$. That $\phi_\mathrm{B}$ enters the error phase should again be of no surprise—as in the PDD case, $\phi_\mathrm{B}$ determines how quickly the noise seen by the system evolves, and hence affects the efficacy of DD which does a good elimination of the noise only when the noise remains nearly unchanged for the full DD sequence.

What is perhaps more surprising is the appearance of the $2^{n(n+2)}$ factor in $\Phi_{\mathrm{SB},n}$, a important feature in the CDD. For fixed $\tau_0 \equiv \tau$ as $n$ increases (e.g., if $\tau$ is the experimental limit for the switch time between consecutive pulses), such factors mean that the error phase eventually increases for large enough $n$: The exponentially decreasing $\phi^n$ factor is eventually overcome by the super-exponentially increasing $2^{n^2}$ factor. This means there is no accuracy threshold, i.e., there is no level of noise that is weak enough such that $\Phi_{\mathrm{SB},n+1} \le \Phi_{\mathrm{SB},n}$ for all $n$. Instead, there is an maximal useful level of concatenation level, beyond which further concatenation actually increases the noise seen by the system. Fig. 2 plots this situation of fixed $\tau_0$, for increasing concatenation level $n$. We observe the initial decrease of $\Phi_{\mathrm{SB},n}$ as $n$ increases, but this turns around eventually.
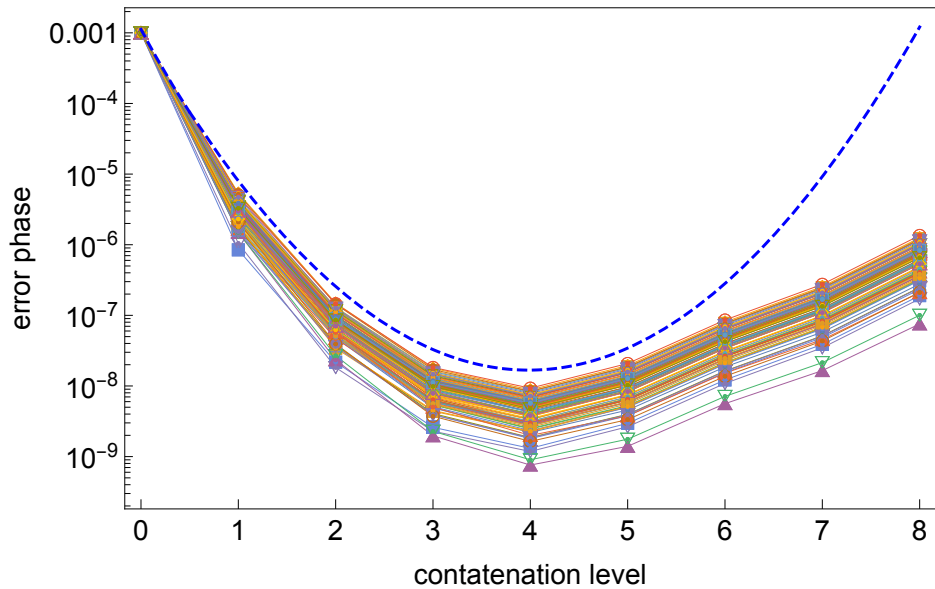
Figure 2: The error phase as a function of the concatenation level. The thin colored lines are obtained from random samples satisfying $\phi_{\mathrm{B}} = \phi_{\mathrm{SB}} = 0.001$. The blue dashed line is our theoretical upper bound for error phase. The maximal concatenation level occurs at 4 in this configuration.

The upper bound (6) alone is insufficient for predicting when will $\Phi_{\text{SB},n+1} < \Phi_{\text{SB},n}$. As that would require both an upper and lower bound. For higher concatenation level to make sense, increasing concatenation level should decrease the error phase. To derive a sufficient condition of it, we examine the recursive relation between concatenation levels, which suggests

$$
\begin{aligned}
B_{n+1,0} &= 4B_{n,0} \\
B_{n+1,1} &= -4\mathrm{i}[B_{n,0}, B_{n,1}] \\
B_{n+1,2} &= -2\mathrm{i}[B_{n,0}, B_{n,1}] \\
B_{n+1,3} &= 0
\end{aligned}
\tag{8}
$$

For both $B_1$ and $B_2$ to decrease in size, it suffice to require the map $-4\mathrm{i}[B_{n,0}, \cdot]$ a contraction map. By requiring its norm to be smaller than 1, we obtain the bound for the maximal concatenation level:

$$
n_{\max} \leq -\log_4(\phi_{\text{B}}) - 3/2.
\tag{9}
$$

Note that the existence of a maximal useful concatenation level does not technically contradict the statement that higher-order decoupling is achieved by higher level CDD. Using decoupling order to quantify the degree of noise removal requires, in the first place, the convergence of the Magnus series, so that the $(n+1)$th- and higher-order terms in the series are a small correction to the first $n$ terms. However, if the DD scheme is designed such that the total time for the sequence grows exponentially with $n$, as is the case for CDD if $\tau_0$ is fixed as $n$ increases, eventually, we exceed the convergence criterion and the decoupling order stops being a reasonable indicator of successful noise removal.

We can, however, analyze a different situation: to have fixed $\tau_n \equiv T$, so that the CDD sequence takes the same amount of time, regardless of $n$. This requires $\tau_0 = \frac{T}{4^n}$ for each $n$, so that the pulses are applied at shorter and shorter time intervals as the concatenation level increases. This can be the practical choice if one is doing computation where computational gates, which can only be applied at the end of a complete DD sequence in order to not interfere with the noise averaging process, have to be applied at a particular clock rate. Increasing $n$ to increase the noise removal capabilities must not increase the total time taken for the DD sequence. For $\mathsf{CDD_n}$, the time is only $4^{-n}$ fraction of the full evolution. We define

the fixed bath operators
$$A_i \equiv 4^n B_{0,i}. \tag{10}$$

Eq. (5) still holds, but we need to rewrite it in terms of the fixed bath operators. This leads to

$$
\begin{aligned}
\Omega_n \approx{} & \sigma_0 \otimes A_0 \\
& + \sigma_1 \otimes (-\mathrm{i})^n 2^{-n^2-n} \, \mathrm{ad}_{A_0}^n (A_1) \\
& + \sigma_2 \otimes (-\mathrm{i})^n 2^{-n^2-2n} \, \mathrm{ad}_{A_0}^{n-1} \big( [A_0, A_2] - \mathrm{i}\{A_1, A_3\} \big).
\end{aligned}
\tag{11}
$$

Apparently, $\Phi_\mathrm{B} = \phi_\mathrm{B}$ remains unchanged. The error phase is reduced to

$$\Phi_{\mathrm{SB},n} \lesssim 2^{-n^2} \phi_\mathrm{B}^n \, \phi_\mathrm{SB} \, f_n\!\Big(\frac{\phi_\mathrm{SB}}{\phi_\mathrm{B}}\Big). \tag{12}$$

The upper bound is a super-exponentially decreasing function in $n$. This suggests that noise can be arbitrarily suppressed by increasing the concatenation level. This requires, of course, the ability to do faster and faster pulse switching. Eventually, of course, this becomes limited by the technological—not fundamental—constraint of how fast we can switch between pulses, and thus how short $\tau_0$ can be in an experiment.

## 1.2 Noisy pulses

As CDD is derived from PDD, the noisy CDD is also derived from the noisy PDD. Let us suppose that the noisy PDD maps the Lie algebra generator through $\widetilde{\mathcal{D}}$:

$$\widetilde{\Omega}_\mathsf{PDD} = \widetilde{\mathcal{D}}(\Omega). \tag{13}$$

For noisy $\mathsf{CDD_n}$, since we are using the same set of noisy gates as PDD,

$$\widetilde{\Omega}_{n+1} = \widetilde{\mathcal{D}}(\widetilde{\Omega}_n). \tag{14}$$

The threshold condition becomes:

$$\Phi_\mathrm{SB}(\widetilde{\mathcal{D}}(\widetilde{\Omega}_n)) \leq \Phi_\mathrm{SB}(\widetilde{\Omega}_n). \tag{15}$$

This is exactly the breakeven condition for noisy PDD with the free evolution generator $\Omega$ replaced by $\widetilde{\Omega}_n$. Let us characterize the gate noise level by $\eta$, and write the breakeven condition for PDD:

$$\eta \leq \eta(\phi_\mathrm{B}, \phi_\mathrm{SB}). \tag{16}$$

The CDD threshold condition becomes

$$\eta \le \eta(\widetilde{\Phi}_{\mathrm{B},n}, \widetilde{\Phi}_{\mathrm{SB},n}). \tag{17}$$

The left hand side is fixed for a given set of gates, while $\widetilde{\Phi}_{\mathrm{B},n}$ and $\widetilde{\Phi}_{\mathrm{SB},n}$ changes with the concatenation level $n$. Hence one useful way to look at the problem is to start with an error-tolerance diagram and view the pair $(\widetilde{\Phi}_{\mathrm{B},n}, \widetilde{\Phi}_{\mathrm{SB},n})$ as points evolving on the graph. Besides the graph itself, we also need to known the update rule $(\widetilde{\Phi}_{\mathrm{B},n}, \widetilde{\Phi}_{\mathrm{SB},n}) \to (\widetilde{\Phi}_{\mathrm{B},n+1}, \widetilde{\Phi}_{\mathrm{SB},n+1})$ for such evolution. This is, by construction of CDD, the same rule for the PDD transformation $(\phi_{\mathrm{B}}, \phi_{\mathrm{SB}}) \to (\widetilde{\Phi}_{\mathrm{B},1}, \widetilde{\Phi}_{\mathrm{SB},1})$. According to the study on PDD, we can write it as:

$$\widetilde{\Phi}_{\mathrm{B},n+1} \approx 4\widetilde{\Phi}_{\mathrm{B},n}, \tag{18}$$

$$\widetilde{\Phi}_{\mathrm{SB},n+1} \lesssim \widetilde{\Phi}_{\mathrm{B},n}\widetilde{\Phi}_{\mathrm{SB},n} + c. \tag{19}$$

We visualized such evolution in Fig.

For the quantum memory setting, the $\Phi_{\mathrm{B}}$ part is constant and the error phase updates according to:

$$\widetilde{\Phi}_{\mathrm{SB},n+1} \le \alpha\,\widetilde{\Phi}_{\mathrm{SB},n} + c. \tag{20}$$

Solving the sequence iteration rule leads to the general formula:

$$\widetilde{\Phi}_{\mathrm{SB},n} = \alpha^n \phi_{\mathrm{SB}} + \frac{c}{1 - \alpha}. \tag{21}$$

This indicates that despite the threshold do exists, increasing the concatenation level indefinitely will not achieve arbitrarily noise suppression. The smallest error phase that can be obatined is the CDD error phase limit.

9

# A    Analyzing the CDD dynamics

The unitary dynamics $U_n$ for $\mathsf{CDD_n}$ can be completely described by its Hermitian generator (dimensionless Hamiltonian), $\Omega_n \equiv i \log(U_n)$. It is more convenient to focus on the generators and regard DD as transformation among the generators.

At the ground level, we have the bare Hamiltonian $\Omega \equiv \tau H$. CDD is equivalent to PDD at the first level, producing $\Omega_1 \equiv \Omega_{\mathsf{PDD}}$, which can be further expressed as a series $\sum_{m=1}^{\infty} \Omega_1^{(m)}$ through Magnus expansion. To characterize this process, we introduce the $\mathcal{F}$ and $\mathcal{F}^{(m)}$ maps, defined through:

$$\Omega_1 = \mathcal{F}(\Omega), \quad \Omega_1^{(m)} = \mathcal{F}^{(m)}(\Omega), \tag{22}$$

with $\mathcal{F} = \sum_{m=1}^{\infty} \mathcal{F}^{(m)}$ in correspondence with the Magnus series. In particular, $\mathcal{F}^{(1)}$ and $\mathcal{F}^{(2)}$ are already given in the main text by Eq. (**??**). Higher level $\mathsf{CDD_n}$ is defined by the iteration map

$$\Omega_n = \mathcal{F}(\Omega_{n-1}) = \sum_{m=1}^{\infty} \mathcal{F}^{(m)}(\Omega_{n-1}). \tag{23}$$

Backtracking the map from level $n$ to level 0, we have $\Omega_n = (\mathcal{F})^n(\Omega)$. In the end, $\Omega_n$ can be expanded as a Magnus series in the bare Hamiltonian.

$$\Omega_n = \sum_{m=1}^{\infty} \Omega_n^{(m)}(\Omega), \tag{24}$$

where $\Omega_n^{(m)}$ is the $m$th order term for $\Omega_n$ in $\Omega$ such that $\Omega_n^{(m)} \sim \|\Omega\|^m$.

The full $\Omega_n$ is typically difficult to solve. To estimate $\Omega_n$, it suffice to find an analytically solvable estimator $\widehat{\Omega}_n$ which reflects leading order behavior of the full series. The leading order behavior is required separately for the pure-bath part and the system-bath coupling part, as these two parts can be of different orders. Formally, we define an estimator-error pair

$$\Omega_n = \widehat{\Omega}_n + \delta\Omega_n. \tag{25}$$

For a faithful estimator, the error term $\delta\Omega_n$ must be of higher order smallness than the estimator both in the pure-bath part and the coupling part.

To better quantify this statement, we take the two-component norm for both $\delta\Omega_n$ and $\widehat{\Omega}_n$, and demand

$$|||\delta\Omega_n||| \equiv \begin{pmatrix} \delta\alpha_n \\ \delta\beta_n \end{pmatrix} \ll |||\widehat{\Omega}_n||| \equiv \begin{pmatrix} \widehat{\alpha}_n \\ \widehat{\beta}_n \end{pmatrix}, \tag{26}$$

where the comparison is implied for both components and should be understood by comparing the leading powers of the polynomials in $\alpha$ and $\beta$: a "smaller" polynomial should have a larger leading power.

The original CDD paper (Ref.[1]) constructed an estimator by keeping the first two Magnus terms for each iteration step. In our notation:

$$\widehat{\Omega}_n \equiv (\mathcal{F}^{(1)} + \mathcal{F}^{(2)})(\widehat{\Omega}_{n-1}) = \left(\mathcal{F}^{(1)} + \mathcal{F}^{(2)}\right)^n(\Omega). \tag{27}$$

After a little bit algebra, the estimator can shown to be specified by the formula:

$$\begin{aligned} \widehat{\Omega}_n = {} & \sigma_0 \otimes 4^n B_0 \\ & + \sigma_1 \otimes (-\mathrm{i})^n 2^{n(n+1)} \operatorname{ad}_{B_0}^n(B_1) \\ & + \sigma_2 \otimes (-\mathrm{i})^n 2^{n^2} \operatorname{ad}_{B_0}^{n-1}\left([B_0, B_2] - \mathrm{i}\{B_1, B_3\}\right). \end{aligned} \tag{28}$$

For simplicity, we make the recognition $\|B_0\| \simeq \|B_i\| \simeq \tau$, where $\simeq$ represents that the both sides are of leading power smallness when writing as a polynomial in $\tau$. After taking the two-component norm of the estimator, we have,

$$\begin{pmatrix} \widehat{\alpha}_n \\ \widehat{\beta}_n \end{pmatrix} \simeq \begin{pmatrix} \tau \\ \tau^{n+1} \end{pmatrix}. \tag{29}$$

To show that $\widehat{\Omega}_n$ is indeed faithful, $\delta\alpha_n$ and $\delta\beta_n$ need to be of higher order smallness compared to $\widehat{\alpha}_n$ and $\widehat{\beta}_n$. We claim that:

$$\begin{pmatrix} \delta\alpha_n \\ \delta\beta_n \end{pmatrix} \lesssim \begin{pmatrix} \tau^3 \\ \tau^{n+2} \end{pmatrix}. \tag{30}$$

We prove this assertion by mathematical induction.

At $n = 1$, i.e. for PDD, the estimation error comes solely from the Magnus series truncation. The leading order is the third order:

$$|||\delta\Omega_1||| \simeq |||\Omega_1^{(3)}||| \simeq \begin{pmatrix} \tau^3 \\ \tau^3 \end{pmatrix}. \tag{31}$$

At higher truncation level, the estimation error is related with the lower level terms by:

$$\delta\Omega_{n+1} = \mathcal{F}(\Omega_n) - (\mathcal{F}^{(1)} + \mathcal{F}^{(2)})(\widehat{\Omega}_n)$$
$$= \mathcal{F}^{(1)}(\delta\Omega_n) + \sum_{m=2}^{\infty} \mathcal{F}^{(m)}(\Omega_n) - \mathcal{F}^{(2)}(\widehat{\Omega}_n), \tag{32}$$

where we have used the fact that $\mathcal{F}^{(1)}$ is a linear map. To properly bound the size of the error term, we take the two-component norm on both sides of Eq. (32):

$$\||\delta\Omega_{n+1}\|| \lesssim \||\mathcal{F}^{(1)}(\delta\Omega_n)\|| + \||\sum_{m=3}^{\infty} \mathcal{F}^{(m)}(\Omega_n)\|| \tag{33}$$
$$+ \||\mathcal{F}^{(2)}(\Omega_n) - \mathcal{F}^{(2)}(\widehat{\Omega}_n)\||.$$

We now need to show $\||\delta\Omega_{n+1}\|| \lesssim (\tau^3, \tau^{n+3})$. The original CDD paper by Khodjasteh and Lidar has only showed that the higher order Magnus series $\sum_{m=3}^{\infty} \mathcal{F}^{(m)}(\widehat{\Omega}_n)$ is small. However Eq. (32) suggests that this argument is not sufficient: as the estimate error $\delta\Omega_{n+1}$ comes not only from the truncation error of the same level, but also propagates from the lower level error $\delta\Omega_n$.

We now examine the size of these terms. The first term in (33) is straightforward,

$$\||\mathcal{F}^{(1)}(\delta\Omega_n)\|| = \begin{pmatrix} 4\delta\alpha_n \\ 0 \end{pmatrix} \simeq \begin{pmatrix} \tau^3 \\ 0 \end{pmatrix}.$$

The second term involves a infinite sum of Magnus terms higher than third order, which is lead by the third order term, whose size can be bounded by

$$\||\mathcal{F}^{(3)}(\Omega_n)\|| \lesssim \begin{pmatrix} \beta_n^2(\alpha_n + \beta_n) \\ \beta_n(\alpha_n + \beta_n)^2 \end{pmatrix} \simeq \begin{pmatrix} \tau^{2n+2} \\ \tau^{n+3} \end{pmatrix},$$

where we have used $\alpha_n \simeq \widehat{\alpha}_n$ and $\beta_n \simeq \widehat{\beta}_n$ according to the induction hypothesis. The above estimation can be shown by directly calculating the third order Magnus term for PDD. Or observing that for the higher PDD Magnus terms, at least one $B_{i\neq 0}$ is involved in the coupling part and

at least two $B_{i \neq 0}$ is involved in pure bath part. To bound the third term in (33), we can calculate the difference $\mathcal{F}^{(2)}(\widehat{\Omega} + \delta\Omega) - \mathcal{F}^{(2)}(\widehat{\Omega})$ using the expression for the second order Magnus term from Eq. (??). In the end,

$$\||\mathcal{F}^{(2)}(\Omega) - \mathcal{F}^{(2)}(\widehat{\Omega})|\| \lesssim \begin{pmatrix} 0 \\ \delta\alpha_n\beta_n + \alpha_n\delta\beta_n \end{pmatrix} \simeq \begin{pmatrix} 0 \\ \tau^{n+3} \end{pmatrix}.$$

Since all three terms are bounded by $(\tau_0^3, \tau_0^{n+3})^{\mathsf{T}}$, so is their sum. This proves our assertion Eq. (30). $\qquad\square$

Having showed its faithfulness, it must be the case that $\widehat{\Omega}_n$ contains at least the leading order terms of the full $\Omega_n$. We are now confident in using the estimator Eq. (28) to study the full $\Omega_n$ of $\mathrm{CDD}_n$.

# References

[1] Kaveh Khodjasteh and DA Lidar. Fault-tolerant quantum dynamical decoupling. *Physical review letters*, 95(18):180501, 2005.